UNIVERSITY OF SOUTHAMPTON

Uncertainty of Variance Estimators in Analytical and Process Variability Studies

by

Marion J Chatfield

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

in the

Faculty of Social, Human and Mathematical Sciences

Mathematical Sciences

June 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Mathematical Sciences

Doctor of Philosophy

UNCERTAINTY OF VARIANCE ESTIMATORS IN ANALYTICAL AND PROCESS VARIABILITY STUDIES

by Marion J Chatfield

This thesis demonstrates that the half-t distribution is the prior of choice for estimating uncertainty of variance estimators in routine analysis of analytical and process variance components studies.

Industrial studies are often performed to estimate sources of variation e.g. to improve and quantify measurement or process capability. Understanding the uncertainty of those estimators is important, especially for small studies. A Bayesian analysis is proposed – providing a flexible methodology which easily copes with the complex and varied nature of the studies and the varied quantities of interest.

The prior is a fundamental component of a Bayesian analysis. The choice of prior is appraised and the coverage of the credible intervals obtained using six families of priors is assessed. A half-t prior (with several degrees of freedom) on the standard deviation is recommended in preference to a uniform or half-Cauchy prior, when some information exists on the magnitude of variability 'core' to the process or analytical method. Whilst a half-t prior has been previously proposed, through extensive simulation it is demonstrated that it is the prior of choice for estimating uncertainty of variance estimators in routine analysis of analytical and process variation studies. The coverage of 95% credible intervals for variance components and total variance is 93% (approximately) or above across a range of realistic scenarios. Other priors investigated, including Jeffreys', a FLAT prior and inverse gamma distributions on stratum variances available in PROC MIXED[1] in the SAS/STAT® software, are less satisfactory. This evaluation is novel: for one-way variance component designs there is very limited evaluation of the half-t prior when estimating the uncertainty of the variance component estimators; for the two-way or more complex none has been found. Since the coverage issues were primarily for the mid-level variance component, evaluation of designs more complex than one-way is important. Highest posterior density intervals are recommended with the metric of the parameter being important. Additionally, a scale based on the intra-class correlation coefficient is proposed for plotting the credible intervals.

[1]SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Marion J Chatfield, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

UNCERTAINTY OF VARIANCE ESTIMATORS IN ANALYTICAL AND PROCESS VARIABILITY STUDIES

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signed: .....

Date: .....

Acknowledgements

I am very grateful for the advice, support and understanding of many people during the 8 years I have been studying for the Ph.D. In particular, I would like to express my sincere appreciation and gratitude to my supervisor Professor Steve Gilmour for his generosity with advice, support and his impressive statistical knowledge, as well as his patience and ever present encouragement during the lengthy period of study. He has contributed substantially to my education. I'd also like to thank the staff at the University of Southampton including Antony Overstall, Dave Woods, Robin Mitra and Kulvir Bouri for always being so helpful, despite my rather invisible presence at university due to the limitations of studying part-time whilst working. The research has been made possible by the financial support for the fees by GSK (GlaxoSmithKline), and by the support, encouragement and advice from my managers during this period, Mike Denham and Gillian Amphlett, to whom I am very grateful. The work has also benefited greatly from encouragement from other GSK colleagues and, in particular, I would thank Phil Borman, Simon Bate, Sonya and Peter Godbert, John Peterson and Nicky Best for their support and encouragement, and Sandra Harling for her help in producing photos of the pharmaceutical process. I am also very thankful to my son Daniel for enabling LaTeXon my laptop and providing advice on other IT issues. Finally, my biggest thanks are to Ken, my husband, for his love and support, for taking over my share of the household management (not a trivial task with four children), and without whom, this would not have been possible.

To my family
My wonderful husband Ken, my children Daniel, Kirsty, Rowan and Alex, and in memory of my parents

Chapter 1

Introduction to Variation Studies

"Understanding variation is the key to success in quality and business."
- W. Edwards Deming

Contents

1.1 Aims and Focus of the Research

In industry, scientific studies are performed that estimate variances, including the components from a number of sources of variation. In addition to estimating the total variance to be expected in the process and/or measurement system, these studies break down the variation into a number of sources so that the magnitudes, or relative magnitudes, of the components of variance can be quantified.

The utility of this quantification is

- to allow focus of efforts to reduce the largest sources of variation.
- a step in estimating other quantities of interest.
- to aid the design of future studies or data collection to effectively estimate quantities of interest - either when they contribute to the quantity of interest e.g. the total variation or when they are nuisance sources of variation to the estimation.

For example, in the pharmaceutical industry it is required to measure the chemical and physical composition of materials. The precision of the measurement methodology i.e. the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling and measurement of the material, is very important to assuring the quality of the materials and ultimately the medicine which the patient receives. Experimental studies are performed to understand sources of variation and estimate the precision of the measurement methodology. Further details are provided in Section 1.2.

As well as requiring statistical methods to provide reasonable point estimates of the components of variance, it is also valuable to understand the uncertainty around those estimates. This is particularly important given that studies are often small compared to the size likely to be required to produce adequately precise estimates of the variance components. In industry understanding sources of variation in a process or measurement system is useful in assessing, improving and quantifying analytical or process capability e.g. through measurement systems analysis studies such as gage reproducibility and repeatability (gage R&R). If the uncertainty of the variance estimators is unknown, efforts to improve analytical or process variation may be misdirected or the variance considered acceptable or unacceptable when it may not be.

The aim of this research project is to provide a statistical approach which can be used routinely in industry to quantify the uncertainty of variance estimators in analytical and process variability studies.

In this chapter an introduction to variation studies in the pharmaceutical industry, the application area which motivated this research project, is provided in Section 1.2. The research project will draw upon my experience working for the science-led global health-care company GSK (GlaxoSmithKline), one of the top ten pharmaceutical companies in the world, and which has funded the fees for this research. The general statistical model applied to variation studies and an example model for an experimental study examining measurements from a chemical analytical method used to assess the chemical quality of materials is introduced in Section 1.3. The chemical analytical method example is used to illustrate aspects that will need to be taken into account when solutions are proposed to meet the aim of this project. For that example considerations related to the application area are discussed in Section 1.4 and statistical issues encountered with the statistical analysis of such studies are given in Section 1.5.

Section 1.6 provides an overview of the research performed and the structure of the thesis. To further understand the occurrence of some of the issues listed in Section 1.5, simulations using a frequentist analysis are performed in Chapter 2. An overview of the literature concerning this topic is given in Chapter 3. It will be seen that a frequentist analysis is unlikely to provide an adequate solution and that a Bayesian approach is proposed. Chapters 4 to 8 of the thesis are concerned with developing a Bayesian approach which can be recommended to be used routinely for these types of studies and

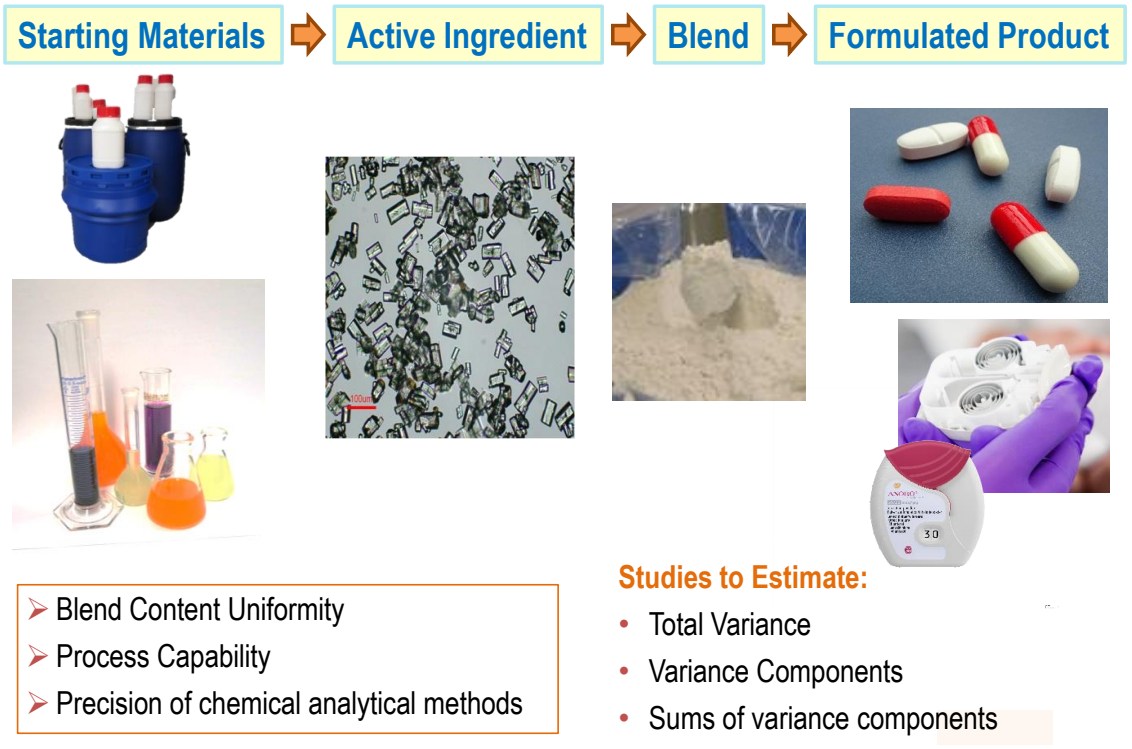
the advocated approach is applied to some example applications in Chapter 9. Final conclusions including a summary of the original contributions made and proposals for future work are given in Chapter 10.

1.2 Introduction to Variation Studies in the Pharmaceutical Industry

In the pharmaceutical industry the formulated medicine which the patient receives is typically the end result of a complicated process. This manufacturing process often starts with commercially available chemicals ("starting materials") and via a series of chemical reactions produces the active ingredient which has the desired efficacy. It is important that the active ingredient is of the right chemical quality and has appropriate physical attributes e.g. particle size. The active ingredient is then combined (blended) with other ingredients to produce the formulated medicine which is in a form which the patient can receive and will remain stable over time e.g. tablets or output from a medical device such as an inhaler. Again the appropriate chemical and physical composition of the drug product is important to achieve a safe and efficacious medicine for the patient. Variability in the inputs and manufacturing conditions (though tightly controlled), results in some variability in the attributes of the materials during the various stages of the manufacturing process. Thus there is a need to measure and quantify many attributes during the manufacturing process to gain process understanding, to ensure appropriate quality and demonstrate that the medicine meets specifications. Some stages of the process are illustrated in Figure 1.1.

Thus in the pharmaceutical industry many situations exist where there is a need to quantify product and analytical method variability and identify sources contributing to that variability. For example: blends and batches are sampled to ensure the patient will receive the same amount of active ingredient ("content uniformity"); the ability of an analytical method to measure accurately and with good precision is quantified and improved where possible; it is required to demonstrate that when changes are made to processes or analytical methods the product or measurements are equivalent. Process variation studies assess the contribution and magnitude of various sources of variation to the attributes of the product (or upstream materials), whilst analytical variation studies assess the variation of the analytical method (measurement system) itself. An example of a process variation study is given by Hofer and Rauk (2017) where Bayesian methodology is used to evaluate content uniformity capability. The variation due to batch, location within a batch and between dosage units at a location are assessed. A similar study is described in Chapter 4 and will be used as a basis for the simulations performed for this thesis.

FIGURE 1.1: **Illustration of pharmaceutical process**



Acknowledgement: Thanks to Sandra Harling, GSK, for assistance with photography and to GSK for use of the inhaler photo

Understanding analytical variation is extremely important as there is often non-negligible variation associated with the reported measurement. The precision of the measurement methodology is very important to assuring the quality of the materials (the "control strategy") and ultimately the medicine which the patient receives. This is not only important to the pharmaceutical company but is also subject to scrutiny by the regulatory authorities. Some definitions used to describe different levels of precision according to their source of variation, and the regulatory and quality motivation for performing analytical variation studies are now given.

The precision of an analytical method is the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under specified conditions (ICH, 1994).

Three types of precision are recognised for an analytical method and defined by ICH (1994) as:

1. **Repeatability** expresses the precision under the same operating conditions over a short interval of time.
2. **Intermediate precision** expresses within-laboratories variations: different days, different analysts, different equipment, etc.

3. **Reproducibility** expresses the precision between laboratories (collaborative studies, usually applied to standardization of methodology).

Experimental work is required to meet method intermediate precision validation requirements for regulatory approval in the pharmaceutical industry. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH, 1994) state the following:

> "The applicant should establish the effects of random events on the precision of the analytical procedure. Typical variations to be studied include days, analysts, equipment, etc. It is not considered necessary to study these effects individually. The use of an experimental design (matrix) is encouraged. The standard deviation, relative standard deviation (coefficient of variation) and confidence interval should be reported for each type of precision investigated."

Note that the term precision is not taken as having precise definition/formula, the guideline says it is usually expressed as the variance, standard deviation or coefficient of variation, of a series of measurements.

Thus experimental studies are performed to estimate the precision of the measurement methodology (often called "precision studies"). Intermediate precision studies can be quite minimal e.g. Walfish (2006) suggests a minimum of two analysts on two different days. For Japanese regulatory submissions 5 degrees of freedom (for factors at the intermediate precision level) are typically required (Nijhuis and Van den Heuvel, 2007), which can be obtained from one laboratory or across two, e.g. a study of two analysts, two instruments and two columns over six days meets this requirement. Usually the individual effects of analyst, day etc. are ignored when estimating intermediate precision. However, as the analytical methods are often quite complex in themselves, understanding the sources of analytical variation beyond the intermediate precision and repeatability levels required by ICH is frequently very important. The application of "quality by design" to analytical methods (Borman et al., 2007) resulted in the desire to incorporate risk assessment into the evaluation and to gain some method understanding. Thus variation ("ruggedness") studies are performed as an extension to the intermediate precision exercise with the aim of (within the limited resource):

- identifying special cause variation which should then be eliminated,
- assessing common cause variation including estimation of variance components due to various sources,
- assessing whether the precision of the method is likely to meet requirements.

The term "ruggedness" is used to show the intent of the study to assess whether the method is rugged (generally the term rugged is defined as strongly made and capable of withstanding rough handling), in contrast to a "precision" study where the focus is on estimating the precision or variance. In practice, similar designs may be used for each

and ruggedness studies also have the intent of estimating the variance. Further details of the use of ruggedness studies within GSK are given by Borman et al. (2011).

This section now gives further details to provide an understanding of this application area and the issues involved. One common type of chemical analytical method is briefly described - one that uses a High Performance Liquid Chromatographic (HPLC) system. The designs of a variation (precision) study used to understand the sources of variation appropriate to an HPLC method are discussed, with two typical examples being provided.

A very common chemical analytical method in the pharmaceutical industry is one which uses an HPLC system. Many people are familiar with a school experiment where the colours of ink are separated by applying water to an ink blot placed on a piece of blotting paper. In essence, the HPLC system is a more sophisticated version of this. A "column" packed with material replaces the blotting paper. Instead of an ink blot, a small volume of a prepared solution of the sample ("preparation") is injected ("injection") on to the column. Then a liquid (called mobile phase) is forced through the column by a high pressure pump causing the sample components to flow through the column. As the sample components exit the column their amount is measured by a detector. The chemical components of the sample flow through the column at different rates and thus exit at different times enabling them to be distinguished from each other. The equipment (HPLC instrument, column etc.) will be set up to analyse a number of injections in sequence, this being called an "analytical run".

Borman et al. (2011) describe analytical method ruggedness studies which provide an experimental evaluation of noise factors such as analyst, instrument or stationary phase batch (column) and identify whether they significantly affect the variability of the results produced by the analytical method. The design of a simple example analytical ruggedness design (Example 1) where only one batch of material has been analysed is given in Figure 1.2. For each Site/Analyst/Instrument/Column combination two measurements are made.

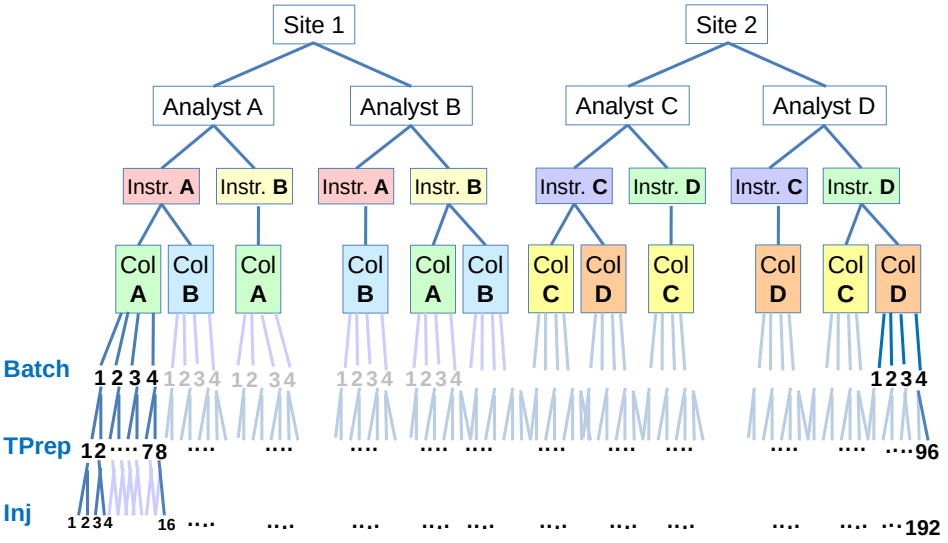
FIGURE 1.2: **Example design of analytical ruggedness study - Example 1**

	Site 1						Site 2					
Analyst	A			B			C			D		
Instrument	A		B	A	B		C		D	C	D	
Column	A	B	A	B	A	B	C	D	C	D	C	D

Some studies may have a large number of random sources of variation being considered. For example the design shown in Figure 1.2 only has two measurements per column/instrument/analyst combination. However, with more measurements there might be multiple days/analytical runs, multiple sample preparations and multiple injections for

each combination allowing further sources of variation to be assessed. A larger example of a ruggedness design (denoted Example 2) is shown in Figure 1.3. A total of 192 measurements are made (injections) under various operations of the analytical method on 4 batches of material as illustrated. For example, by tracing the path back up the tree, it can be seen that the measurement from injection 4 has come from test preparation 2, was made on batch 1, using equipment column A and instrument HPLC A, by analyst A on site 1. For a combination of site, analyst, HPLC instrument and column (Col), two test preparations (TPrep) are made from each of the 4 batches of material and injected twice (Inj). In an ideal world every injection on a particular batch would give the same measured result i.e. the analytical method would be rugged to the effect of factors used in its operation. The aim of the study is to identify which factors such as Site, Analyst HPLC instrument etc. cause the measured results to vary, and then try to reduce those sources of variation. An estimate of how much measurements will vary during routine operation of the method is also required.

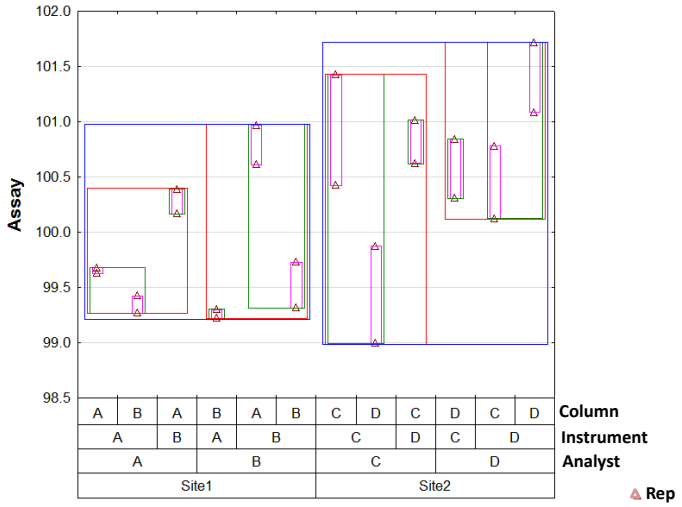
FIGURE 1.3: **Typical ruggedness design - Example 2**



Once the study has been conducted the results will be visualised and a statistical analysis performed. Illustratory results for Example 1 are plotted in Figure 1.4 (see Appendix A, Table A.1 for the data). The replicate measurements of assay (a measure of the amount of active pharmaceutical ingredient) are plotted against the various factors being studied, together with boxes surrounding results sharing the same factor level to aid the eye in identifying the larger sources of variation (a variability plot).

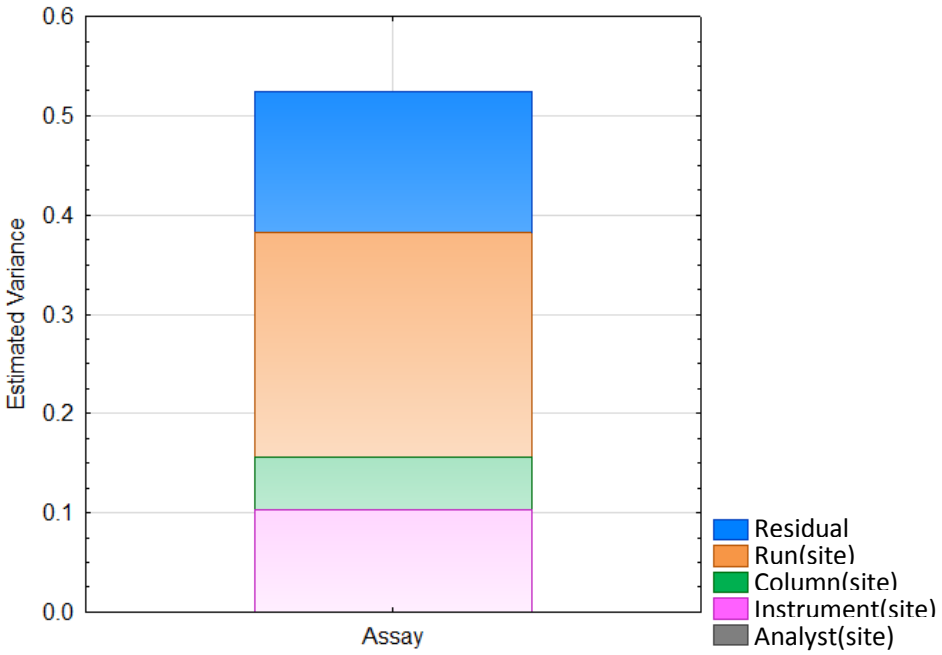
The statistical analysis will attempt to estimate the variance components due to the various sources. These are plotted for Example 1 in Figure 1.5. The largest variance component identified is the analytical run (identified as Run(site)) which can then be evaluated for reasons why the variation is seen. Statistically this is confounded with the

FIGURE 1.4: **Plotting results for a typical ruggedness study - Example 1**



Analyst*Instrument*Column interaction but from scientific knowledge this will almost certainly mostly be variation due to the analytical run. Note Analyst(site) cannot be seen in the plot since it was estimated as zero.

FIGURE 1.5: **Plotting variance components for a typical ruggedness study - Example 1**



Estimated variance components are used to identify larger sources of variation to try to reduce the variability of the method from those sources. However, without knowledge of

the uncertainty around the variance component estimates, efforts may be put into investigating and trying to reduce particular sources of variation when actually the observed larger variance component estimates are just due to chance. The variance components are also used to estimate the long term variability of the method. Again without knowledge of the uncertainty associated with that estimate, the method could be thought to be meeting requirements, or not meeting requirements, due to chance variation in the estimator. Another use of the information gained on the variance components could be in the subsequent design of an equivalence study to assess a change in analytical method - see Borman et al. (2009), example 3.

1.3 Statistical Model for Variation Studies

Analysis of data from a study aiming to estimate variance components will generally be performed using the linear mixed model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{Y}$ is the vector of measurements, $\boldsymbol{\beta}$ represents fixed effects, $\mathbf{X}$ is the known design matrix for the fixed effects, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, represents higher level random effects, $\mathbf{Z}$ is the known design matrix for the random effects and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$ represents the bottom level (residual) variance. The marginal distribution of $\mathbf{Y}$ can be denoted $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ where $\mathbf{V} = Var[\mathbf{Y}] = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$. A more general model has this form with the only assumption being that $\mathbf{V}$ is non-negative definite (referred to as unrestricted model). For our model where the random effects are variance components the assumption is also that $\mathbf{G}$ is non-negative definite (referred to as restricted model). $\mathbf{Z}$ has columns for each level of the random effects and for nested random effects $\mathbf{Z}$ is block diagonal (with appropriate ordering) and consequently $\mathbf{Z}\mathbf{G}\mathbf{Z}'$ is block diagonal. For crossed random effects $\mathbf{Z}$ is not block diagonal but is sparse (e.g. if no random interactions are fitted, each row will have the number of 1's equal to the number of random factors). More information, including the structure of $\mathbf{Z}\mathbf{G}\mathbf{Z}'$, is given in Rasbash and Goldstein (1994) and Bates et al. (2015). Note in subsequent chapters reference will be made to one-way and two-way variance components models. In a one-way model there is one higher level random effect in addition to the bottom (residual) variance, whilst in the two-way model there are two higher level nested random effects.

As an illustration the design in Example 2 is considered - the model has 9 random effects (variance components). The model also has one fixed effect, batch, which is of no interest in itself in this study. The random terms are a mixture of crossed and nested effects of which all have physical meaning (the interaction between Analyst, HPLC Instrument and Column being confounded with an analytical run) except the interactions with Batch. The design is not balanced with respect to Analyst, HPLC

Instrument or Column (e.g. analyst A does not use column B on HPLC Instrument B) to save resource in implementing the design. Note the unbalanced arrangement of these 3 factors within one site is also that which GSK may use to meet the Japanese intermediate precision requirement of 5 degrees of freedom. The data actually collected also has an additional complication in that 2 observations are missing due to the analyst making a mistake in preparing a sample.

The model for the measurement $y_{jkmqbrv}$ made at site j (effect s_j), by analyst k at site j (effect a_{jk}), with instrument m at site j (effect i_{jm}), using column q at site j (effect c_{jq}), from batch b (effect β_b), on test preparation r at site j by analyst k using instrument m and column q from batch b (effect t_{jkmqbr}), and injection v from test preparation r (effect $\epsilon_{jkmqbrv}$) is:

$$y_{jkmqbrv} = \mu+\beta_b+s_j+(s\beta)_{jb}+a_{jk}+i_{jm}+c_{jq}+(aic)_{jkmq}+(aic\beta)_{jkmqb}+t_{jkmqbr}+\epsilon_{jkmqbrv}, \tag{1.2}$$

where μ is the overall mean; β_b is the effect of batch b, for $b = 1, ...4$; $s_j \sim N(0, \sigma_s^2)$ for $j = 1, 2$; $a_{jk} \sim N(0, \sigma_a^2)$ for $k = 1, 2$; $i_{jm} \sim N(0, \sigma_i^2)$ for $m = 1, 2$; $c_{jq} \sim N(0, \sigma_c^2)$ for $q = 1, 2$; $(s\beta)_{jb} \sim N(0, \sigma_{s\beta}^2)$; $(aic)_{jkmq} \sim N(0, \sigma_{aic}^2)$; $(aic\beta)_{jkmqb} \sim N(0, \sigma_{aic\beta}^2)$; $t_{jkmqbr} \sim N(0, \sigma_t^2)$ for $r = 1, 2$; $\epsilon_{jkmqbrv} \sim N(0, \sigma_e^2)$ for $v = 1, 2$, and all random variables are assumed to be independent.

1.4 Application Considerations

In GSK, studies such as those described in Examples 1 and 2 have been successful in identifying special cause variation due to the factors. This has been done pragmatically by using statistics to summarise the data to identify large sources of variation and then visualization and scientific knowledge to identify the cause. Point estimates of variance components and various measures of precision are calculated. Several areas for investigation are described below:

- Confidence intervals (CIs) are not usually provided for parameters related to precision within GSK. Their calculation is very desirable to evaluate the results of the study and to compare the method precision in relation to manufacturing control requirements. Though the ICH (1994) guideline refers to confidence intervals it is not clear whether this refers to the precision related parameter (for example, a confidence interval for σ_r^2), or more likely, to the precision of a single result reported by the method when in use (e.g. the confidence interval $\pm 2\hat{\sigma_y}$). (Note we are interested in σ or σ^2 as they relate to precision which as noted in Section 1.2 is a general concept rather than having a precise definition.) Ermer (2005) states "the author is not aware of any publication on pharmaceutical validation which reports it" referring to the confidence interval on the precision parameter). Though

in Chapter 3 it will be seen that a United States Pharmacopoeial (USP) stimuli article is promoting the use of tolerance intervals.

- Though not required by regulatory authorities, the examination of individual factors contributing to intermediate precision has been practically useful in assessing and improving the ruggedness of analytical methods within GSK. It is appreciated that the number of levels of each factor is extremely small. The ability to evaluate designs and in particular the effect of increasing the levels (where practical) would be useful. If sample sizes are increased what is the accompanying improvement in estimation of variation sources, or power to detect special cause variation? Providing estimates of the uncertainty of the variance estimators enables users of the statistics to make their own judgement in their subsequent decision-making. If the uncertainty is found to be typically too great, this will be a valuable first step in promoting designs which enable better decisions.
- In the design and evaluation of studies, prior analytical knowledge is invaluable. Is it possible to incorporate this into the statistical evaluation?

1.5 Statistical Issues

Considering Example 2 and with some previous knowledge of the application area the following statistical issues were identified:

1. There are only 4 levels of the random effects Analyst, Instrument and Column in total, 2 levels nested within each site - so there are few degrees of freedom for estimating variance components.
2. The design is unbalanced with respect to Analyst, Instrument and Column.
3. There are many variance components to be estimated.
4. The true values for some of the variance components may well be practically zero.
5. If a variance component is small the REML estimate may be zero and some commonly used methods for confidence intervals cannot be used.
6. There are often problems fitting the mixed model e.g. often it cannot be fitted using mean squares (infinite likelihood) or REML with no boundary restrictions on the variance components - the unrestricted model in Section 1.3 (as it doesn't converge). This will be explored further in Chapter 2.
7. For unbalanced data there is not a clear recommendation on either the estimation method or how to construct confidence intervals.
8. An estimate of the long term variation of interest will be required i.e. a sum of the variance components.

9. It may be required to test a fixed effect of site, in which case issues with estimating variance components close to zero could present problems.
10. The estimated variance components and data visualisation have been used to identify special cause variation. Would there be any benefit to treating them as fixed effects initially in the statistical analysis?

Some of these issues are explored further in Chapter 2.

1.6 Outline of Thesis

The aim of this research project is to provide a statistical approach which can be used routinely in industry to quantify the uncertainty of variance estimators in analytical and process variability studies. A large number of aspects/issues which it is desirable to address were identified in Sections 1.4 and 1.5 for which, as will be discussed in Chapter 3, there is not an adequate solution in the current literature. An outline of the thesis is given below. Discussion of which of the above aspects and issues have been addressed during this research, the original contributions made and what further work is proposed is given in Chapter 10.

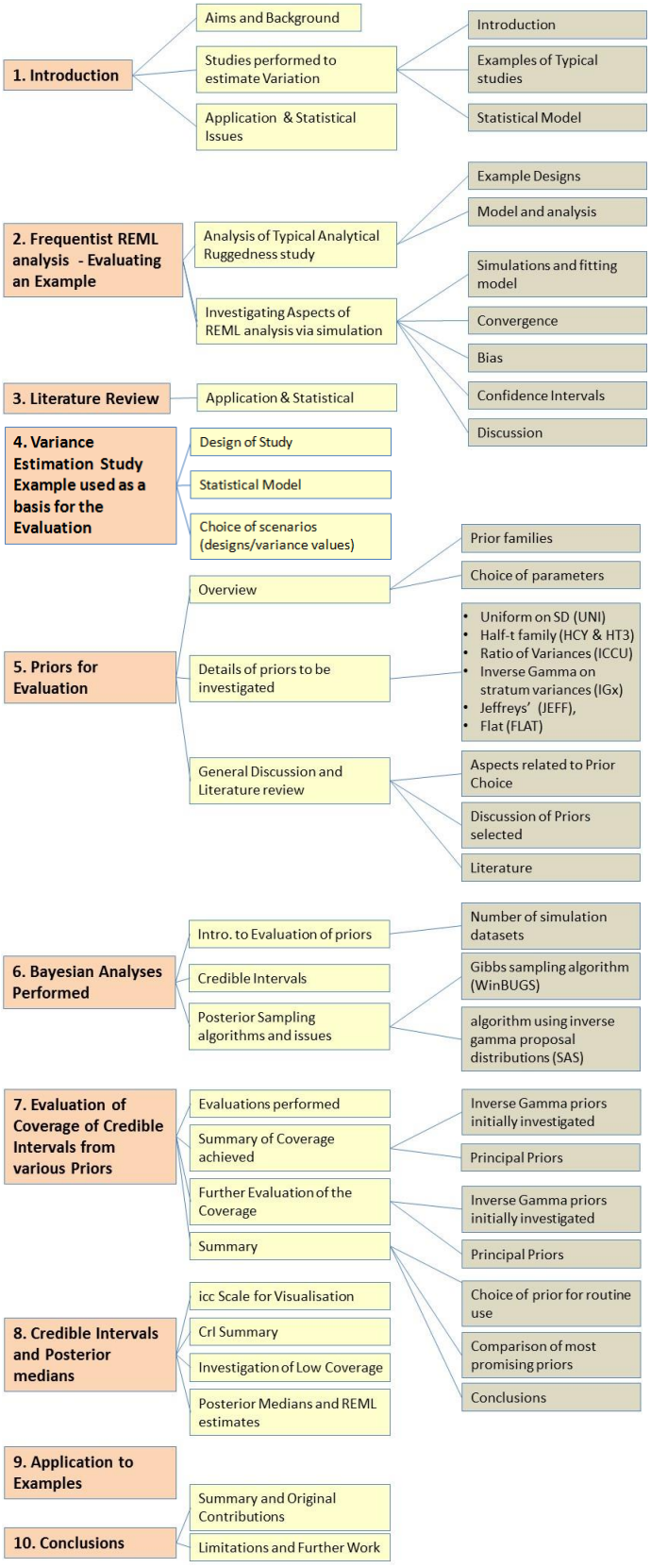
A roadmap of the work performed for this thesis is shown in Figure 1.6.

Chapter 2 describes a frequentist REML (Restricted or Residual Maximum Likelihood) analysis which is/was the type of statistical analysis used within GSK to analyse this type of study. Simulations are also performed to assess how often aspects such as negative estimates of variance components are likely to occur. Chapter 3 provides a review of the literature related to variation studies and their analysis. It describes various frequentist statistical approaches in the literature for estimating variance components and the uncertainty of those estimates and provides references where they have been used for process or analytical variation studies. However, the frequentist approaches are unlikely to be suitable for the very small complex designs typically used for analytical and process variation studies. Literature discussing the advantages of the Bayesian approach and the conclusions drawn from some relevant simulation studies are summarised. A Bayesian approach was decided upon for this research project. The main part of the work performed for this thesis investigating a Bayesian approach to evaluating the uncertainty of variance component estimators is then described in Chapters 4 - 7. The Bayesian approach is investigated through simulating datasets and performing a Bayesian analysis for a number of designs/scenarios which are described in Chapter 4. In order to perform a Bayesian analysis a prior is required. Priors to be evaluated for the estimation of the uncertainty of variance component estimators are discussed in Section 5. Criteria for their choice are developed and discussed, especially in the context of their use in variation studies. Some priors were selected to be vague whilst others were

intended to be mildly informative. The conduct of the Bayesian analyses is described in Chapter 6. Sampling from the posterior is performed using Gibbs sampling in WinBUGS or by sampling using inverse gamma proposal distributions in PROC MIXED in SAS/STAT® software (SAS Institute Inc.: SAS/STAT, 2011)[1] dependent on the prior. Aspects related to sampling from the posterior are described in Section 6.3 for the analyses performed in WinBUGS. These include the assessment of, and ensuring adequate, convergence. Section 6.4 describes the posterior sampling performed in PROC MIXED and associated issues. The algorithm in SAS used to sample from the posterior stops when the acceptance rate is too low. A novel approach to overcome this issue in PROC MIXED is described. The frequentist concept of coverage is then used to evaluate the suitability of the priors and to choose the one(s) to be recommended for routine use in estimating variance components. These results are presented in Chapter 7 where Section 7.2 provides a summary of the results and Section 7.3 provides further details including examining the reasons for the poor coverage of some of the priors. Then an evaluation of the credible intervals (CrIs) for the variance components for the scenarios is provided in Chapter 8. Additionally this chapter contains results of further investigations into possible causes of low coverage. The half-t prior with three degrees of freedom is recommended for routine use to quantify the uncertainty of variance estimators in analytical and process variability studies. Whilst this prior was proposed by Gelman (2006), this research provides an extensive simulation study to support its routine use, identifies that it should be used in situations where Gelman thought a uniform prior would be satisfactory and makes proposals for the scale factor in the context of information on the expected variance related to a factor intrinsic to the scientific situation. In addition, the simulation study provides evidence for not using other priors for these types of study. A Bayesian analysis using the half-t prior is applied to some examples in Chapter 9. Finally conclusions of the research, a summary of original contributions and proposals for future work are given in Chapter 10.

[1]SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

FIGURE 1.6: **Roadmap of thesis**



Chapter 2

Frequentist REML Analysis - an Example

Contents

As mentioned in Section 1.5 I had encountered issues with performing a frequentist REML analysis using PROC MIXED in SAS. When the variance components were not bounded at zero the mixed model often could not be fitted using mean squares or REML. The REML estimate for a variance component could be zero and, in that case, some commonly used methods for confidence intervals cannot be used. This chapter looks at applying a frequentist REML analysis to data from typical analytical ruggedness/ method precision studies and explores these issues further. Small balanced and unbalanced designs with four random effects are investigated. Firstly in Section 2.1 the analysis of four example datasets is discussed. Then in Section 2.2 simulations are performed to gain an understanding of how often difficulties in the REML analysis arise and whether reasons can be identified.

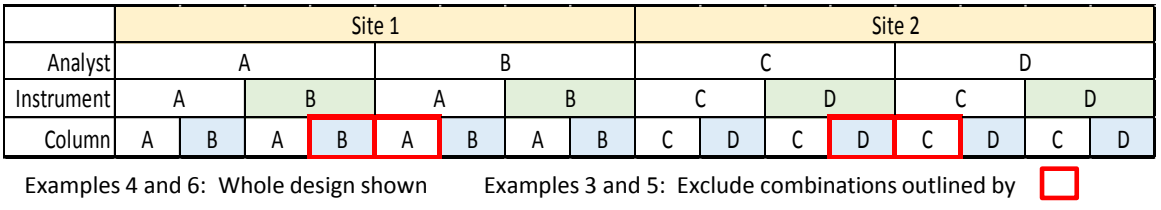
2.1 Analysis of Example Results from a Typical Analytical Ruggedness Study

Typically when data for a design such as that shown in Figure 1.3 is analysed using a bounded mixed model, at least one variance component is estimated as zero. Stroup and Littell (2002) identified that bounding the variance estimates by zero had an impact on the estimates and ultimately any tests of fixed effects. Note the results reported in Section 2.2.3 support this, in that they show a bias in the variance components when their estimation is bounded at zero. Their recommendation (Littell et al., 2006, P150) was to allow estimates to be negative (NOBOUND option using the procedure PROC MIXED in SAS for REML or type 3 option if fitting by the method of moments approach). I had noticed problems with convergence or the Hessian (2nd derivatives) matrix being reported as non-positive definite (and thus no CIs being provided) when NOBOUND was applied in an analysis in PROC MIXED, though these did not occur when the estimates were bounded by zero. Some examples are now analysed.

2.1.1 Analysis of Four Examples

Four example datasets with designs based on that for Example 1 (see Section 1.2) are analysed. The designs are shown in Figure 2.1. Examples 3 and 5 have the same design as Example 1 except they had only one replicate rather than two for each combination of factors. The design is unbalanced with respect to combinations of analyst, instrument and column. A balanced version of the design is also investigated in Examples 4 and 6, where all 8 combinations of analyst, instrument and column are present for each site. Example 5 has the same design as Example 3, and Example 6 has the same design as Example 4 but with alternative data. The data for Example 6 illustrates an issue which occurred with the original GSK data (not included for confidentiality reasons). The data for the examples are given in Appendix A, Table A.2.

FIGURE 2.1: **Designs for Examples 3 to 6**



2.1.2 The Model and REML Analysis Performed

For the four examples the model for the measurement y_{jkmqr} made at site j (effect s_j), by analyst k at site j (effect a_{jk}), with instrument m at site j (effect i_{jm}), using column q at site j (effect c_{jq}), for replicate r (effect ϵ_{jkmqr}) is:

$$y_{jkmqr} = \mu + s_j + a_{jk} + i_{jm} + c_{jq} + \epsilon_{jkmqr}, \tag{2.1}$$

where μ is the overall mean; $s_j \sim N(0, \sigma_s^2)$ for $j = 1, 2$; $a_{jk} \sim N(0, \sigma_a^2)$ for $k = 1, 2$; $i_{jm} \sim N(0, \sigma_i^2)$ for $m = 1, 2$; $c_{jq} \sim N(0, \sigma_c^2)$ for $q = 1, 2$; $\epsilon_{jkmqr} \sim N(0, \sigma_e^2)$ for $r = 1$, and all random variables are assumed to be independent. Note that for the unbalanced design not all combinations of $kmqr$ exist.

The model was fitted using the procedure PROC MIXED in SAS using REML estimation of variance components. Two analyses were performed. In the first analysis (denoted bounded) $\hat{\mathbf{G}}$ is constrained to be within the parameter space $\mathbf{G}$ of the restricted linear mixed model i.e. $\hat{\mathbf{G}}$ is non-negative definite and the variance component estimates are bounded by zero. In the second analysis this constraint is not applied (denoted unbounded) and $\hat{\mathbf{G}}$ is not required to be non-negative definite. This gives an extended REML estimate of $\mathbf{G}$ for the restricted model. This is numerically the same as the REML estimate for the unrestricted model. However, conceptually we consider it different. Though these extended estimates are outside the parameter space (and thus not appropriate estimates in themselves), it is desired to evaluate whether they provide better estimates of linear combinations of variance components or for evaluation of fixed effects. The unbounded analysis is performed in SAS PROC MIXED using the NOBOUND option. PROC MIXED offers the opportunity to provide initial values for the variance components. If the analysis failed to converge using default values in PROC MIXED, initial values were set: 5 for the residual variance and 1 for the other four variance components.

2.1.3 Issues with Fitting the Model using REML

PROC MIXED monitors the convergence of the analysis and performs iterations to satisfy default or specified criterion subject to a maximum number of iterations. It reports the convergence status of the analysis. These are shown in Table 2.1 for the four datasets, and by whether the analysis was bounded or not, and if not, whether initial values were required to reach convergence. The default criterion is based on the Hessian matrix ($\mathbf{H}$).

It is seen that the analyses for the balanced datasets converged whether or not a bounded or unbounded analysis was applied. However, for the unbalanced datasets convergence was reached for the bounded analysis but not necessarily for the unbounded analyses.

TABLE 2.1: **Convergence status of the REML analyses**

Analysis	**Balanced Dataset**		**Unbalanced Dataset**	
	3	5	4	6
Bounded	Converged	Converged	Converged	Converged
Unbounded - no initial values	Converged	Converged	Stopped with Infinite likelihood	Stopped with Infinite likelihood
Unbounded - initial values	Converged	Converged	Converged	Converged, Hessian not positive definite

When an unbounded analysis was performed using the default initial values SAS provided, the likelihood was found to be infinite and thus the analysis was stopped. By providing initial values such as setting those for the higher level variance components to 1 this problem was overcome. However, though the analyses were then reported as converged, for dataset 6 there was an additional message that the Hessian was not positive definite. Other options to perform the analysis were also explored. If the criterion was based on the change in the function itself or site was fitted as a fixed effect, the analysis with the default initial values was reported as converged but there were still issues (it took over 2000 iterations and the Hessian was still reported as not positive definite). Note problems were also found when using the method of moments approach in SAS. Estimates were provided but it was reported that an infinite likelihood was found and, on occasion, that a linear combination of covariance parameters was confounded with the residual variance.

The estimated variance components (Est.) and the confidence intervals (Lower, Upper) obtained for the analyses are shown in Table 2.2. Some additional statistics output by SAS are also provided in the table.

The method by which SAS produces confidence intervals (CIs) depends on whether a bounded or unbounded analysis is performed.

For a bounded analysis, the chi-squared distribution is used to produce an approximate $(1-\alpha)100\%$ CI about σ_e^2:

$$\nu\hat{\sigma}_e^2/\chi^2_{\alpha/2,\nu} < \sigma_e^2 < \nu\hat{\sigma}_e^2/\chi^2_{1-\alpha/2,\nu}, \tag{2.2}$$

where $\chi^2_{\alpha/2,\nu}$ is the $\alpha/2$ percentile of the χ^2 distribution with ν degrees of freedom, and $\nu = 2\hat{\sigma}_e^2/\text{se}(\hat{\sigma}_e^2)$.

For an unbounded analysis an asymptotic normal approximation is used to produce a CI of

$$\hat{\sigma_e}^2 \pm z_{1-\alpha/2}\text{se}(\hat{\sigma}_e^2). \tag{2.3}$$

TABLE 2.2: **Results for the REML analyses which converged**

Dataset			Bounded						Unbounded					
		Variance	Est.	StdErr	ZValue	ProbZ	Lower	Upper	Est.	StdErr	ZValue	ProbZ	Lower	Upper
Balanced	**3**	Site	22.41	*32.61*	*0.69*	*0.246*	4.34	33864.00	22.95	*32.62*	*0.70*	*0.482*	-41.00	86.89
		Analyst(site)	0.00	_	_	_	_	_	-0.88	*0.88*	*-1.00*	*0.320*	-2.60	0.85
		Instr(site)	0.00	_	_	_	_	_	-0.21	*1.39*	*-0.15*	*0.881*	-2.93	2.52
		Column(site)	0.07	*1.39*	*0.05*	*0.479*	_	_	-0.11	*1.48*	*-0.07*	*0.941*	-3.00	2.78
		Residual	4.90	*2.00*	*2.45*	*0.007*	2.52	13.36	5.63	*2.81*	*2.00*	*0.023*	2.57	20.65
	5	Site	0.00	_	_	_	_	_	-2.76	*4.62*	*-0.60*	*0.551*	-11.82	6.30
		Analyst(site)	3.46	*4.82*	*0.72*	*0.237*	0.70	2914.49	5.68	*8.50*	*0.67*	*0.504*	-10.98	22.34
		Instr(site)	0.00	_	_	_	_	_	-2.20	*1.45*	*-1.52*	*0.128*	-5.04	0.63
		Column(site)	0.00	_	_	_	_	_	-0.09	*2.95*	*-0.03*	*0.976*	-5.87	5.69
		Residual	9.32	*3.80*	*2.45*	*0.007*	4.79	25.38	10.84	5.42	2.00	0.023	4.95	39.80
Unbalanced	4	Site	17.20	*26.13*	*0.66*	*0.255*	3.21	49327.00	17.55	26.16	*0.67*	*0.502*	*-33.72*	68.82
		Analyst(site)	1.82	*2.55*	*0.71*	*0.238*	0.36	1634.72	0.53	0.55	*0.96*	*0.335*	*-0.55*	1.60
		Instr(site)	0.00	_	_	_	_	_	-1.16	0.62	*-1.88*	*0.060*	*-2.37*	0.05
		Column(site)	0.00	_	_	_	_	_	1.38	3.29	*0.42*	*0.674*	*-5.06*	7.83
		Residual	2.11	*1.06*	*2.00*	*0.023*	0.96	7.75	3.20	1.67	*1.92*	*0.028*	*1.42*	12.66
	6	Site	0.00	_	_	_	_	_	1.33	4.09	*0.33*	*0.744*	*-6.67*	9.34
		Analyst(site)	6.36	*7.38*	*0.86*	*0.194*	1.51	756.47	3.44	1.72	*2.00*	*0.046*	*0.07*	6.81
		Instr(site)	0.00	_	_	_	_	_	-4.17	2.09	*-2.00*	*0.046*	*-8.26*	-0.08
		Column(site)	0.00	_	_	_	_	_	-0.20	0.10	*-2.00*	*0.046*	*-0.40*	0.00
		Residual	7.62	*3.81*	*2.00*	*0.023*	3.48	27.97	11.96	5.98	*2.00*	*0.023*	*5.46*	43.91

For the bounded analysis it is seen that for each dataset a number of the variance components were estimated as zero and no confidence interval was provided for these variance components. For the software to provide no confidence intervals is reasonable numerically as the method used to produce the confidence intervals has a numerator for the degrees of freedom of $(\hat{\sigma}_e^2)^2$ and this is clearly inappropriate when the estimated variance component is zero. However, from a practical perspective, not to have a confidence interval or to assume that it is $[0, \infty)$ is not useful. It also seems inappropriate that if the estimate for a variance component has a (non-small) positive value an upper confidence limit can be obtained. However, when an estimate is small (approaching zero) the upper limit can become larger than that when the estimate is larger, and when the estimate is zero the upper limit is either unavailable or assumed to be infinity. It is also noted that for dataset 3 and the Column(site) variance component, though the variance component estimate is non-zero, confidence limits were not provided. The PROC MIXED functionality does not provide confidence intervals for linear combinations of variance components. Within GSK this has been done by using a Satterthwaite approximation but the methodology cannot be applied incorporating a variance component estimated as zero.

The unbounded analysis did provide confidence intervals for individual variance components for datasets 3, 4 and 5. However, as discussed in Section 2.1.2 it is considered that these extended estimates are outside the parameter space and thus not appropriate estimates in themselves. They may be useful to provide better estimates of linear combinations of variance components or for evaluation of fixed effects. However, the PROC MIXED functionality does not provide confidence intervals for linear combinations of variance components. For dataset 6 confidence intervals were also provided. However, SAS reported that the Hessian was not positive definite and examination of the ZValues (Wald Z-values) in Table 2.2 shows that for all but the site variance component, a value of 2.00 (highlighted in red) is obtained suggesting that there has been a numerical problem with the analysis and that the results should not be relied upon. If the convergence criterion was changed similar issues occurred. Note problems were also found when using the method of moments approach in SAS. Estimates were provided but the confidence intervals were based on the observed inverse Fisher information matrix $(2\mathbf{H}^{-1})$, i.e. based on the likelihood, with sometimes the standard errors being given as 0 or the z-values being all the same. The software reported that an infinite likelihood was found and, on occasion, that a linear combination of covariance parameters was confounded with the residual variance. Though there are no clear recommendations for unbalanced data, given that the literature tends to favour using REML, and this issue occurred for the real GSK dataset on which the example is based, the issues for this analysis were investigated further.

The linear mixed model was given in Equation (1.1). $\mathbf{V} = Var[\mathbf{Y}] = \mathbf{ZGZ}' + \mathbf{R}$ depends on the unknown vector $\boldsymbol{\phi}$ consisting of the higher level variance components such as those

in the model defined by Equation (1.2) or the model defined in Equation (2.1) and the bottom (residual) variance σ_r^2. REML estimates the random effects by minimising the objective function:

$$-2l_R(\boldsymbol{\phi};\mathbf{y}) = \log|\mathbf{V}(\boldsymbol{\phi})| + \log|\mathbf{X}'\mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{X}| + \left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})\right)'\mathbf{V}(\boldsymbol{\phi})^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\phi})\right) + c_R, \tag{2.4}$$

where $c_R = (n-p)\log(2\pi)$; and p is the rank of $\mathbf{X}$ (see Appendix B). For a mean model (i.e. no other fixed effects) p=1.

In the unbounded analysis for the restricted model Equation (1.1) the estimate of $\mathbf{G}$, $\hat{\mathbf{G}}$, is not constrained to be non-negative definite. Whilst the true variance components must be ≥ 0, in this formulation of the loglikelihood Equation (2.4) and with an unconstrained optimisation, the estimates can be negative (and as mentioned above may be preferable when estimating linear combinations of variance components). The estimate of $\mathbf{V}$, $\hat{\mathbf{V}}$ is non-negative definite in both unbounded and bounded analyses.

An infinite log likelihood could correspond to a solution where $|\mathbf{V}(\boldsymbol{\phi})| = 0$ since as $|\mathbf{V}(\boldsymbol{\phi})| \to 0$, $\log|\mathbf{V}(\boldsymbol{\phi})| \to -\infty$. In investigating Example 6 and examining the eigenvalues as described below, this was found to be the reason for the convergence issues. Whilst the model estimation had converged to a solution, this solution had $|\mathbf{V}(\boldsymbol{\phi})| = 0$. One consequence is that statistics relying on either the Hessian or log likelihood cannot be applied. A review of the design matrix $\mathbf{Z}$ and/or the response data did not provide an immediate reason for the solution having $|\hat{\mathbf{V}}(\boldsymbol{\phi})| = 0$. Modifying the response data provided examples where this is not the case (even when some variance components are negative), for example dataset 4, and so it is not just due to the design $\mathbf{Z}$ and the model chosen.

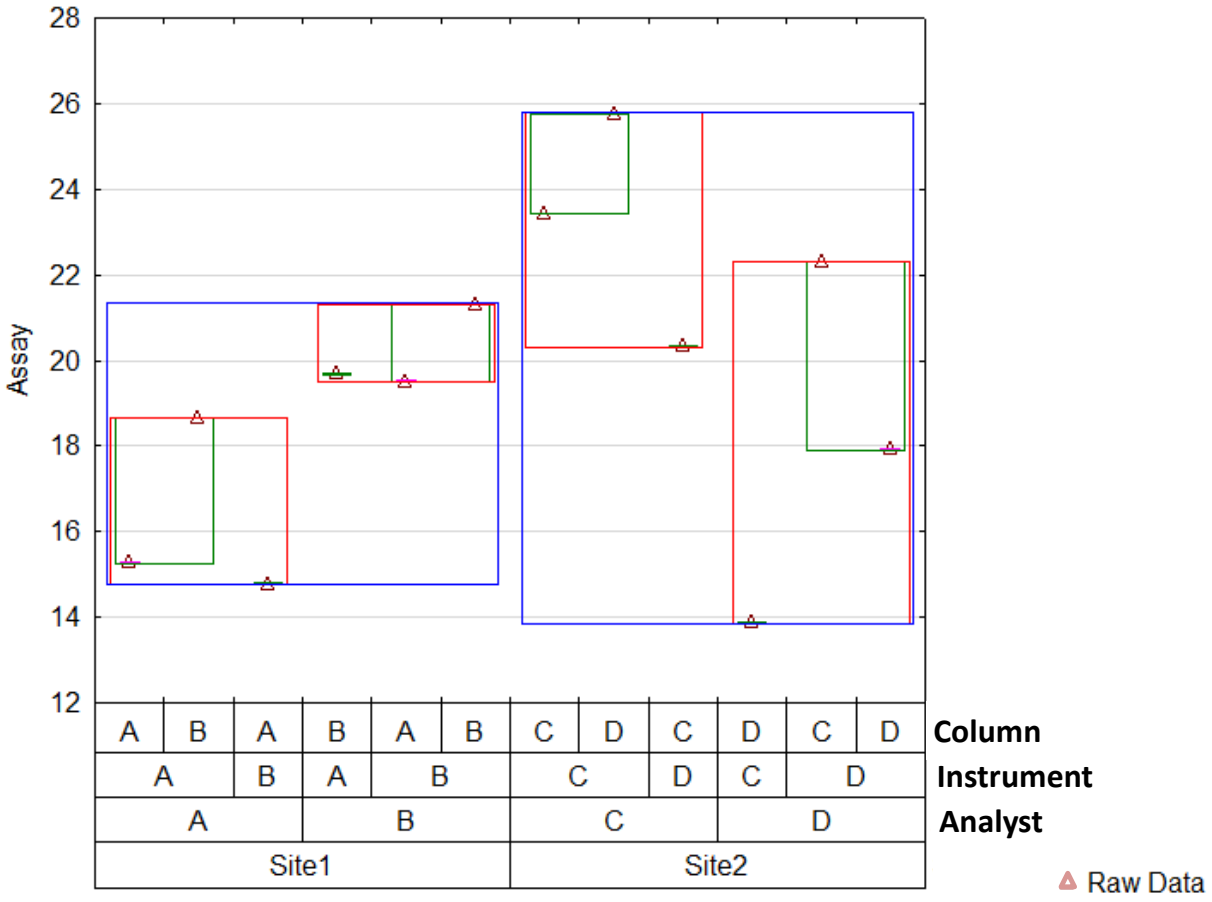
For Example 6 it is not clear whether the response data could reasonably come from a situation where the true variance components are ≥ 0 (i.e. due to unbalanced nature of the design or very limited data to estimate the variance components) or whether it is an indication that the data does not arise from the hypothesised situation and a variance components model is inappropriate. The data are shown in Figure 2.2 and there appears nothing particularly remarkable or abnormal about the data.

The eigenvalues of $\hat{\mathbf{V}}$ for this example were examined. It was found that two simple cases of an eigenvalue being 0 are when $\hat{\sigma_r^2} = 0$ which occurs in the trivial case that all the data values are the same and when $6\hat{\sigma_s^2} + 3\hat{\sigma_a^2} + 3\hat{\sigma_i^2} + 3\hat{\sigma_c^2} + \hat{\sigma_r^2} = 0$. This latter occurs when the site means are equal ($Var(\bar{y}_{site1} - \bar{y}_{site2}) = 2Var(\bar{y}_{site1}) = 2 \times \frac{1}{6}(6\sigma_s^2 + 3\sigma_a^2 + 3\sigma_i^2 + 3\sigma_c^2 + \sigma_r^2)$) It was also found that an eigenvalue will also be 0 when a root of the following characteristic equation equals $\hat{\sigma_r^2}$.

$$16\hat{\sigma_a^2}\hat{\sigma_i^2}\hat{\sigma_c^2} + 8(\hat{\sigma_a^2}\hat{\sigma_i^2} + \hat{\sigma_a^2}\hat{\sigma_c^2} + \hat{\sigma_i^2}\hat{\sigma_c^2})\lambda + 3(\hat{\sigma_a^2} + \hat{\sigma_i^2} + \hat{\sigma_c^2})\lambda^2 + \lambda^3. \tag{2.5}$$

This was the case for Example 6.

FIGURE 2.2: **Plot of the data for Example 6**



Whilst formulae can be provided for three real roots of this equation, it was not easy to see how these related to properties of the data itself. It is known that it occurred for a real GSK study and the data for Example 6 did not appear abnormal and hence a proposed method of analysis will need to accommodate these situations.

2.2 Investigating Aspects of the REML Analysis via Simulation

Simulations were performed to assess how often an infinite likelihood occurs in a realistic scenario and whether the simulations provide clues as to an interpretable property of the data which causes this. The frequency with which the variance components would be estimated as zero using REML for a typical situation was also examined since in those cases the chi-squared approximation does not provide confidence intervals. The simulations were performed in SAS version 9.2. However, except sometimes for the

content of the message accompanying a non-positive definite Hessian, example datasets gave the same results in SAS version 9.3.

2.2.1 Simulations and Fitting model

20,000 datasets were simulated with 12 or 16 experiments as per the unbalanced and balanced designs in Examples 3-6. Variances of σ_s^2=2, σ_a^2=1, σ_i^2=1, σ_c^2=1, σ_r^2=5 were used (denoted scenario "F_2_111_5"). Whilst the variance components are small in themselves, their sum is 5 the same as the residual variance. This is realistic in that it is aligned with a rule of thumb that the intermediate precision is typically twice the repeatability (though perhaps less realistic in that the site variance would be hoped to be minimal).

It was found that initially specifying the parameter estimates of the variance components with positive realistic values aided the model fitting. If no initial estimates are specified, SAS starts with all higher-level variance components as 0 (just providing an initial estimate of the residual variance) which often resulted in a note that the likelihood was infinite and no iterations were performed).

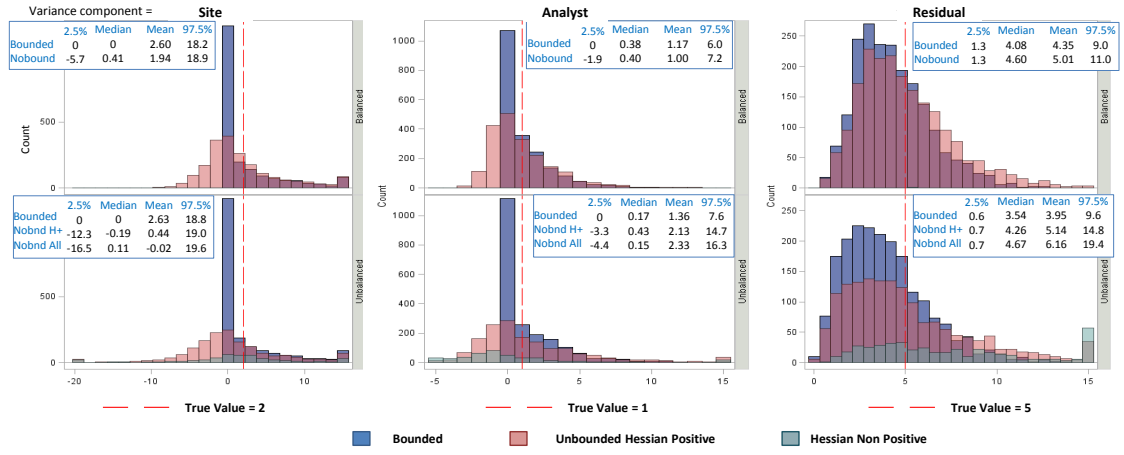
2.2.2 Convergence

- For the bounded analysis all analyses converged. For the balanced dataset of 16 values only 8.9% had a positive definite $\hat{\mathbf{G}}$ (all variance components were estimated as positive). For the unbalanced dataset of 12 values only 5.1% had a positive definite $\hat{\mathbf{G}}$ (probably lower than for the balanced dataset due to reduced number of values). This indicates that any analysis of such studies must cater for datasets where the REML estimates of the variance components are not positive.
- For the unbounded analysis but with the balanced dataset, 0.07% (13/20000) did not have a positive definite Hessian. This was found to be due to the site means being very nearly equal (which results in σ_s^2 being estimated as 0).
- For the unbounded analysis but with the unbalanced dataset, 21.9% did not converge with a positive definite Hessian. SAS reported one of three outcomes: 12.3% converged with non-positive definite Hessian; 5.9% did not converge; and 3.7% stopped due to an infinite likelihood (though all had started iterating). Some of the datasets from each type of outcome were examined and it was found that the estimates of σ_a^2,σ_i^2,σ_c^2,σ_r^2 satisfied Equation (2.5) that is required for an eigenvalue of $\hat{\mathbf{V}}$ to be zero (thus$|\hat{\mathbf{V}}| = 0$). For these datasets the outcome of fitting and type of message given by SAS often depended on the initial estimates of the variance and thus there seemed no purpose in differentiating between the three outcomes.

2.2.3 Bias

SAS Institute Inc.: The MIXED Procedure (2013) recommend that unbounded analysis be used in order that estimates of whole plot error (in this case sums of variance components) are not biased when non-positive variance components estimates are encountered. To evaluate this for scenario "F_2_111_5" and to consider whether bias is a problem, the REML estimates obtained from the bounded and unbounded models are compared. This was also evaluated for the scenario σ_s^2=2, σ_a^2=1, σ_i^2=5, σ_c^2=10, σ_r^2=5 (denoted "F_2_1510_5"). The REML estimates of the variance components and sums of the higher-level variance components ($\sigma_s^2+\sigma_a^2+\sigma_i^2+\sigma_c^2$ and $\sigma_a^2+\sigma_i^2+\sigma_c^2$) are plotted in Figures 2.3, 2.4 and 2.5 (for plotting purposes they are censored at 15).

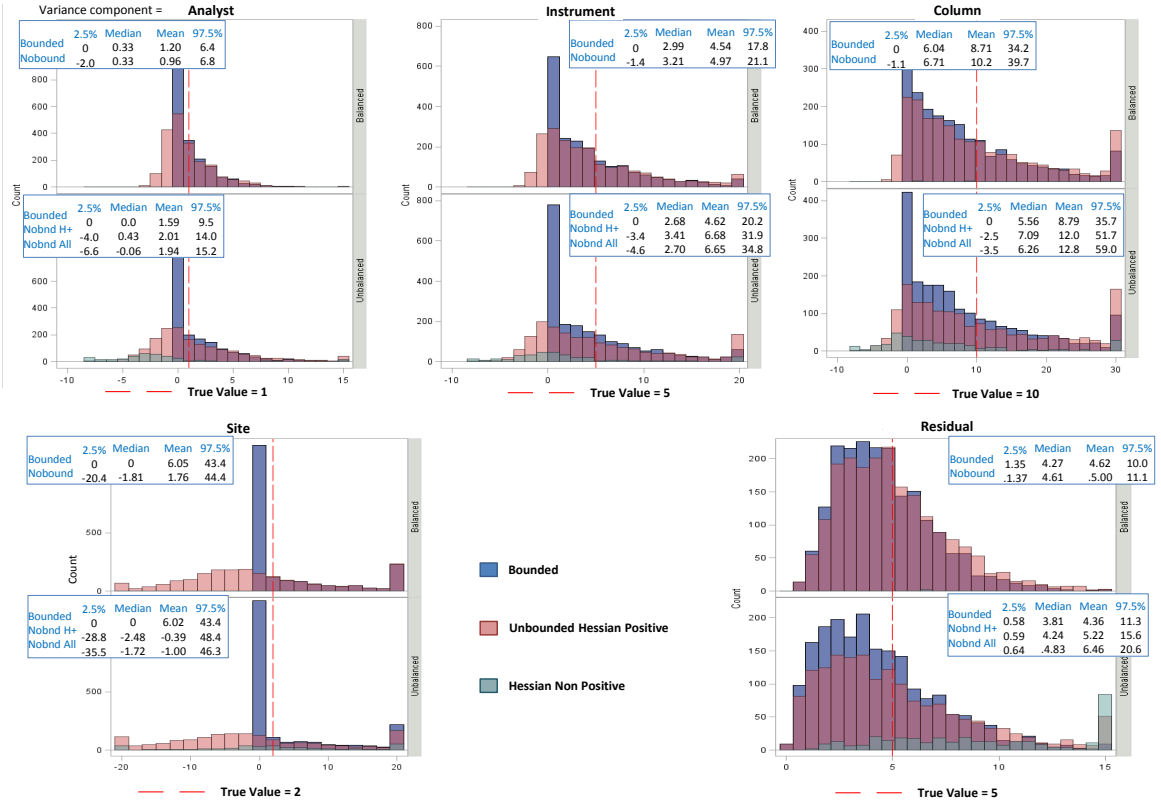
FIGURE 2.3: **Distribution of variance components estimates for scenario F_2_111_5**



It is seen from the summary statistics provided in Figures 2.3, 2.4 and 2.5 that for the balanced dataset the bounded estimators are biased whereas the unbounded are not. However, this did not hold for the unbalanced dataset where even the unbounded estimators are biased. It should also be noted that even when a variance component is as large as 10 (double the residual variance), with only 4 levels evaluated in the design (as for Column) the REML estimate will be non-positive reasonably often (see variance component σ_c^2 in Figure 2.4). When a variance component is small (see variance component σ_a^2 in Figure 2.3 which has a true value of 1) high estimates can be observed - more than 2.5% of estimates are above 6 and for the unbalanced unbounded analysis, above 15. The tails of the distribution were heavier for the unbalanced dataset which, though not verified, is thought largely due to the imbalance rather than fewer experiments.

Note whilst the bias of estimates is often considered, the concept may not be very useful in the context of a skewed distribution such as that seen for the variance component estimates. It may be that alternatives are more appropriate such as consideration of the median or the distribution of the estimates.

FIGURE 2.4: **Distribution of variance components estimates for scenario F_2_1510_5**



2.2.4 Confidence Intervals

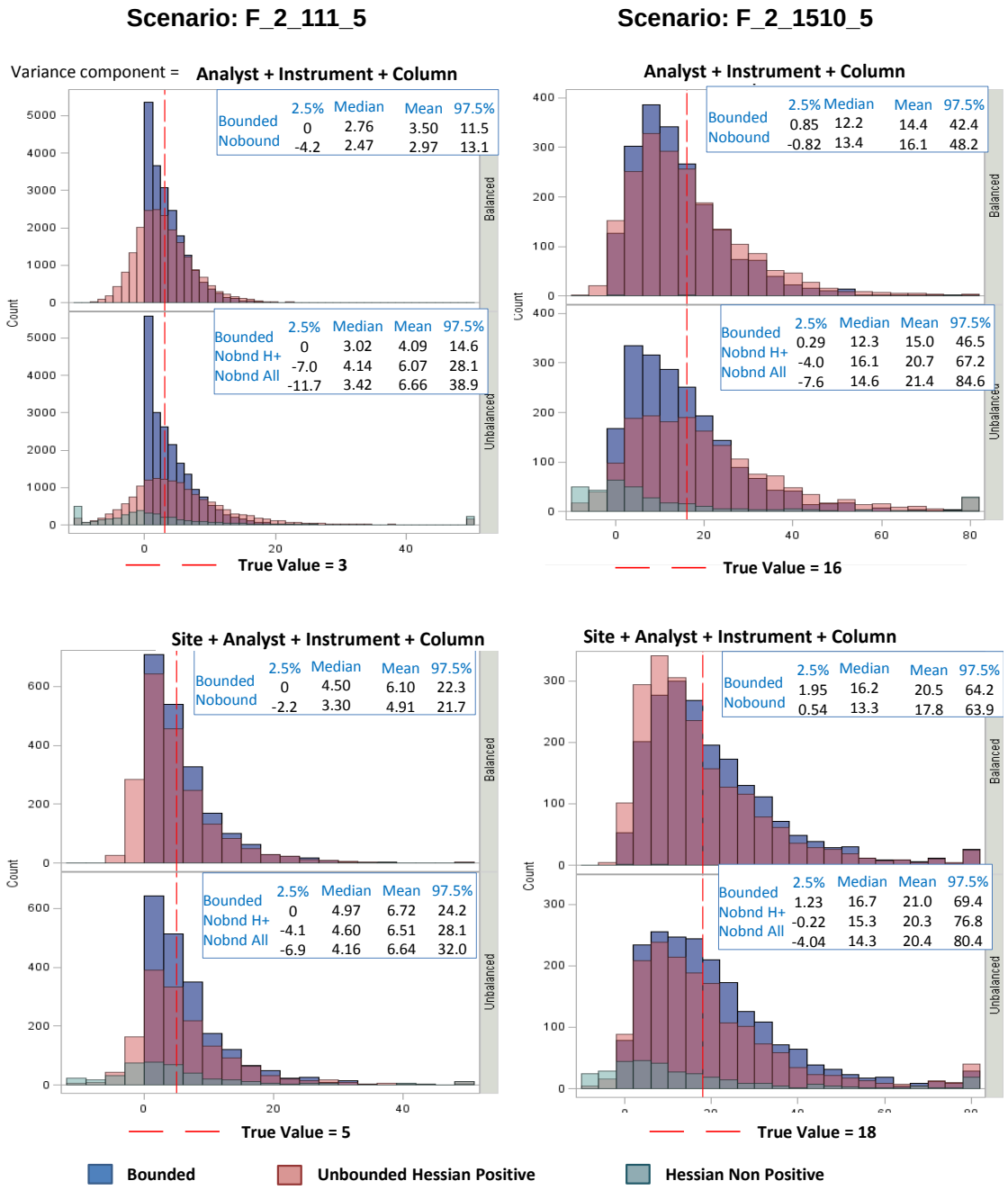
Table 2.3 shows the 95% confidence limit multipliers for simple estimates of variance and standard deviation (SD) based on a chi-squared distribution. It is reasonable that with a more complex design and model, the confidence limits will be even wider when estimating variance components.

TABLE 2.3: **95% confidence limit multipliers for variance and SD**

Number of		95% Lower CL		95% Upper CL	
Data Points	DoF	Var	SD	SD	Var
2	1	0.20	0.45	31.9	1018.3
3	2	0.27	0.52	6.3	39.5
4	3	0.32	0.57	3.7	13.9
5	4	0.36	0.60	2.9	8.3
6	5	0.39	0.62	2.5	6.0
7	6	0.42	0.64	2.2	4.8

Table 2.4 summarises the CIs produced by SAS for the Analyst variance component σ_a^2. Note that in the designs used for the simulations there are effectively 2 degrees of freedom available for estimating each of Analyst, Instrument, and Column variance components for which a simple SD estimate would have multipliers of 0.52 and 6.3 for

FIGURE 2.5: **Distribution of estimates for sums of higher-level variance components**



lower and upper confidence limits respectively (see Table 2.3). For the bounded analysis a CI is produced < 60% of the time. For the unbalanced, unbounded analyses, results are shown for all CIs produced and for those where the Hessian was positive definite. Again an appreciable number of simulated datasets fail to have accompanying CIs. The upper confidence limit is very high for the bounded analysis and in contrast seems low for the unbounded analysis (especially when compared to that for the simple SD estimate), though neither method of analysis provides good coverage. These results are in line with Lynn et al. (2010) who also suggested other software algorithms provide better CIs since they applied an asymptotic normal approximation to $\log(\hat{\sigma}_e^2)$. It should be noted that the notional 2 degrees of freedom available for estimating Analyst, Instrument, and Column variance components most likely provides estimates with such large uncertainty that they are practically useless. However, initially our interest is in a methodology for estimating the uncertainty of variance components, particularly sums of variance components. If the uncertainty is large that will encourage the use of designs which can provide better estimates. The wide spread of estimates seen in Figure 2.4 illustrates why the CIs are wide, for example, estimates of 0 are not infrequent even for $\sigma_c^2 = 10$.

TABLE 2.4: **Median 95% confidence intervals for σ_a^2 and σ_a for scenario F_2_111_5**

		CI	95% Lower CL		95% Upper CL		
Dataset	Analysis	produced	Var	SD	SD	Var	Coverage
Balanced	Bound	58%	0.26	0.51	81	6560	87%
Balanced	Nobound	100%	-3.33	-	1.92	3.67	77%
Unbalanced	Bound	51%	0.36	0.60	78	6041	80%
Unbalanced	Nobound	90%	-4.23	-	1.82	3.32	68%
Unbalanced	Nobound(H+)	78%	-4.47	-	2.07	4.27	74%

2.3 Summary

The analyses of four example datasets typical of a small variation study were investigated in Section 2.1 using a REML model. Bounded and unbounded models where variance components are restricted to being non-negative or not, respectively, were applied as described in Section 2.1.2. Once suitable initial values are provided (if necessary), the analyses achieved the convergence criteria based on the Hessian (2nd derivatives) matrix. However, for one of the unbalanced datasets and the unbounded analysis the Hessian was not positive definite, though there was nothing particularly abnormal about the data. For the bounded analyses confidence intervals (CIs) were not obtained when variance components were zero (or in one case where the variance component was small). For the unbounded analyses CIs were obtained, but are not considered appropriate given that the true variance components will be non negative. In the situation where the Hessian

was not positive definite CIs were provided but the statistics provided by PROC MIXED suggest that there is a numerical problem and the CIs should not be relied upon.

A simulation study was then performed to investigate whether a non-positive definite Hessian matrix is a very rare occurrence or something which is likely to be encountered in routine use. It was also assessed how often REML estimates of variance components are zero.

For scenario F_2_111_5 we observed that for the unbalanced dataset and unbounded estimation only 78% of datasets converged with positive definite Hessian. Even if the true variance components for Analyst, Instrument and Column were increased from 1 to 5 (i.e. equal to residual variance), the %converged with positive definite Hessian only increased to 83%. Given such a large percentage of analyses failed to converge with a positive definite Hessian when an unbounded analysis was performed, this suggests that Example 6 (which was considered typical) and the original GSK dataset which motivated this investigation are not particularly unusual and that the variance components model is not inappropriate.

Though Stroup and Littell (2002) recommend that unbounded REML variance component estimates should be used when testing fixed effects, it is seen that for these types of study and if the dataset is unbalanced this is often not feasible to implement as a non-positive definite Hessian is obtained. It was also seen that if the dataset was unbalanced the variance component estimators were not unbiased even for the unbounded analysis. Thus the model and analyses used in later chapters will be the bounded analysis described in Section 2.1.2, corresponding to $\hat{G}$ in Equation (1.1) being non-negative definite. It is noted that the lmer function in the lme4 package (Bates, 2010) in R does not provide negative variance components in the model. It is acknowledged that for the bounded analysis, the estimators are biased. However, this thesis is concerned with the uncertainty of the variance component estimators and thus the estimate to be used when testing fixed effects is considered outside the scope of this research project.

For the bounded analysis even for the balanced dataset in 42% of the simulated datasets no confidence interval is produced as the REML estimate is 0. In fact the counter intuitive situation occurs that, with a fixed s.e., as an estimate of the variance component gets closer to 0 the upper confidence limit can get larger. Thus this investigation has shown that the chi-squared confidence interval produced by SAS for a bounded analysis does not provide a solution which can be used routinely for these types of analyses. A methodology is required which can accommodate datasets which would produce a zero REML estimate for a variance component.

Chapter 3

Literature Review

The aim of this thesis is to provide a statistical approach which can be used routinely in industry to quantify the uncertainty of variance estimators in analytical and process variability studies. In Chapter 1 this aim was introduced together with background to the application area and the issues encountered. Firstly this chapter references literature relevant to the application area of process and analytical variation studies. Following that, a number of frequentist approaches based on likelihood or method of moments methodology are critiqued. Given the complexity of industrial variation studies the frequentist methods seemed unlikely to provide a satisfactory approach for routine statistical analysis. Hence Bayesian methodology was considered and the advantages of taking this approach are discussed. A necessary component of the Bayesian approach is the choice of a prior. Some relevant simulation studies which compare frequentist methods with Bayesian methodology and/or assess the appropriateness of priors for credible interval estimation are then described. The simulation studies referenced provide some information on priors they evaluated and a list of the main priors suggested by the literature for variance components is provided. However, discussion on the choice of priors, those evaluated in this research project and the related literature is predominantly given in Chapter 5. Some of the literature references published later than 2010 are discussed in the conclusions to the research project (Chapter 10).

In many applications, scientific studies are performed in which variation arises from a number of sources and the size, or relative size, of the components of variance must be quantified. This allows efforts to be focused on reducing the largest sources of variation, it can be a step towards estimating other quantities of interest, or it can aid the design of future studies to estimate quantities of interest. Components of variance may contribute to the ultimate quantity of interest, e.g. the total variation, or be nuisance sources of variation. As well as requiring reasonable point estimates of variance components, it is important to understand the uncertainty associated with those estimates. This is particularly important given the typical small size of the studies - likely to be insufficient to produce adequately precise estimates of the variance components.

In industry understanding sources of variation in a process or measurement system is useful in assessing, improving and quantifying analytical or process capability e.g. through measurement systems analysis studies such as gage reproducibility and repeatability (gage R&R). A classic gage R&R used in industry is a two-way crossed random effects design with parts crossed with operators with replicate measurements. Gaining an understanding of the uncertainty associated with the variance component estimation has been promoted in the literature for these designs e.g. Vardeman and VanValkenburg (1999) and Burdick et al. (2005), with frequentist methods being proposed. However, Johnson and Deaner (2014) describe the use of expanded gage R&Rs which incorporate additional factors to those routinely used in typical gage R&Rs. In the pharmaceutical industry many applications exist through the need to quantify both product and analytical method variability such as: designing schemes for batch sampling; quantifying and improving precision of analytical methods; and designing appropriate studies to demonstrate equivalence of processes or analytical methods. Borman et al. (2011) describe analytical method ruggedness studies which identify whether potential analytical sources of variation appreciably affect the method's precision. In Borman et al. (2009), example 3, the information gained on the variance components from a ruggedness study is used in the design of an equivalence study. Some studies may have a large number of random sources of variation being considered. Borman et al. (2011) explain that the estimate of the total analytical variance is often then compared to the specification limits for the product (the precision to tolerance ratio) to ensure that the analytical variation does not take up too much of the allowable tolerance. For processes, estimates of variability are used to assess the capability of the process to remain within the drug product specifications (for example see K.M. Kurian (2008); Wu (2008)), often in conjunction with the process mean. Tolerance intervals are used to quantify the typical variability for a process and can be used when proposing product specifications. Though most applications seen in the literature have at most one higher-level factor, in practice the data will have many sources of variation applying. For example, one batch of active ingredient may be used in several batches of drug product which is then used to make a large number of tablets from which a few are selected to be tested. In Chapter 1 it was noted historically there has not been a requirement to report the uncertainty associated with an analytical method variance estimate within the pharmaceutical industry. However, there are signs that expectations are changing - a United States Pharmacopoeial (USP) stimuli article promotes the derivation of two-sided beta-content tolerance intervals (intervals that contains at least $100\beta\%$ of population with given confidence level γ) for the result of the method (Barnett et al., 2016) illustrating the approach with confidence levels of 50% and 90%. However, the spreadsheet tool provided to support the article only addresses one variance (repeatability) and does not incorporate higher random effect(s). A further USP article on measurement uncertainty for the pharmaceutical industry (Weitzel et al., 2018) states "The pharmaceutical industry typically reports a point estimate (a 'reportable value') without its uncertainty" and quote from

the standard ISO21748:10:

> "Knowledge of the uncertainty associated with measurement results is essential to the interpretation of the results. Without quantitative assessments of uncertainty, it is impossible to decide whether observed differences between results reflect more than experimental variability, whether test items comply with specifications, or whether laws based on limits have been broken. Without information on uncertainty, there is a risk of misinterpretation of results. Incorrect decisions taken on such a basis may result in unnecessary expenditure in industry, incorrect prosecution in law, or adverse health or social consequences.

Statistical methods for estimating variance components are provided in the literature. One way to analyse variation studies is through an analysis of variance or method of moments approach providing least-squares-based estimates (Cox and Solomon, 2003). Another is through maximum likelihood, usually restricted (or residual) maximum likelihood (REML) (Cox and Solomon, 2003), where the likelihood that is being optimised is a projection of the data so that it no longer contains the fixed effects.

In deciding on the broad approach to take for this research project which aims to quantify the uncertainty of variance estimators, the literature was reviewed in 2010 (though some later relevant papers are also mentioned here). A number of methods for constructing confidence intervals (CIs) have been proposed in the literature. Wald-type CIs and profile likelihood have been suggested for a REML analysis. Satterthwaite, modified large sample (MLS) and generalized CIs have been suggested if the analysis is through a method of moments approach. Bootstrapping is another possible approach.

Firstly let us discuss CIs related to a likelihood approach e.g. REML. In considering Wald-type CIs it should be noted that their implementation can differ between software and even according to the choice of analysis. Lynn et al. (2010) compared these across REML performed in four commercial software packages and find differences in performance - mainly due to the parameterization employed. The literature however, suggests that Wald-type CIs are particularly inappropriate for small samples (Hox, 2002) (the simulations in Chapter 2 also confirmed this). Though profile likelihood is suggested by Bates (2010) (and was implemented in R in a developmental version of lmer at the time of the review) this is also likely to suffer from small sample problems as the optimality properties of likelihood methods are asymptotic. It also doesn't take into account the uncertainty of the other parameters which are profiled out (optimised over). Likelihood ratio tests and CIs also need adjustment when the variance components estimates are on the boundary of the parameter space; from the simulations performed in Chapter 2 it was seen that this is likely to occur for these studies. Some literature exists on overcoming this issue: Crainiceanu and Ruppert (2004) proposed a fast algorithm for computing

the exact null distribution of the restricted likelihood ratio test for models but only with a single variance component, and (subsequent to 2010) Wang and Chen (2012) have computed this for models with independent random effects each depending on a single variance parameter. In the example used by Bates (2010) some light is shown into why the parameterization used by the different software was important when applying Wald-type CIs. The square root of the likelihood statistic was seen to be more linearly related to $\log(\sigma)$ than σ or σ^2. Browne and Draper (2006) investigated five likelihood methods in addition to the Wald CI. However, most of the methods required a non-zero variance component estimate. Browne and Draper point out that because the methods use maximisation rather than integration over the other parameters of a highly-skewed likelihood surface this is a fundamental reason why likelihood-based methods are likely to underperform for small samples. Bootstrap CIs are also based on large samples, so there is no reason to expect them to work well for small samples and parametric bootstrapping will depend on the point estimates, which is likely to be a particular problem as REML estimates of zero are frequently expected.

Secondly I examine method of moments approaches. Burdick and Graybill (1992) provided a good summary of the literature as at 1992, with Burdick et al. (2005) and Sahai and Ojeda (2005) providing a more recent comprehensive discussion for various designs with the latter focusing on unbalanced designs. However, except in the simplest case, any CIs proposed for σ_h^2 (effect h denoting a higher-level random effect i.e. one which is not the bottom-level effect) or sums of such variance components, are approximate (Burdick et al., 2006). They are also based on modifications of large sample methods, though Burdick and Graybill (1992) state "the methods presented in this book work well for all sample sizes. The intervals are generally approximate, but in most cases maintain the stated confidence level". These methods are often proposed in the context of gage R&R studies (Vardeman and VanValkenburg, 1999; Burdick et al., 2003; Gong et al., 2005), though almost entirely for two-way mixed or random effect models (though Adamec and Burdick (2003) provide an example for three random factors). The analytical variation studies described in Chapter 1 are more complex, having many variance components, nested and crossed random effects, may not be balanced as well as having small sample sizes. In 2011 the R package "varcompci" (Civit et al., 2011) became available - it implemented methodology based on Burdick and Graybill (1992) for individual variance components for any mixed model involving five or fewer factors. The package's aim was to fit mixed models to balanced datasets, though it was able to fit models to unbalanced datasets. As Hess and Iyer (2002) noted in a previous paper "The methods considered here remain largely untested for unbalanced data" (note the SAS macro referred to is not available as it had an error). The R package "varcompci" is no longer available from the CRAN repository ("archived on 2014-09-19 as vignette locations were never updated for R 3.1.0"). A paper by Nijhuis and Van den Heuvel (2007) discusses Welch, Satterthwaite, and MLS CIs applied to precision estimates, though just for a three factor design and the investigations assume a balanced design. Generalized CIs

which extend the definition of a CI to cater for situations where nuisance parameters are present have been proposed for balanced designs (Weerahandi, 1993; Iyer and Mathew, 2002; Iyer and Patterson, 2002; Ye and Wang, 2007). Bootstrapping has also been applied with the method of moments approach but the implementation is not trivial. Tong and Brennan (2007) discuss various methods for bootstrapping in the context of educational studies having 2 or 3 factors and their implication for bias. Wang and Li (2003) apply bootstrapping to a gage R&R study, though acknowledge that for most practical applications and a nominal coverage of 95% the actual coverage may be around 85% - referring to Prada-Sáichez and Cotos-Yánez (1997).

A useful review of the literature and methodology related to small studies with one higher level random effect and the bottom random effect, is provided by McNeish and Stapleton (2016a).

In the context of the literature on CIs, industrial designs can be complex: have small sample sizes; can have many variance components, both nested and crossed in the design; and may not be balanced, even for the bottom-level factor. Thus it is not clear how well in practice the above methods will apply to such designs and, even if applicable, often the methodology will be difficult for practitioners to extend to the more complex designs. Given that the above approaches are unlikely to be suitable for the very small designs being used and for which an estimate of the uncertainty is particularly important, a Bayesian approach is an attractive alternative. Bates (2006) recommends using a Markov Chain Monte Carlo sample to evaluate the properties of individual coefficients. A Bayesian approach does not require specific derivations of confidence intervals for particular design types and easily incorporates fixed factors to form a mixed model analysis. Another important advantage of a Bayesian approach is that intervals for quantities of interest derived from the variance components are easily obtained. Further discussion of the attractiveness of using a Bayesian approach is provided in Weaver et al. (2012). In addition, whilst not a topic for this thesis, a Bayesian approach is attractive in that it provides a framework for making use of previous data through the use of more informative priors.

In applying a Bayesian approach it is necessary to choose a prior. When analysing a single study it may be appropriate to apply a range of priors, assess the effect on the results of the analyses and choose the best interpretation and conclusions from those. However, for small studies the choice of prior could have a substantial effect for a particular dataset and thus interpretation and conclusions may be difficult. Also, if a Bayesian approach is to be used routinely for a type of study in industry, it may be desirable to find a class of priors which will perform reasonably without an in-depth investigation of each dataset e.g. for analytical ruggedness studies as illustrated in Figure 1.2 or for batch sampling studies described in Section 4.1. Many vague priors have been suggested in the literature for variance components: Jeffreys' prior (Jeffreys, 1946); reference priors (Bernardo, 1979; Berger and Bernardo, 1992); uniform-shrinkage prior (Daniels, 1999);

inverse gamma on the variance, IG(ϵ, ϵ) where ϵ is small (Lunn et al., 2000); uniform, half-t or folded noncentral-t on σ (Gelman, 2006). However, there is limited evaluation of their suitability in the situation where they are used routinely and thus their longer term properties are of interest, especially in the context of estimating the uncertainty of the estimator.

The following papers provide useful information on simulation studies for models with random effects where likelihood and Bayesian methods are compared and coverage properties of intervals for parameters based on variance components are provided:

- Lambert et al. (2005) assessed 13 priors for one higher-level random effect for a meta-analysis of odds ratios from 5, 10 or 30 units (studies) for between-unit standard deviations of 0.001, 0.3 and 0.8. Four of the priors which were uniform on the variance metric were particularly poor with a small number of units. For the other priors, gamma and logistic on the precision metric, uniform and half-normals on SD and uniform on log variance there was no one prior that was particularly recommended from the study. They noted that "Some of the prior distributions used here are clearly unrealistic in that they give support to unfeasibly large values for the between unit standard deviations." and recommended "investigation of prior distributions that are vague within a realistic range for the data set under consideration". All the priors had extremely poor coverage ($< 5\%$) for the between-unit standard deviations of 0.001 which they explained was due to an upward bias as the posterior samples are positive. The credible intervals used were percentile intervals.
- Browne and Draper (2006) compared Bayesian and likelihood-based methods for non-informative inverse gamma and proper uniform priors on the variance components. In the case of the two-level variance components model (one higher-level random effect and the bottom-level random effect), the general conclusion was that the uniform prior gave close to nominal levels across the study designs with 12, 18 or 24 units at the higher level and at the parameter settings examined (though none reached 95%). However, when there were only 6 units at the higher level the coverage for the uniform prior was in the region of 91%. They also noted that intervals for the uniform prior were extremely wide and further work is needed to see if other prior specifications might yield narrower but still well-calibrated intervals. They found that the inverse gamma prior did not perform as well and under covered in some situations - particularly for small values of the intraclass correlation (for 6 units the coverage was less than 89%). An additional reason would be that the credible intervals used were percentile intervals.
- Though for a different situation (random slopes/coefficients regression model), Browne and Draper (2000) compared two Wishart priors and an improper uniform prior on the covariance matrix for the intercept and slope for a higher level random effect. The simulation study investigated the higher level random effect (school) at

12 and 48 levels. For 12 levels and nominal 95% intervals, the improper uniform prior had coverage as low as 82%, and one of the Wishart priors (which was "gently data-determined") had coverage less than 80%, for some component of the covariance matrix. The alternative Wishart prior was better but still had coverage as low as 88%.

- Also for the random slopes/coefficients regression model), Tsai and Hsiao (2008) compared three reference priors (two based on Jeffreys and a uniform shrinkage prior) with REML. One of the simulations was for a six response, six random effects model with varying true values between 1 and 4 with the error variance fixed at 1. The study had 15 subjects. Their conclusion was that the Bayesian analyses performed well in contrast to REML. Based on the distance metrics used (squared Euclidean norm and squared error risk) they concluded the two Jeffreys' priors performed best. Coverages for Jeffreys' priors were 1.00 for all six random effects compared to 0.98 for the random intercept and 0.83-0.91 for the other five random effects for the uniform shrinkage prior. They used the R package "lme" to perform the REML analysis and the CIs obtained had very low coverage 0.09-0.23 (assumed to be Wald CIs though not explicitly stated in the paper).

Given the advantages of taking a Bayesian approach and the concern of applicability of frequentist methods to the complex industrial studies it was decided to adopt a Bayesian approach in the research project. It is noted that this decision was made for practical reasons rather than from a philosophical perspective - this is further discussed in Section 7.1.

It should be noted that the following study published after 2010 found in favour of frequentist methods for estimating the intraclass correlation coefficient. This will be further commented upon in the research conclusions.

- Ionan et al. (2014) investigated estimating the intraclass correlation coefficient (in range 0.71-0.99) for measurement reliability for a two-way crossed random effects model with 48 or 96 biological samples and 3 or 6 levels for lab. They concluded non-informative inverse gamma priors on individual variance components performed poorly. Though they considered an improper uniform prior on the variance component standard deviation was better, they concluded that if there were less than 8 labs, the generalized CI and modified large sample methods were better in the scenarios they had examined for estimating intervals for the intraclass correlation.

Given the attractiveness of a Bayesian approach but lack of assessment of the performance of priors in the literature for small designs, this thesis evaluates a range of priors in order to assess their suitability if used routinely. An example study used for the evaluation is a two-way variance components design of the type which can be used in

assessing batch uniformity in pharmaceutical industry. The two-way variance components (three level) design was chosen as in industry there are often multiple variance components even though the literature is focused on models with only two levels. The review in McNeish and Stapleton (2016a) only found one paper for more than two levels when looking the effect of small sample sizes on model estimation. Though only a two-way variance components design is to be examined, the findings are expected to extend to variation studies which have more than two higher-level effects. A two-way variance components design can be obtained from one with more random effects by taking means over random effects lower in the design or ignoring random effects higher in the design. The priors which are to be studied are described in Chapter 5. Criteria for their choice are developed and discussed, especially in the context of their use in studies estimating sources of variation. Some priors were selected with an intention that they were very vague, whilst others were intended to be mildly informative. Literature related to the choice of prior is discussed in Section 5.3.

In this research project an evaluation of the chosen priors is performed for a batch sampling design described in Section 4.1 and for various alternative scenarios described in Section 4.3. The Bayesian analysis and simulations performed are described in Chapter 6. The frequentist concept of coverage is then used to evaluate the suitability of the priors in Chapter 7. An evaluation of the variance components estimates and their uncertainty for some of the priors is given in Chapter 8. The reasons for the poor coverage are further investigated in Sections 7.3 and 8.4.

Chapter 4

Scenarios and Model used for the Evaluation of A Bayesian Approach for Variance Estimation

Contents

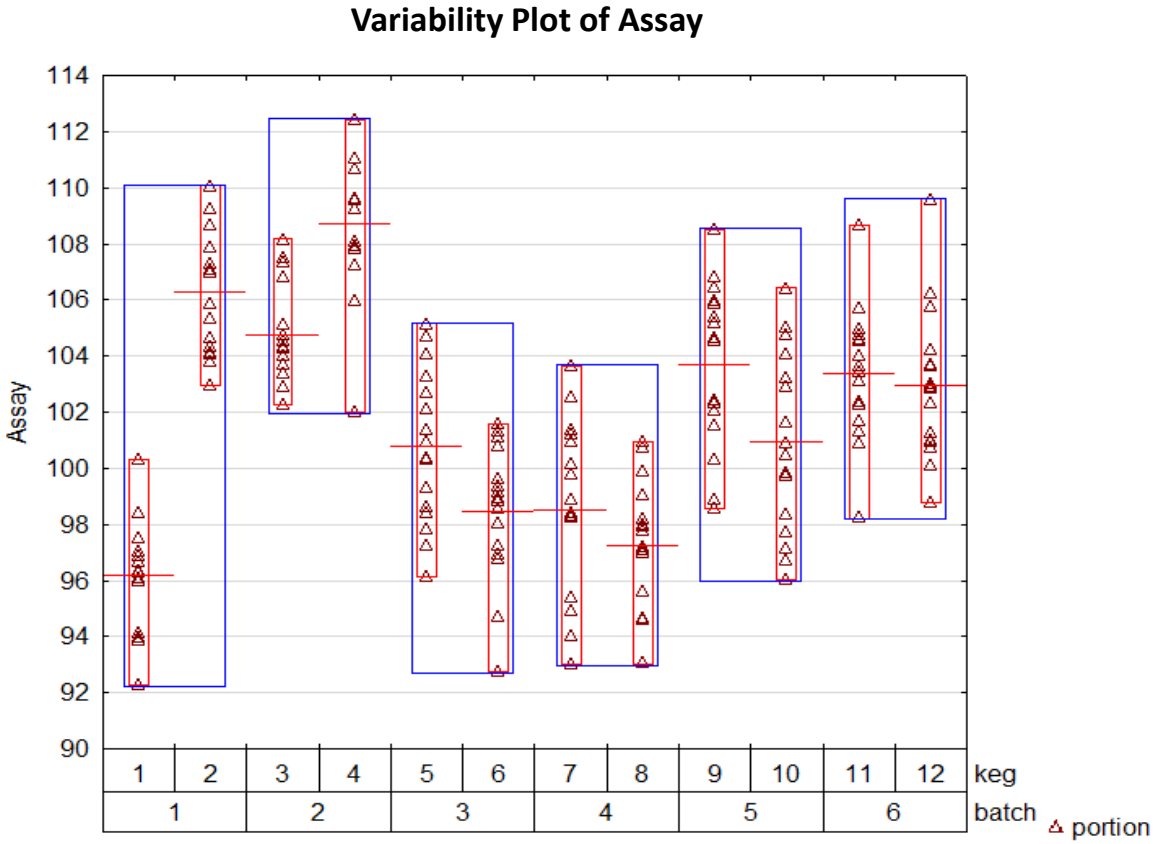
A Bayesian approach was chosen to estimate the uncertainty of variance components estimates as discussed in Section 1.5. This chapter describes the scenarios to be used for that evaluation. A real batch sampling study was used as the basis for the evaluation of the Bayesian analysis and its study design is described in Section 4.1. The statistical model which is assumed is given in Section 4.2. Modifications to the design of the original study were made together with varying the true variance component values to give a range of scenarios to be evaluated. These are described in Section 4.3. The Bayesian analyses performed for these scenarios are described in Chapters 5 (choice of priors) and 6 (analyses performed), and the results are given in Chapters 7 (coverage of credible intervals) and 8 (summary of credible intervals).

4.1 Design of Batch Sampling Study used as a Basis for the Evaluation

To study the effect of the prior on the estimation of the uncertainty of variance components estimators a two-way variance components design was chosen (see Section 1.3), i.e. it has three random effects in total. The actual designs evaluated were based on a real

batch sampling study which consisted of 6 batches of material, 2 kegs taken from each batch and 16 portions taken from various parts of the keg. The aim of the study was to estimate the between and within batch variation in the assay of the active ingredient and to assess variation between kegs. The statistical analysis assumes that the results from portions within a keg come from a random distribution and there is no systematic effect of location within a keg. Figure 4.1 shows an example data set - simulated due to commercial sensitivity.

FIGURE 4.1: **Example batch sampling study (simulated data)**



The results are desired to be applicable to other applications and designs with different numbers of variance components. For example, a two-way variance components analytical precision study could have a number of analysts performing a number of analytical runs over several days, and on each day/run analyse a number of preparations. In practice there may be more than 3 random effects but for exploration of the effect of the choice of prior the chosen study design has three. It was decided not to extend to more than three random effects as this was likely to make the evaluation unwieldy and wouldn't provide much further gain or insight. The evaluation was not restricted to two random effects (on which much of the frequentist literature concentrates) as this is unlikely to provide results extending to the studies of interest. To ensure applicability

to other applications the values explored are relative to the bottom-level variance (assumed to be around 6). Very weak knowledge is assumed that the higher-level variance components are likely to be within the range 0.5 to 24 corresponding to $\frac{1}{12}$ and 4 times the portion variance. When the variance is broken down into many potential sources, often many variance components will be close to 0. Since interest is in not only the uncertainty of the estimates of the variance components, but also their contribution to the total variance, it is important to evaluate a value as small as $\frac{1}{12}$ of the portion variance. In addition it was seen in Chapter 2 that small variance components caused issues in the frequentist analysis. It is noted that many other applications will not have as many levels for the bottom-level factor (16 portions per keg in the original study) and thus a smaller number will also be studied.

4.2 Statistical Model

A linear mixed model generally used in the analysis of data from a study aiming to estimate variance components was given in Equation 1.1. This model is assumed for the analysis of the example study described in Section 4.1. Note this example is called a two-way variance components model where there are two higher-level random effects / variance components in addition to the bottom level effect/(residual) variance.

The model for the measurement y_{ijp} made on batch i (effect b_i), keg j from batch i (effect k_{ij}) and portion p from keg j and batch i (effect ϵ_{ijp}) is

$$y_{ijp} = \mu + b_i + k_{ij} + \epsilon_{ijp}, \tag{4.1}$$

where μ is the overall mean, $b_i \sim N(0, \sigma^2_{batch})$, $i = 1, ...B$, $k_{ij} \sim N(0, \sigma^2_{keg})$, $j = 1, ...K$, $\epsilon_{ijp} \sim N(0, \sigma^2_{portion})$, $p = 1, ...P$, and all random variables are independent.

In the remainder of the thesis the variance components corresponding to batch, keg and portion may be abbreviated by σ^2_b (σ^2_{batch}), σ^2_k (σ^2_{keg}), and σ^2_p ($\sigma^2_{portion}$). In addition σ^2_h is used to represent either of the higher-level variances σ^2_{batch} or σ^2_{keg} and σ^2_{tot}=σ^2_b+σ^2_k+σ^2_p. For the original study B=6, K=2 and P=16.

4.3 Scenarios to be Evaluated

Seven scenarios based on the two-way variance components design described in Section 4.1 will be used in the evaluation of the priors. The number of levels of each factor and values of σ^2_k were varied. The scenarios are listed in Table 4.1 where the number of batches (B), the number of kegs per batch (K) and the number of portions per keg (P) are shown and the true variance component values are also given. For each scenario,

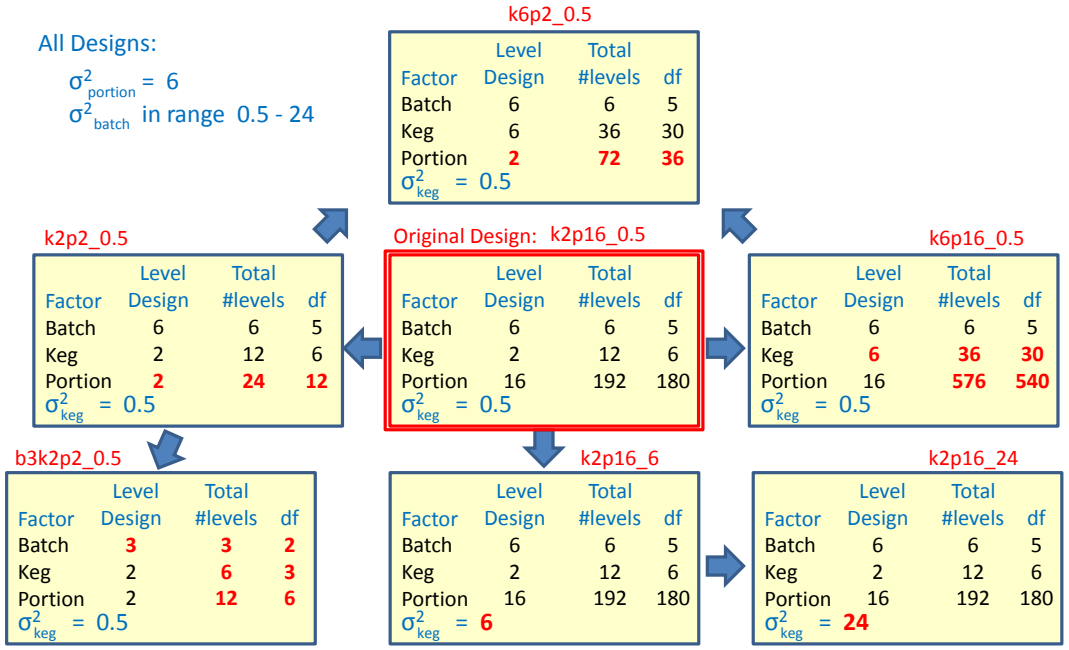
the σ_b^2 values investigated (0.5, 6, 24) encompassed values around σ_p^2 for which it was considered important to estimate uncertainty, as discussed in Section 4.1. The scenarios are also represented in Figure 4.2 showing how they relate to each other and the size of the design. The abbreviation for each scenario is given at the top of the box containing the details for each scenario. Details for the original batch sampling study design (k2p16_0.5) are given in a central box (outlined in red). Level Design shows the number of factor levels per factor level above. To give an understanding of design size, the total number of levels for each factor and the degrees of freedom which would be available for a traditional ANOVA analysis are provided. Changes between the designs evaluated are shown in bold (red) with arrows indicating the designs being compared.

The model to be fitted to the data from each scenario was given in Section 4.2 and the parameter values are given in Table 4.1.

TABLE 4.1: **Model parameter values for each scenario**

Scenario	B	K	P	σ^2_{batch}	σ^2_{keg}	$\sigma^2_{portion}$
b3k2p2_0.5	3	2	2	0.5, 6, 24	0.5	6
k2p16_0.5	6	2	16	0.5, 6, 24	0.5	6
k2p16_6	6	2	16	0.5, 6, 24	6	6
k2p16_24	6	2	16	0.5, 6, 24	24	6
k2p2_0.5	6	2	2	0.5, 6, 24	0.5	6
k6p16_0.5	6	6	16	0.5, 6, 24	0.5	6
k6p2_0.5	6	6	2	0.5, 6, 24	0.5	6

FIGURE 4.2: **Scenarios to be used in the evaluation of priors**



Chapter 5

Priors for Evaluation

Contents

In order to perform a Bayesian analysis using the model described in Section 4.2, priors were required for each model parameter to be estimated. This chapter explores various aspects of the choice of the priors for estimating variance components and describes those chosen for the evaluation of methods of estimating the uncertainty of variance component estimates. Six families of priors were chosen to be screened and are listed in Section 5.1.1. An overview of how their distribution parameters were selected so that the results of the evaluation are applicable to other designs and applications is given in Section 5.1.2. The specific priors investigated are listed in Table 5.1 and described further in Section 5.2. A review of the literature and discussion of the choice of priors is provided in Section 5.3. The Bayesian analysis performed using these priors is described in Chapter 6 and the results are given in Chapters 7 (coverage of credible intervals) and 8 (summary of credible intervals).

5.1 Overview

5.1.1 Families of Priors Chosen

Sets of priors were chosen for investigation which related to six families of priors for the parameters for the higher level variance components (e.g. σ_h) corresponding to the model given by Equation (4.1). The families investigated were:

- Uniform prior on σ_h
- Half-T prior on σ_h
- Beta prior on ratio $\frac{\sigma_h^2}{\sigma_h^2+\sigma_p^2}$
- Inverse Gamma prior on stratum variance (linear combination of variance components proportional to the between-group variance for that stratum)
- Jeffreys' prior on variances associated with random effects
- Flat prior resulting in a posterior proportional to the likelihood

An extremely diffuse normal prior, $\mathcal{N}(0, \sigma^2 = 10^{10})$, was placed on the overall mean for the first three families of priors listed above and an improper uniform prior across the real line for the remainder. Note a prior $\pi(\theta)$ is said to be improper if $\int \pi(\theta)\, d\theta = \infty$. A proper uniform prior on the parameter for the bottom-level random effect, σ_p, was used for the first three families of priors listed above and for the remaining families an alternative, related to the priors chosen for the parameters for the higher-level random effects, was used.

The specific priors investigated for the variances are listed in Table 5.1 together with the abbreviations which will be used in the remainder of this thesis. The priors are described further in Section 5.2 and discussion of the choice is given in Section 5.3.

TABLE 5.1: **Summary table of priors and their abbreviations**

Abbrev.	Section	Brief Description
UNI	5.2.1	Uniform prior, U[0,8.66] on the batch and keg SD and U[0,12.25] on the portion SD.
HCY	5.2.2	Half-Cauchy (scale=8.66) on keg and batch SD and U[0,12.25] on portion SD
HT3	5.2.2	Half-t (scale=8.66, 3 df) on keg and batch SD and U[0,12.25] on portion SD
ICCU	5.2.3	Beta(1,1), i.e. U[0,1], prior on $\frac{\text{variance component}}{\text{variance component + portion variance}}$ for each of keg and batch variance components. U[0,12.25] on portion SD.
IG	5.2.4	Vague Inverse Gamma priors on all stratum variances.
- IG0		IG(1,1)
- IG1		IG(0.1,0.1)
- IG2		IG(0.01,0.01)
- IG3		IG(0.001,0.001)
IIN	5.2.4	Mildly informative Inverse Gamma distribution prior on portion variance IG(1,12), and vague IG(0.001,0.001) on keg and batch stratum variances.
IαSm	5.2.4	Mildly informative Inverse Gamma prior on portion variance IG(1,12), and mildly informative Inverse Gamma priors on the keg and batch stratum variances. These are described in Table 5.2.
JEFF	5.2.5	Jeffreys' prior
FLAT	5.2.6	Uniform prior on variance components

5.1.2 Choice of Prior Parameters

For many of the prior families listed in Section 5.1.1 a choice of parameter value was required. In Section 4.1 it was suggested that it is known that the higher-level variances (σ_h^2) are likely to be within the range 0.5 to 24. On the SD metric this corresponds to a range 0.7 to 4.9. The parameters of the prior distributions were chosen with the aim that prior support was wider than this range, ensuring they were vague (low information priors). It was desired that the values outside the support of the prior or outside an interval containing a high percentage of the prior distribution were considered to be either infeasible or practically that there was no need to distinguish between a value on the boundary and one outside e.g. σ_h^2=24 is so large compared to $\sigma_{portion}^2$=6, that whether σ_h^2 is estimated as 24 or 30 makes no practical difference. An additional consideration was the likely size of the higher-level variances (σ_h^2) compared to the bottom-level or intrinsic variance ($\sigma_{portion}^2$). One reasonable assumption is that they are of comparable size. However, for analytical method precision studies a rule of thumb often used is that the higher-level sources of variation sum to approximately that of the bottom-level variance or intrinsic variance (i.e. intermediate precision is approximately double the

repeatability). The likely values chosen above align with this, though it is acknowledged that the batch sampling study is a different application. Thus the choice of some priors took this aspect into account.

For example, the upper limit (U) for the uniform prior (UNI) was chosen to be a multiple of a likely value for the SD, provided U also reasonably exceeded the required range discussed above. A multiplier of 5 was used since it was considered that it would make little practical difference to distinguish between values 5 or more times the most likely value. For $\sigma_{portion}$ the likely value is approximately $\sqrt{6}$ (=2.4) giving U=12.25. On the assumption that the sum of the two higher-level variances is similar to the bottom-level variance, the σ_h likely value is $\sqrt{3}$ (=1.7) and thus U=8.66. The limits for both the bottom-level SD and higher-level SD also adequately encompass the range 0.7 to 4.9 (the range required to cover values it is important to distinguish between as discussed above).

5.2 Details of the Priors to be Investigated

A summary of the sets of priors investigated was provided in Table 5.1. The priors are now described in more detail. In Section 5.3.3 the importance of the metric is discussed and thus some priors are plotted on both SD and variance metrics. The formula

$$f(\sigma^2) = \frac{1}{(2\sigma)} f(\sigma) \tag{5.1}$$

is used, which is derived via the Jacobian for a transformation of a density function - see Appendix C, Section C.1.

5.2.1 Uniform on SD (UNI)

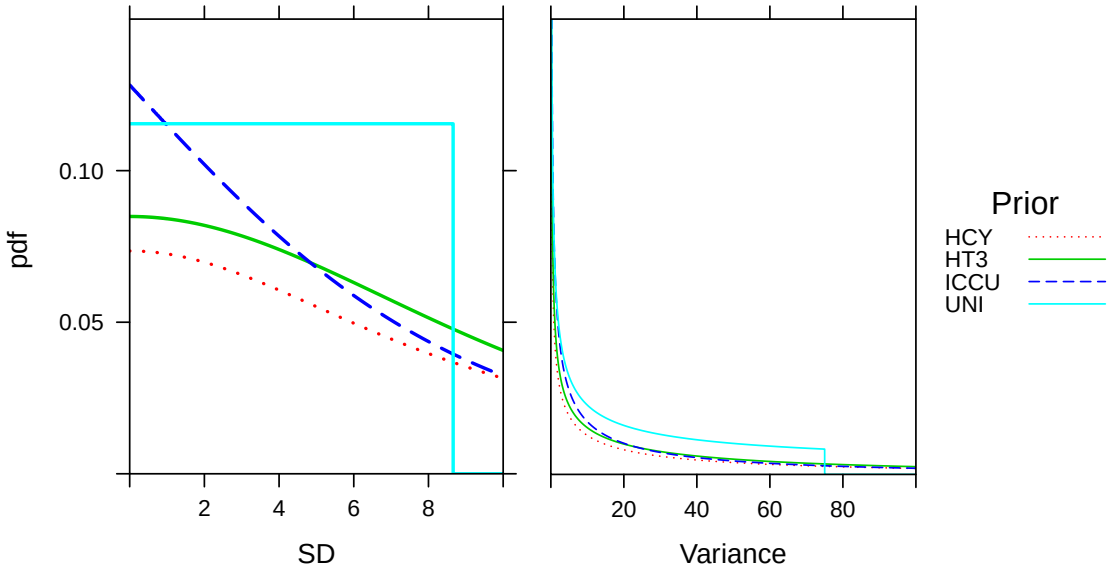
For the uniform prior on σ, the pdf is given by $f(\sigma) = 1/U$, $\sigma < U$, 0 otherwise (denoted by Uni[0,U]).

The upper limit (U) for the uniform prior was chosen to be a multiple of a likely value for the SD, provided this also reasonably exceeded the required interval for support of the prior. The choice of U[0,8.66] on the batch and keg SD and U[0,12.25] on the portion SD was discussed in Section 5.1.2. The pdf of the uniform prior on the SD for the higher-level random effects is plotted in Figure 5.1 on the SD and variance metrics.

5.2.2 Half-t Family on SD (HCY and HT3)

A non-standardised half-t distribution (a generalised version of the Student's t-distribution with a scale parameter and location parameter set to zero) is applied as a prior to the

FIGURE 5.1: **HCY, HT3, ICCU and UNI priors on σ_h^2 and σ_h metrics**



higher-level random effect standard deviations. The probability density function is given by:

$$f(\sigma_h) \propto \frac{1}{s}\left(1+\frac{1}{\nu}\left(\frac{\sigma_h}{s}\right)^2\right)^{(-(\nu+1)/2)}, \tag{5.2}$$

where s is the scale factor and ν is the degrees of freedom. The half-t restricts the distribution to $\sigma_h \geq 0$.

Two versions of the half-t prior are used. HCY has 1 degree of freedom (the half-Cauchy distribution) whilst HT3 had 3 degrees of freedom. The scale parameter was set to $5\sqrt{3}$ = 8.66, the value used for the upper limit of the uniform distribution in Section 5.2.1. For HCY this represents the half width of the distribution at half of the maximum and also represents the median value of the half-Cauchy distribution. For HT3 61% of the half-t distribution lies at smaller values illustrating its less heavy tail.

The priors are plotted in Figure 5.1 for the pdf on the SD metric and on the variance metric.

5.2.3 Ratio of Variances (ICCU)

For the "ratio" prior, a beta distribution is placed on the ratio of σ_h^2 to the sum of σ_h^2 and the "bottom-level" variance (the consistency intraclass correlation coefficient) i.e.

$r = \frac{\sigma_h^2}{\sigma_h^2+\sigma_p^2} \sim Beta(\alpha, \beta)$ for the batch sampling study. The probability density for the beta distribution is given by:

$$f(r) \propto \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \tag{5.3}$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha)+\Gamma(\beta)}$ and α and β are the shape parameters.

This prior aims to compare higher-level variance components with a "core" or intrinsic source of variability which in the example is the "bottom-level" variance. The $Beta(1,1)$ prior which corresponds to Uni[0,1] was investigated. The prior distribution on the ratio and Uni[0,12.25] prior on σ_p implies a resulting prior on the higher-level variance components. This prior is plotted in Figure 5.1 for the pdf on the SD metric and on the variance metric. The use of the beta distribution allows for incorporating a variety of shapes for the prior, though only the uniform prior is investigated in this research project due to the extensive time the simulations take to perform.

5.2.4 Inverse Gamma on Stratum Variances

5.2.4.1 Summary

Inverse gamma prior distributions, IG(α,β) (where α and β are the shape and scale parameters respectively) are placed on linear combinations of the variance components representing the expected mean squares from a traditional ANOVA table (and thus with coefficients usually corresponding to effective sample sizes). Note they are not directly placed on the individual higher-level variance components which is typically the case in the literature. The inverse gamma distribution has a probability density function $f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$. The transformations used for the example design described in Section 4.1 are: $\phi_p = \sigma_p^2$ (portion variance); $\phi_k = P\sigma_k^2 + \sigma_p^2$ (referred to as keg stratum variance); $\phi_b = KP\sigma_b^2 + P\sigma_k^2 + \sigma_p^2$ (referred to as batch stratum variance). The inverse gamma prior distributions are placed independently on the three stratum variances. As will be described in Sections 5.3 and 6.4, though the priors as described could allow negative variance component values, the implementation ensures that posterior samples of variances are non-negative as aligned with the model given in Equation (4.1).

Three groups of inverse gamma priors were investigated:

- **Vague priors on all stratum variances** Four were investigated where $\alpha = \beta =$ 1, 0.1, 0.01 or 0.001 (referred to as prior sets **IG0, IG1, IG2, IG3** respectively, the number indicating the decimal place of the "1"). The priors on the stratum variances are plotted in Figure 5.2.

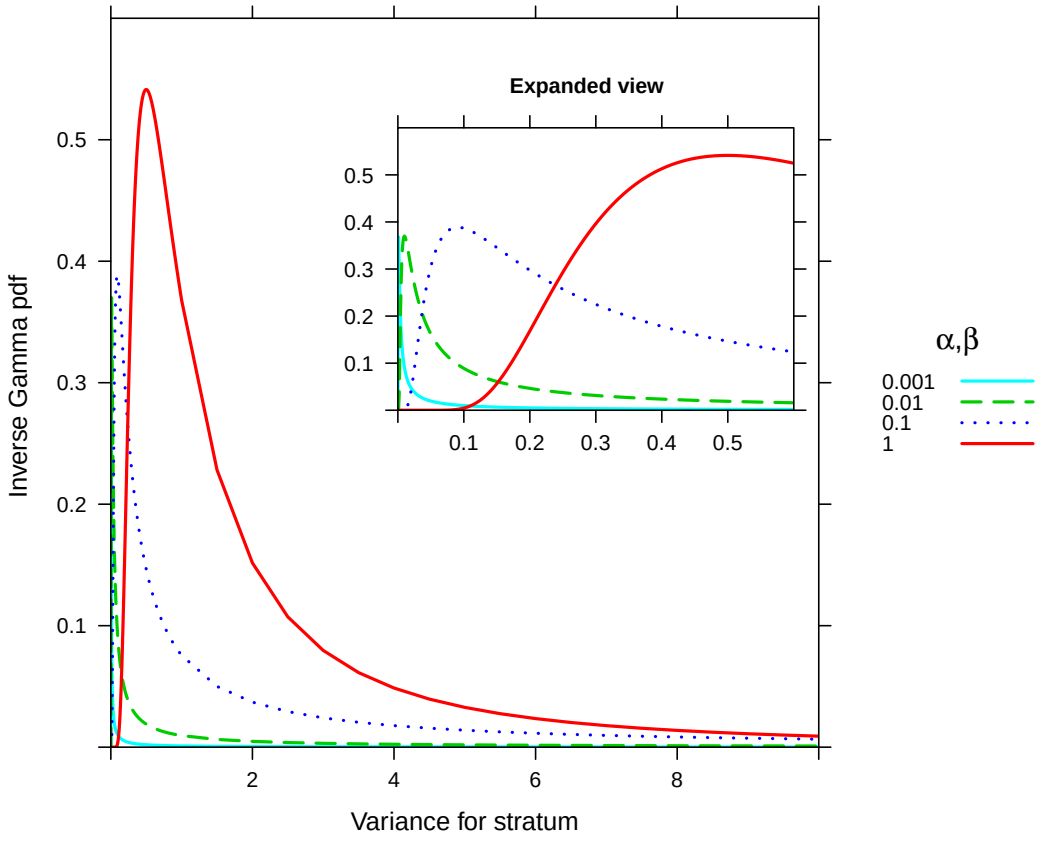
- **Vague priors on most stratum variances with a mildly informative prior on an intrinsic stratum variance level** It is often the case that there will be some knowledge about an intrinsic level of variance - often the bottom level. Here a mildly informative prior IG(1,12) was chosen for $\sigma^2_{portion}$ which gives a distribution with a mode of 6. The choice of β and α is discussed in Section 5.2.4.2. The IG(1,12) prior on $\sigma^2_{portion}$ together with IG(0.001,0.001) on the keg and batch stratum variances are denoted the **IIN** prior.
- **Mildly informative priors on all stratum variances** Priors intended to be mildly informative on the higher-level stratum variances were chosen to assess their performance together with IG(1,12) on $\sigma^2_{portion}$. Inverse Gamma distributions were chosen where their mode or median equalled those for the keg or batch stratum variances corresponding to likely values for σ^2_{keg} or σ^2_{batch} of 3 or 6. As the priors are placed on the stratum variances they are dependent on the design. The priors investigated are listed in Table 5.2 where the prior abbreviation IαSm is constructed by including a character for α which represents the scale parameter of the IG prior, S represents whether the mode (S=O) or median (S=E) statistic is set to the likely value and m represents the likely value. The choice of β and α are discussed in Section 5.2.4.2.

TABLE 5.2: **Mildly informative IG distributions for batch sampling design with 2 kegs/batch and 16 portions/keg**

Abbrev.	σ^2_k	σ^2_b	ϕ_k	ϕ_b	Statistic	α_k	β_k	α_b	β_b
I2O3	3	3	54	150	Mode	2	162	3	600
I2O6	6	6	102	294	Mode	2	306	3	1176
I1O3	3	3	54	150	Mode	1	108	1	300
I1O6	6	6	102	294	Mode	1	204	1	588
I1E3	3	3	54	150	Median	1	37.4	1	103.97
I1E6	6	6	102	294	Median	1	70.7	1	203.97
IhE6	6	6	102	294	Median	0.5	23.2	0.5	66.88

All three of the above sets of priors were investigated as, though SAS software documentation (SAS Institute Inc.: The MIXED Procedure, 2013) indicates that the DATA= option (through which inverse gamma prior distributions are specified) provides an opportunity to specify an informative prior, it was wished to also investigate the use of priors which may be vague.

FIGURE 5.2: **Vague inverse gamma priors, IG(α,β), $\alpha = \beta$, placed on $\sigma^2_{portion}$ and stratum variances**



5.2.4.2 Choice of β and α

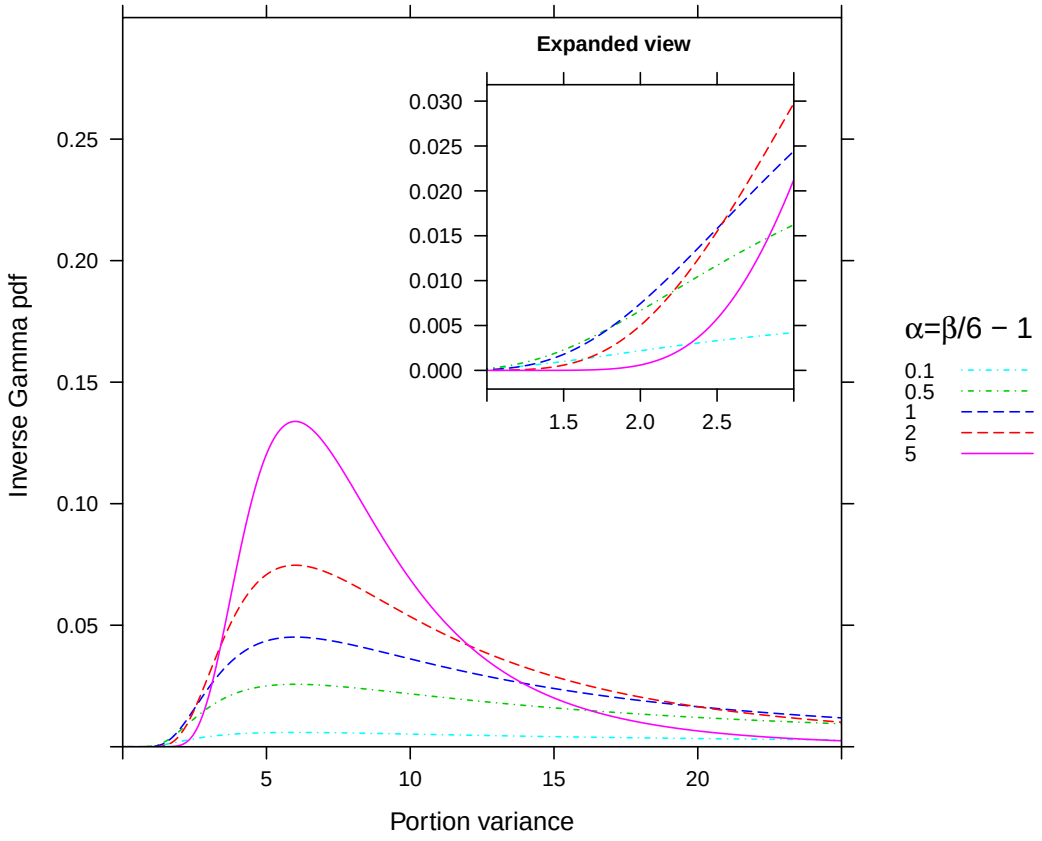
Mildly informative prior IG(α,β) on portion variance It is often the case that there will be some information about $\sigma^2_{portion}$. Given that the most likely value for $\sigma^2_{portion}$ is 6, it seems reasonable to choose an inverse gamma distribution with a mode of 6. The mode of the inverse gamma density is at $\frac{\beta}{\alpha+1}$. Figure 5.3 shows various inverse gamma priors with a mode of 6. A choice of α=1 for the inverse gamma prior seems reasonable since it places less weight on values very close to 6 than those with larger α values whilst still giving greater probability to values between 1.5 and 3 compared to smaller α values. It is noted that this prior is informative with respect to small values as it gives very low probability for values less than 1.5 (as do all the priors with a mode of 6). However, in the situation where there is information on $\sigma^2_{portion}$ it seems reasonable that existing information would exclude a true value less than one quarter of the most likely value and that departures from what is expected will result in greater, not smaller, variance being observed. Thus the mildly informative prior chosen for $\sigma^2_{portion}$ was IG(1,12). This together with IG(0.001,0.001) on the keg and batch stratum variances are denoted the IIN prior.

Mildly informative priors IG(α,β) on batch and keg stratum variances For the batch sampling design with 2 kegs per batch and 16 portions per keg (K=2 and P=16) the most likely values for the keg and batch stratum variances would be 102 ($\phi_k = P\sigma^2_k + \sigma^2_p$) and 294 ($\phi_b = KP\sigma^2_b + P\sigma^2_k + \sigma^2_p$) respectively if the likely values for all variances of random effects are 6. If the likely value for the portion variance is 6 and those for the higher-level variances are 3 the likely values for the stratum variances are 54 and 150 respectively. The priors investigated are listed in Table 5.2 where the prior abbreviation IαSm is constructed by including a character for α which represents the scale parameter of the IG prior, S represents whether the mode or median statistic is set to the likely value and m represents the likely value. The mode of the inverse gamma density is at $\frac{\beta}{\alpha+1}$ and the median is given by solving

$$0.5 = \frac{\Gamma(\alpha, \frac{\beta}{\text{median}})}{\Gamma(\alpha)},$$

where the right hand side of the formula is the cumulative distribution function and Γ is the gamma or incomplete gamma function. ϕ_h is the value of the h stratum variance if its constituent variances are set to their mode or median value (h = b or k) and α_h and β_h represent the IG parameter values which achieve a mode or median for the corresponding stratum variance. Variances of the higher-level random effects within the range 0.5 to 24 for the original study correspond to a range of 8.5 to 408 for the keg stratum variance and a range of 24.5 to 1176 for the batch stratum variance.

FIGURE 5.3: **Example mildly informative inverse gamma priors with mode of 6 on $\sigma^2_{portion}$**



The inverse gamma priors on the keg stratum variance are shown in Figure 5.4 and those on the batch stratum variance are shown in Figure 5.5.

FIGURE 5.4: **Mildly informative inverse gamma prior on keg stratum variance**

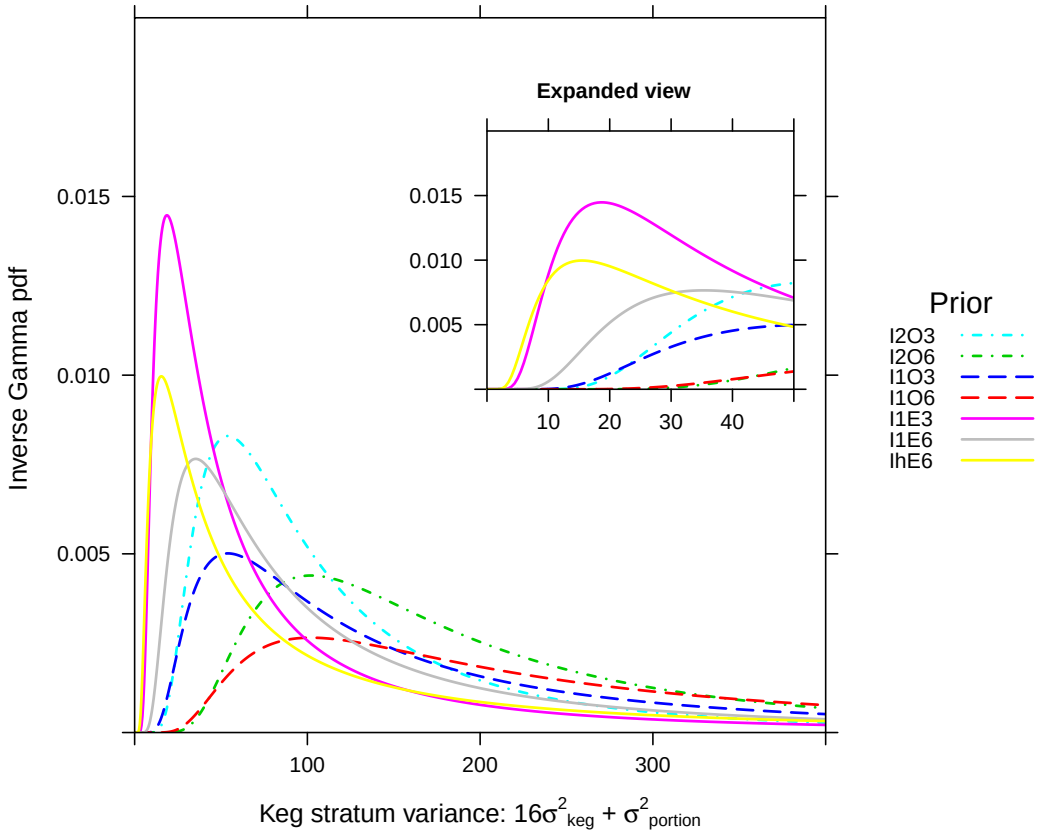
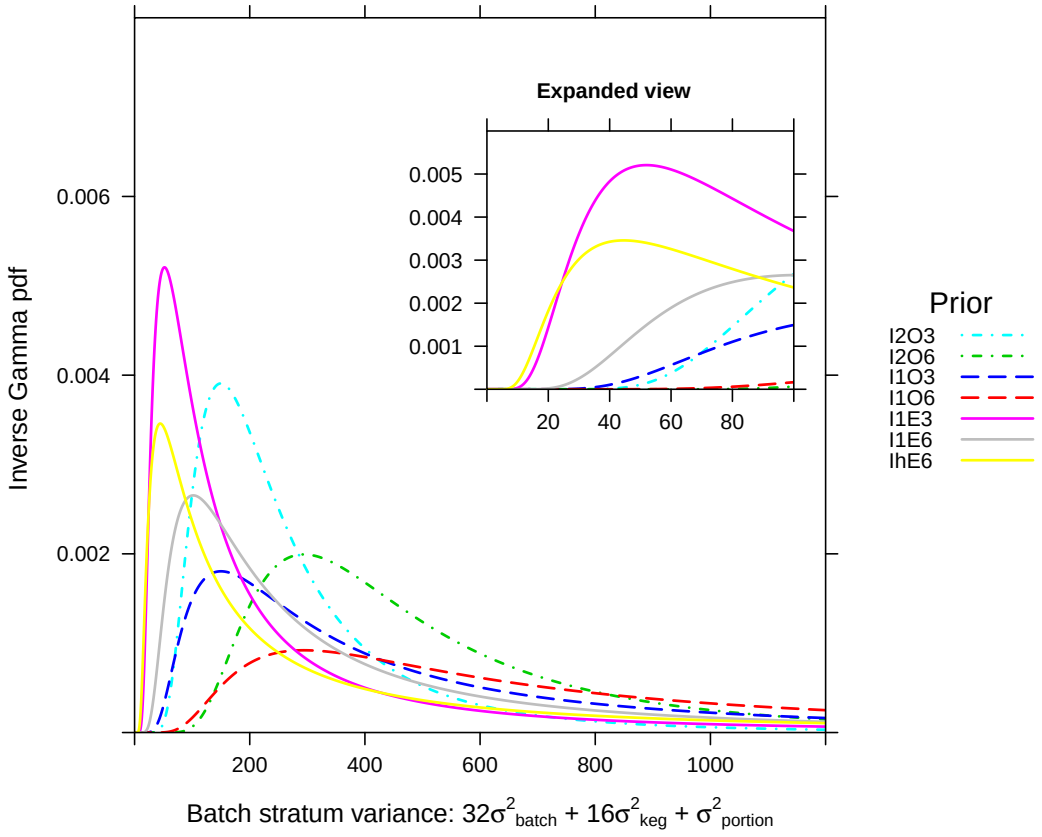


FIGURE 5.5: **Mildly informative inverse gamma prior on batch stratum variance**



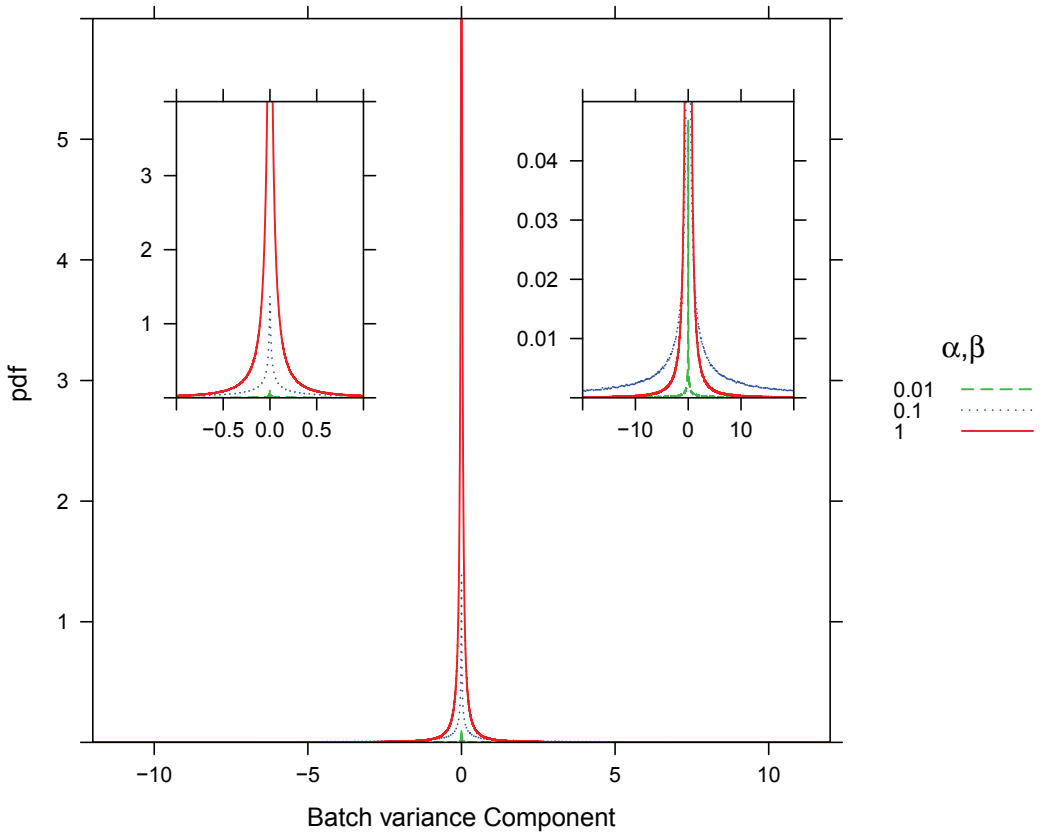
5.2.4.3 Implied Marginal Distribution on Higher-Level Variance Components

The independent prior distributions on the stratum variances imply resultant marginal distributions on the higher-level variance components. For the vague IG priors placed on all stratum variances (ϕ_p, ϕ_k, ϕ_b) and, for the moment assuming an unrestricted model (see Section 1.3), the resultant marginal distributions were obtained for the original batch sampling design with 6 batches, 2 kegs/batch and 16 portions/keg. Figure 5.6 shows the implied prior marginal distributions on the batch variance component for the various priors and Figure 5.7 shows the marginal distributions on the keg variance component (to aid viewing expanded views of parts of the distributions are inset). We then wished to show the marginal distributions on the standard deviation metric. A difficulty occurred when trying to find the distribution on σ_{batch}. Since the Inverse Gamma prior on linear combinations of variance components only constrains the stratum variances to be positive, it allows negative values for individual variances e.g. σ^2_{batch}. When σ^2_{batch} is negative, σ_{batch} will be undefined. In the simulations used to obtain the density, when $\sigma^2_{batch} < 0$, then σ_{batch} was set to $-\sqrt{-\sigma^2_{batch}}$ for the purpose of plotting. The marginal distribution on the batch component on the SD metric (σ_{batch}) given this convention is shown in Figure 5.8. The marginal distributions when IG(0.001,0.001) is placed on stratum variances are not plotted due to difficulties obtaining the density from simulated values since IG(0.001,0.001) is so heavy tailed. On the graph it would appear flat if the density had been obtained. Placing the same vague inverse gamma on all the strata seems to invoke a very informative prior for the keg or batch variance components if they are very close to zero. However, the steepest part of the density curve is for values within ± 0.5 i.e. outside the main range of interest. Also if the data is such that the bottom level $\sigma^2_{portion}$ is well estimated, this apparent informative prior marginal distribution for the batch and keg variance components may not in practice substantially affect the posterior.

For the vague IG priors placed on the higher-level stratum variances (ϕ_k, ϕ_b) and IG(1,12) on $\sigma^2_{portion}$, Figure 5.9 shows the implied prior marginal distributions on the keg variance component assuming an unrestricted model for the original batch sampling design. The implied marginal distributions on the batch variance component are the same as in Figure 5.6 given that ϕ_b is a linear combination of σ^2_{batch} and ϕ_k.

For the mildly informative IG priors placed on the higher-level stratum variances (ϕ_k, ϕ_b) and IG(1,12) on σ^2_p, Figure 5.10 shows the implied prior marginal distributions on the batch variance component and Figure 5.11 shows those for the keg variance component assuming an unrestricted model for the original batch sampling design. It is seen that, for the chosen priors, placing the prior on the stratum variances results in negative values for the higher-level variance component estimates not being excluded by the priors. When vague priors were placed on the higher level variance component

FIGURE 5.6: **Marginal distribution for σ^2_{batch} given vague priors on ϕ_b and ϕ_k (with inset expanded views of the distributions)**



stratum variances, about half of the marginal distribution for the higher-level variance components has negative values, whereas with the more informative priors the majority of the distribution takes positive values. One approach to the issue of negative values for the higher-level variance component estimates not being excluded by the priors, could be to allow negative variance component estimates (the unrestricted model described in Section 1.3). Another is to restrict variance component estimates to be non-negative (the restricted model described in Section 1.3). Given that we expect any higher-level variance components to be truly ≥ 0 the latter approach is taken. The posterior sampling algorithm rejects any sample from the posterior which has a negative estimate for a variance component. This is further discussed in Sections 5.3.3 and 6.4.1.1.

FIGURE 5.7: **Marginal distribution for σ^2_{keg} given vague priors on ϕ_k and $\sigma^2_{portion}$ (with inset expanded views of the distributions)**

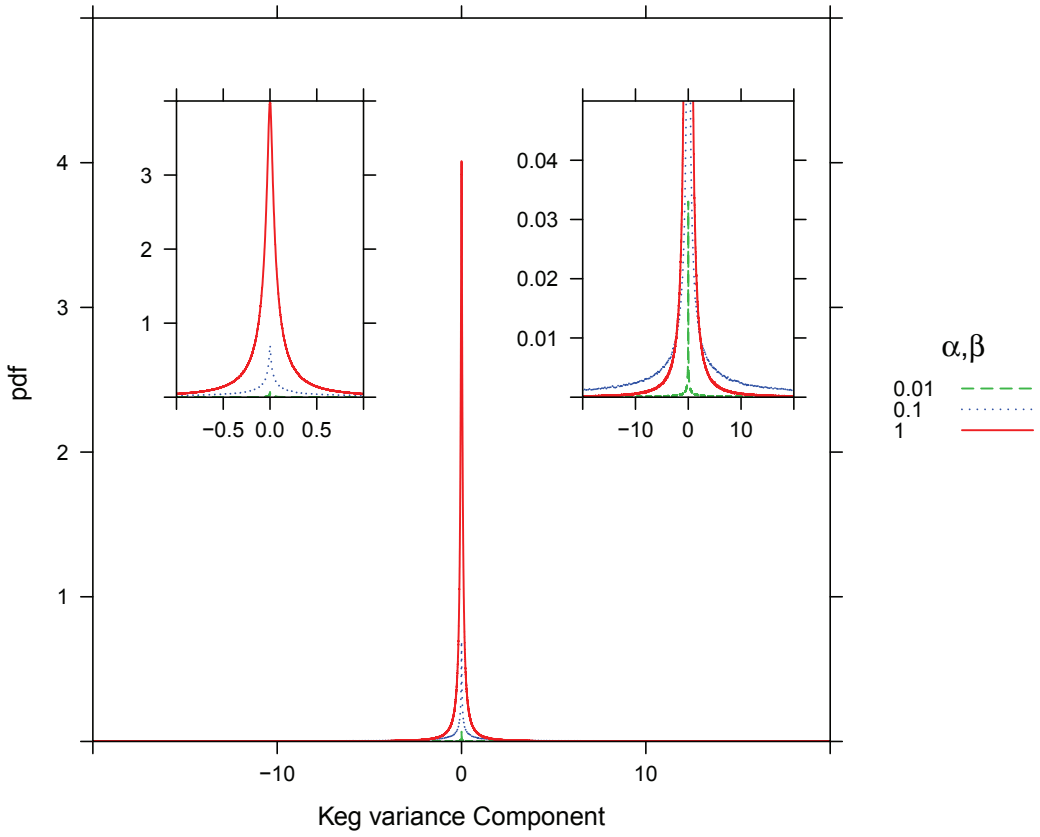


FIGURE 5.8: **Marginal distribution for σ_{batch} given vague priors on ϕ_b and ϕ_k (with inset expanded views of the distributions)**

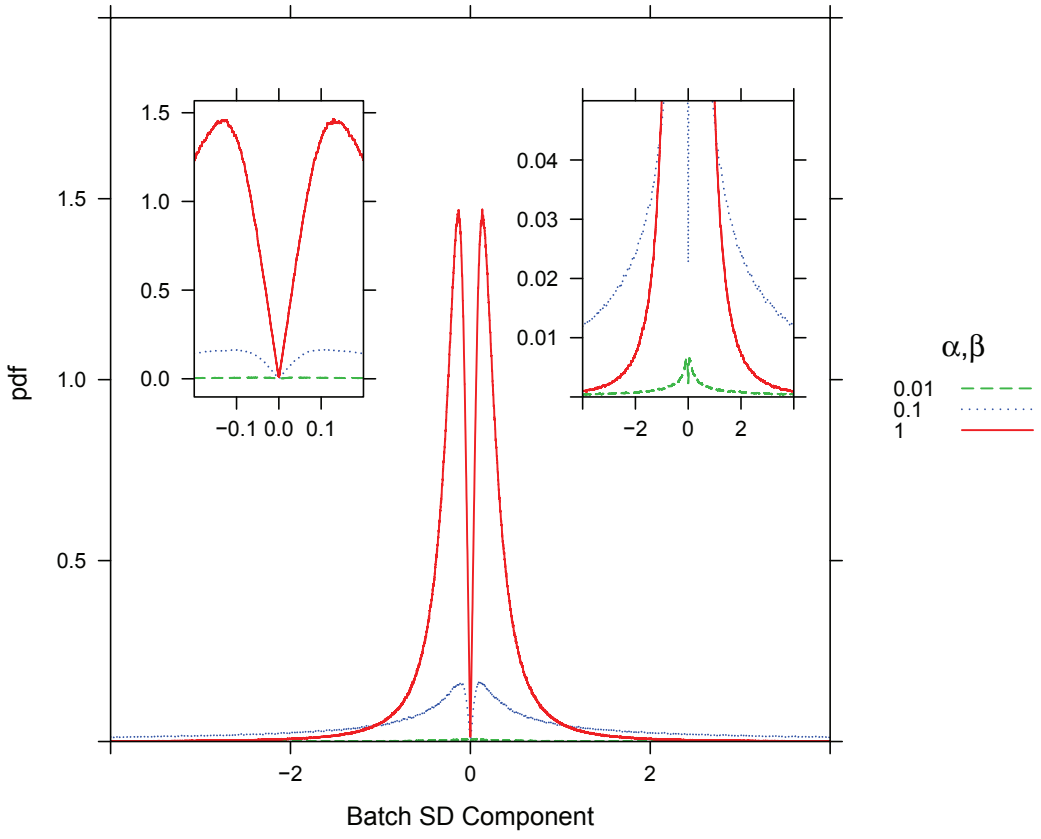


FIGURE 5.9: **Marginal distribution for σ^2_{keg} given vague priors on ϕ_k and IG(1,12) on $\sigma^2_{portion}$**

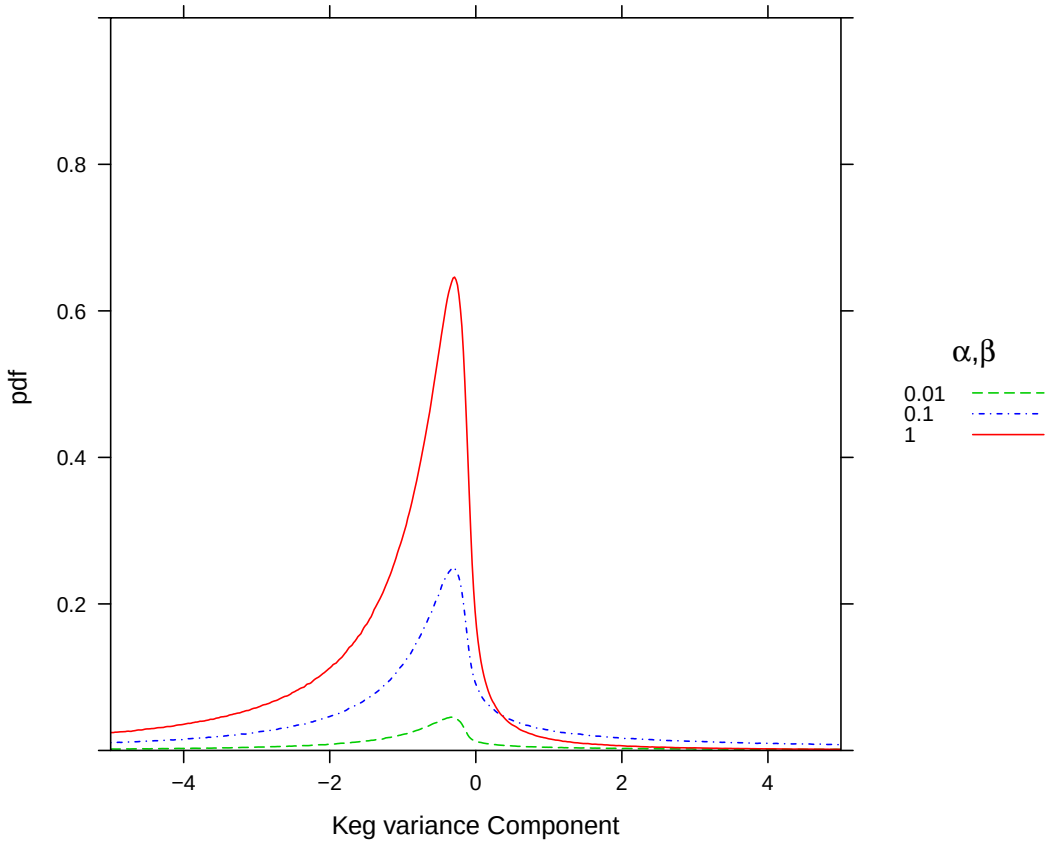


FIGURE 5.10: **Marginal distribution for σ^2_{batch} given mildly informative IGs on ϕ_b and ϕ_k**

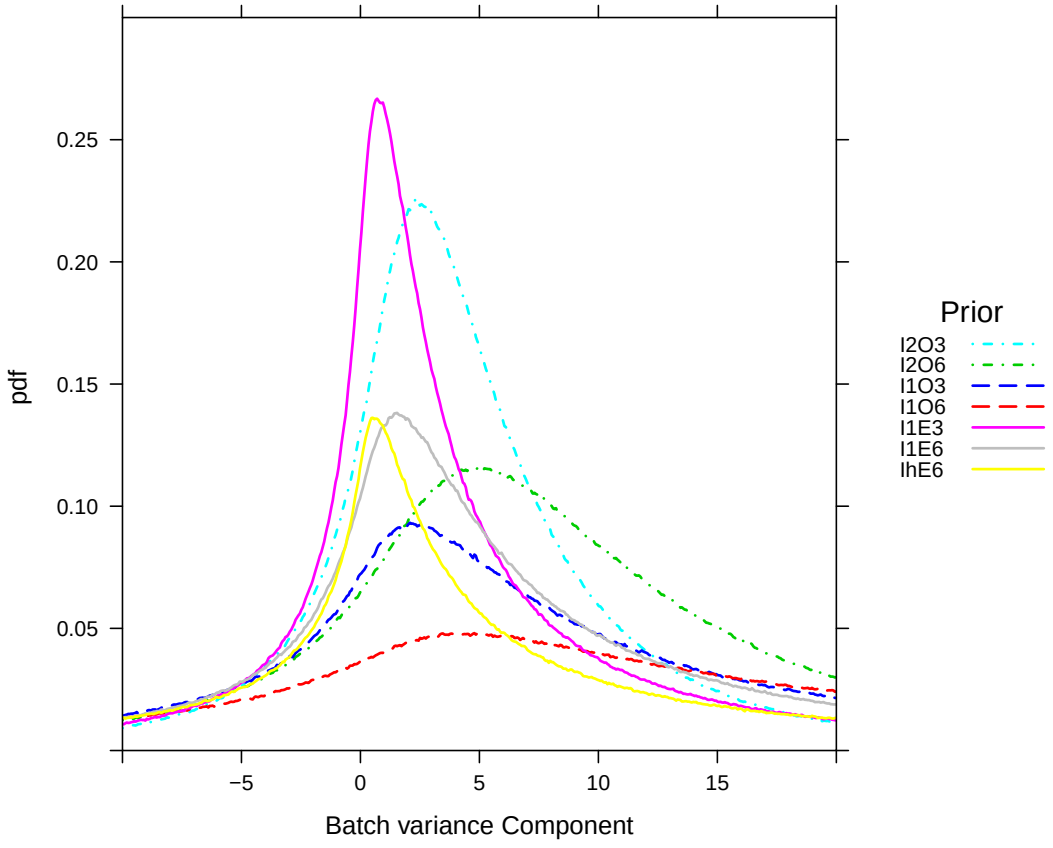
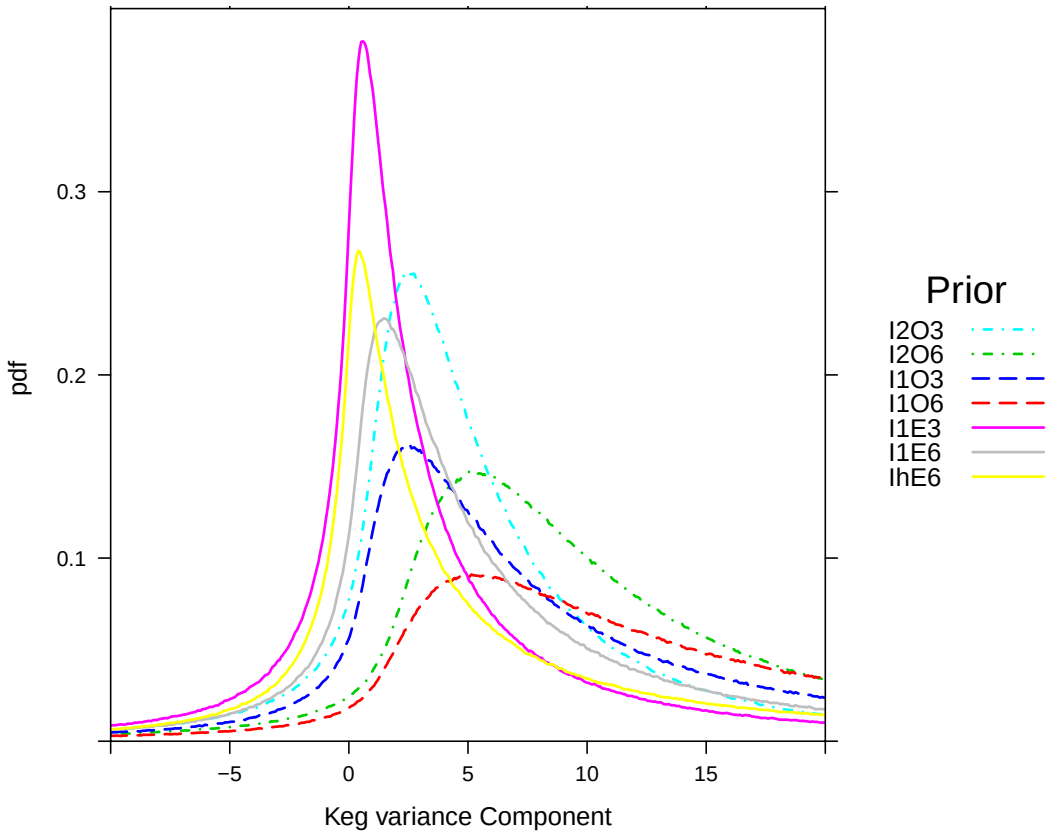


FIGURE 5.11: **Marginal distribution for σ^2_{keg} given mildly informative IGs on ϕ_k and IG(1,12) on $\sigma^2_{portion}$**



5.2.5 Jeffreys' Prior (JEFF)

5.2.5.1 Jeffreys' Priors for Variance Components Models

Jeffreys (1946) aimed to provide a rule for finding an non-informative prior for any parametric model. Jeffreys' prior was designed by Jeffreys (1946) to be invariant under one to one parameter transformations. Using the notation of Kass and Wasserman (1996) Jeffreys' prior is given by $\pi_\theta(\theta) \propto det(\mathbf{I}(\theta))^{1/2}$ where the Fisher information $\mathbf{I}(\boldsymbol{\theta})$ is

$$\mathbf{I}(\theta)_{ij} = E\left(-\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\right), \tag{5.4}$$

where l = log likelihood.

Jeffreys proposed that location and scale parameters should be treated separately as, though the information matrix would generally include both location and scale effects, doing this leads to priors on the scale parameters which are considered undesirable (Jeffreys, 1946; Kass and Wasserman, 1996).

The derivation of Jeffreys' priors for the simple normal distribution, one-way and two-way nested variance components models are given in Appendix D, Sections D.1, D.2.1 and D.3.1 respectively.

For a one-way balanced variance components model with measurements taken from P portions from each of K kegs from one batch with two variance components (between keg variance component σ^2_{keg} and between portions variance $\sigma^2_{portion}$) Jeffreys' prior is given by:

$$\pi(\sigma_k^2, \sigma_p^2) = \sqrt{|I(\sigma_k^2, \sigma_p^2)|} \propto \frac{1}{\sigma_p^2(\sigma_p^2 + P\sigma_k^2)}. \tag{5.5}$$

For a two-way balanced variance components model with measurements taken from P portions from each of K kegs from each of B batches with three variance components (σ^2_{batch}, σ^2_{keg}, $\sigma^2_{portion}$) Jeffreys' prior is given by:

$$\pi(\sigma_b^2, \sigma_k^2, \sigma_p^2) = \sqrt{|I(\sigma_b^2, \sigma_k^2, \sigma_p^2)|} \propto \frac{1}{\sigma_p^2(\sigma_p^2 + P\sigma_k^2)(\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2)}. \tag{5.6}$$

It is derived from the determinant of the information matrix for the balanced two-way variance components model given by Searle (1970), p.514 (see Section D.3). The design of the original batch sampling study has B=6, K=2, P=16 (see Section 4.1).

5.2.5.2 Exploration of Distribution of Jeffreys' Priors

For comparison with other priors it was desired to find the density function of the Jeffreys' prior and/or look at the marginal distributions for σ^2_{keg} and σ^2_{batch}.

In Appendix D, Sections D.2.2 and D.3.2 it is shown that the marginal distributions for σ^2_{keg} for Jeffreys' priors for either the one-way or the two-way nested variance components model are undefined. The integral of the prior distributions over the range of support values is undefined/infinite and thus the prior is improper. Jeffreys' priors conditioned on $\sigma^2_{portion}$ were also investigated and found to be undefined. It was also attempted to look at the distributions $\pi\left(\frac{\sigma^2_k}{\sigma^2_p}, \sigma^2_p\right)$ and $\pi\left(\frac{\sigma^2_b}{\sigma^2_p}, \frac{\sigma^2_k}{\sigma^2_p}, \sigma^2_p\right)$ but these were also undefined.

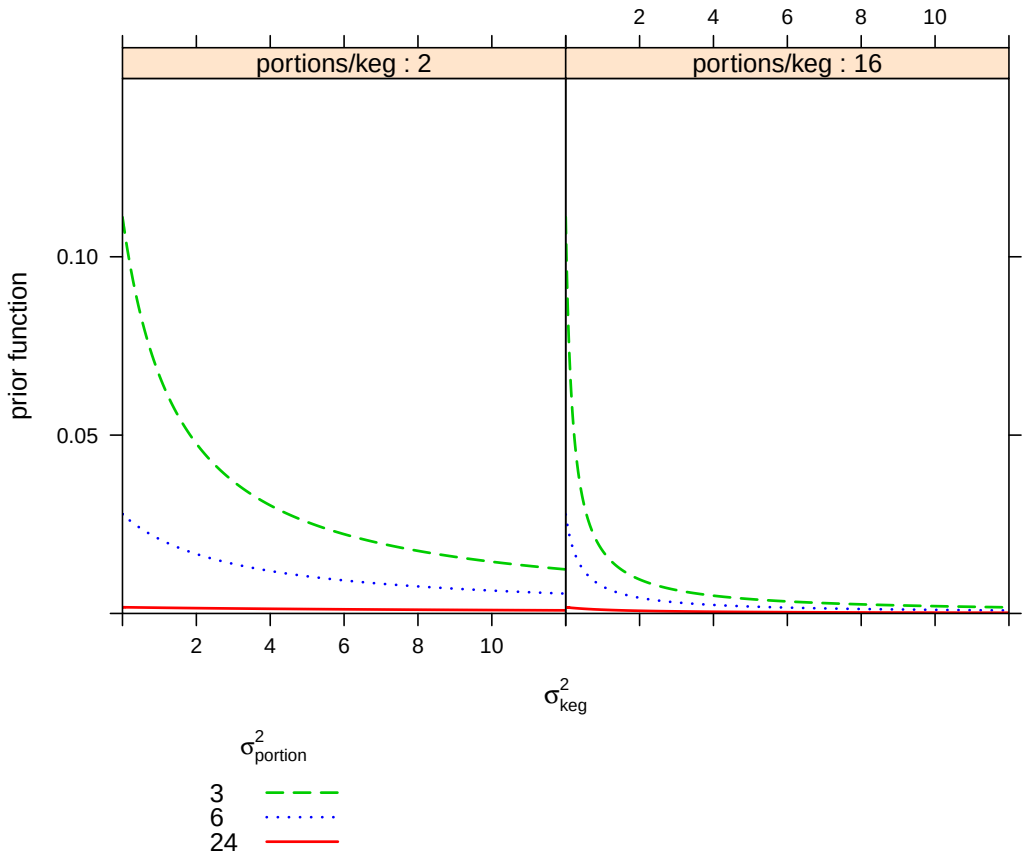
One avenue for exploring the distribution of Jeffreys' prior is to examine joint prior distributions $\pi(\sigma^2_k, \sigma^2_p)$ and $\pi(\sigma^2_b, \sigma^2_k, \sigma^2_p)$. However, these are improper and thus comparisons with other priors will be difficult. An alternative is to examine the distribution subject to σ^2_b or $\sigma^2_k \leq$ U where U is sufficiently large to cover values of interest.

Though the study on which the evaluations are to be performed is a two-way variance components design, the one-way variance components design is explored first to get a clearer understanding of the distribution implied by the Jeffreys' prior.

One-way variance components design The joint prior function for σ^2_{keg} and $\sigma^2_{portion}$ for the one-way variance components design is plotted in Figure 5.12. It is plotted for 16 portions/keg (as in the original batch sampling study design) and 2 portions/keg. It is plotted for $\sigma^2_{portion}$=6 (the value it is known to be around - see Section 4.1) and half and double this value. Values of σ^2_{keg} up to at least 24 are of interest but for clarity in the region where there is most change in the function σ^2_{keg} is only plotted to 12. From Equation (5.5) it is seen that the Jeffreys' prior function does not depend on the number of kegs and that the y intercept is independent of the number of portions/keg. The joint prior function is improper. As mentioned above, an alternative way to examine the function is to plot the density subject to $\sigma^2_{keg} \leq U_k$ where U_k is a large enough value that all reasonable values are included. This is calculated using Equation (D.4) and shown in Figure 5.13 for U_k=10,000.

It is seen from Figures 5.12 and 5.13 that the impact of the prior on the posterior will be to pull the posterior estimate of σ^2_{keg} closer to zero in the cases where the true value is non-trivially above zero (note if the maximum likelihood or REML estimate is 0 or very close then due to the miscalibration effect the posterior estimate may be larger). As $\sigma^2_{portion}$ decreases the curve becomes steeper and thus the impact is to pull the estimate of σ^2_{keg} even closer to zero. This is likely to be reasonable with the variation due to higher level sources being expected to be smaller if the bottom level source of variation is smaller. Jeffreys' prior thus provides a way of scaling the prior for the higher level source of variance to the bottom level source. Increasing the number of portions/keg from 2 to 16 results in the curve becoming steeper and thus the impact of the pull of the estimate of σ^2_{keg} towards zero will be greater with larger numbers of portions/keg. This is less reasonable for a prior intended to be vague.

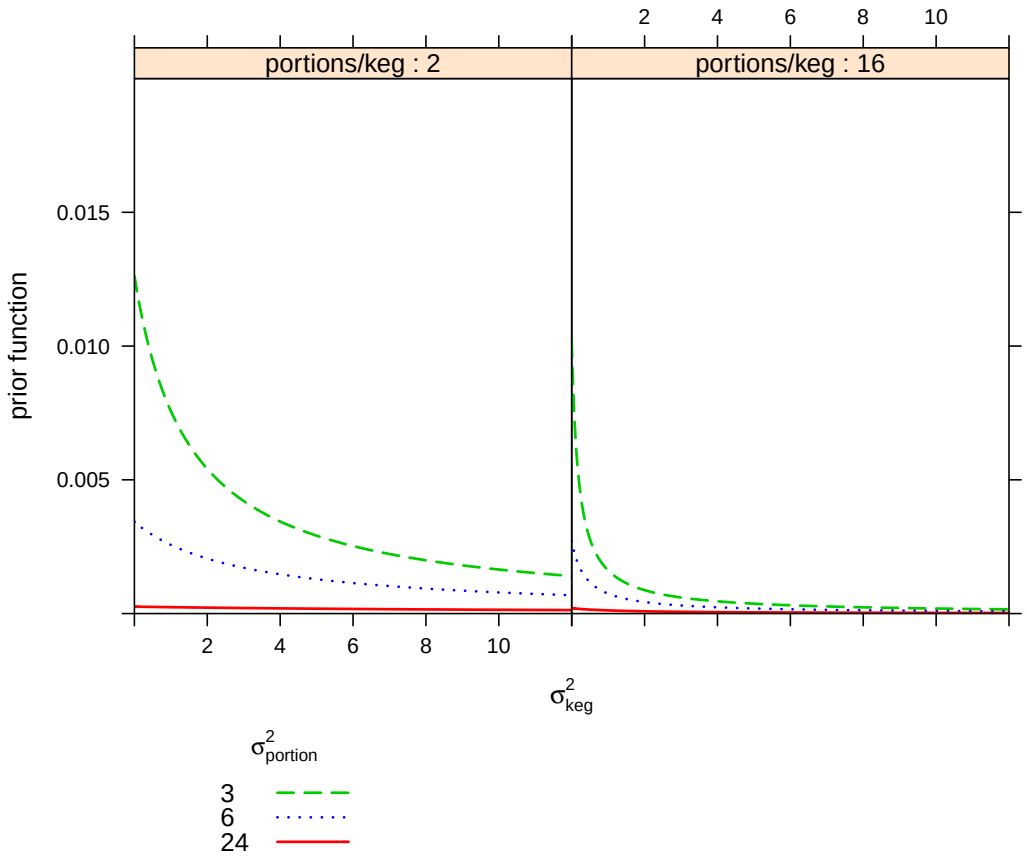
FIGURE 5.12: **Jeffreys' prior for σ^2_{keg} for varying $\sigma^2_{portion}$ for 2 and 16 portions/keg (1-way)**



If the design was used to estimate an overall mean and it was anticipated that the keg variance component was much smaller than the portion variance then it would make sense to provide much more replication at the portion level and thus the prior will be aligned with that prior knowledge. However, the numbers of kegs and portions/keg may well be due to the practicalities of the situation and design and thus this is not a reasonable assumption. Indeed in the original batch sampling study the reason for the large number of portions per keg was to be able to demonstrate that there were not spatial effects within the keg.

Two-way variance component design Jeffreys' prior for the two-way nested variance components design is now examined. The prior function is improper. The marginal distributions for σ^2_{keg} and σ^2_{batch} are undefined as is the distribution conditioned on $\sigma^2_{portion}$. Given this, the prior function will be examined in a similar way to the one-way variance components design by plotting the density subject to $\sigma^2_{batch} \leq U_b$ where U_b is a large enough value that all reasonable values are included. In addition $\sigma^2_{portion}$ will be

FIGURE 5.13: **Jeffreys' prior for σ^2_{keg} for varying $\sigma^2_{portion}$ for 2 and 16 portions/keg, given $\sigma^2_{keg} \leq$ 10,000 (1-way)**



set to a fixed value M_p. A value of M_p=6 is used which is the value $\sigma^2_{portion}$ is known to be around - see Section 4.1.

The distribution $\pi(\sigma^2_b, \sigma^2_k | \sigma^2_p = M_p, \sigma^2_b \leq U_b)$ is defined and is given by:

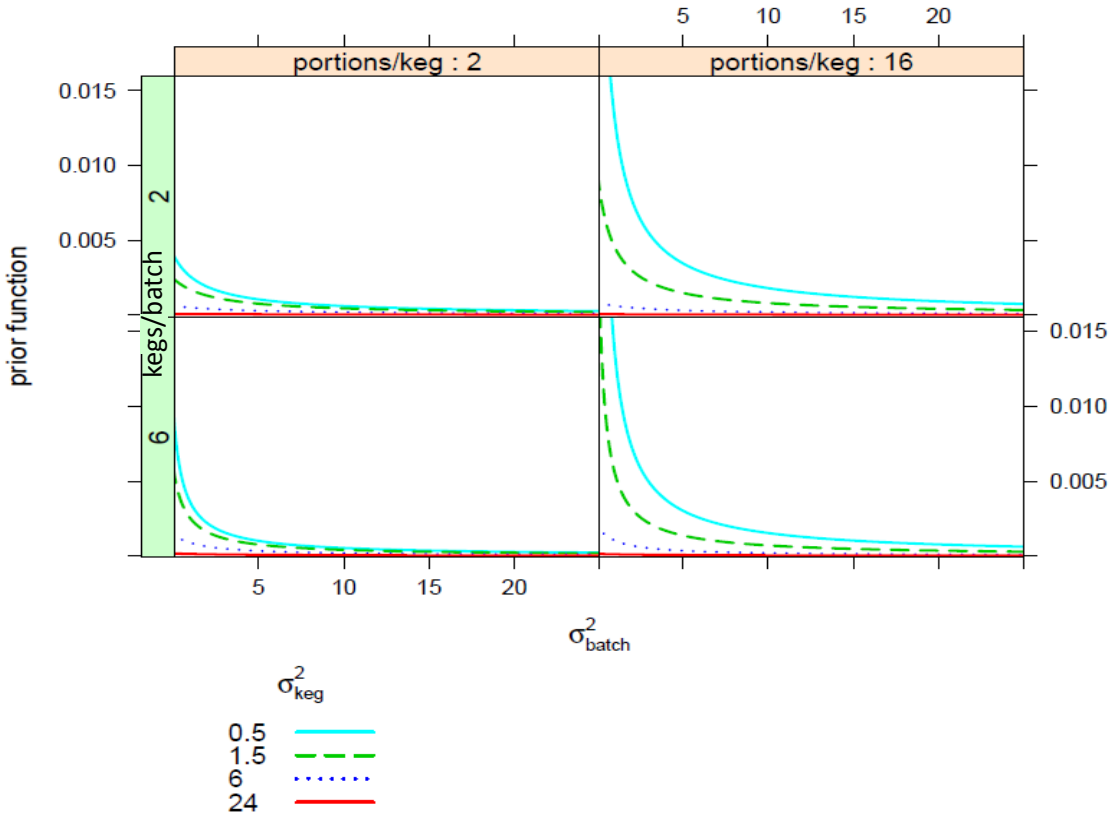
$$\pi(\sigma^2_b, \sigma^2_k | \sigma^2_p = M_p, \sigma^2_b \leq U_b) = \frac{\frac{1}{M_p(M_p+P\sigma^2_k)(M_p+P\sigma^2_k+PK\sigma^2_b)}}{KU_b\left[\frac{Li_2(-\frac{1}{z})}{a}\right]^{\infty}_{L_z}}, \tag{5.7}$$

where M_p is the chosen value of σ^2_p, $z = \frac{M_p+P\sigma^2_k}{PKU_b}$, $a = M_p P^2 K^2 U_b$, $L_z = \frac{M_p}{PKU_b}$ and the poly logarithm function $Li_2(x) = \Sigma^{\infty}_{m=1} x^m/m^2$. See Appendix D, Section D.3.3 for its derivation.

Thus the distribution of $\pi(\sigma^2_b, \sigma^2_k | \sigma^2_p = M_p, \sigma^2_b \leq U_b)$ is used to investigate Jeffreys' prior. The prior is plotted on the variance scale in Figure 5.14 and on the standard deviation

scale in Figure 5.15 for $M_p = 6$ and U_b=10,000. It is plotted for a number of scenarios for the design (number of kegs/batch and number of portions/keg) and for varying σ^2_{keg} and σ^2_{batch}.

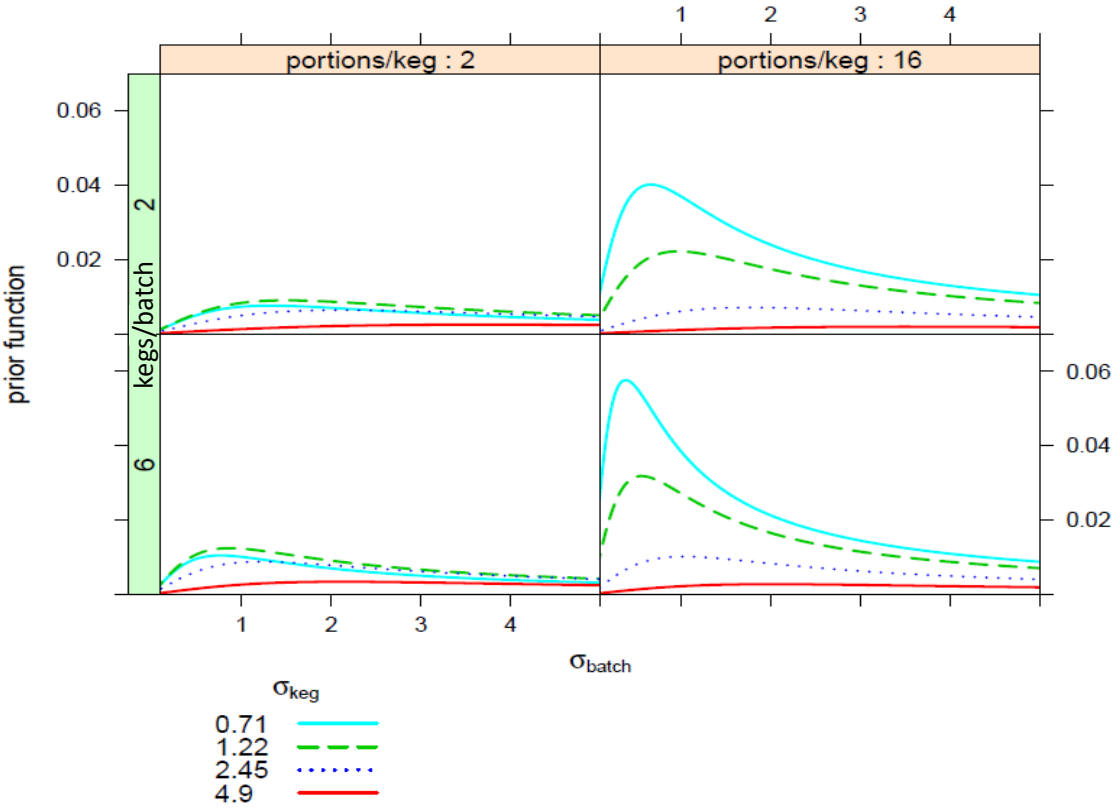
FIGURE 5.14: **Jeffreys' joint prior for $\sigma^2_{batch}, \sigma^2_{keg}$, for $\sigma^2_{portion} = 6$, given $\sigma^2_{batch} <$ 10,000, for various kegs per batch and portions per keg**



On the variance metric for each scenario it is seen that the density is highest as $\sigma^2_{keg} \to 0$ or $\sigma^2_{batch} \to 0$. However, on the standard deviation metric it is seen that the highest density is at small but not zero values of σ_{keg} and σ_{batch}.

It is seen from Figure 5.14 that the impact of the prior on the posterior will be to pull the posterior estimate of σ^2_{batch} closer to zero in the cases where the true value is non-trivially above zero. As σ^2_{keg} decreases the curve becomes steeper and thus the impact is to pull the estimate of σ^2_{batch} even closer to zero. In some applications this may be reasonable with the variation due to one higher-level source being expected to be smaller if other higher-level sources of variation are small. However, in other applications this is likely to be unreasonable. In the batch sampling application variability between kegs from the same batch is likely to be due to different factors from those which affect variability between batches and thus this seems a less reasonable assumption.

FIGURE 5.15: **Jeffreys' joint prior for $\sigma_{batch}, \sigma_{keg}$, for $\sigma^2_{portion} = 6$, given $\sigma^2_{batch} <$ 10,000 for various kegs per batch and portions per keg**

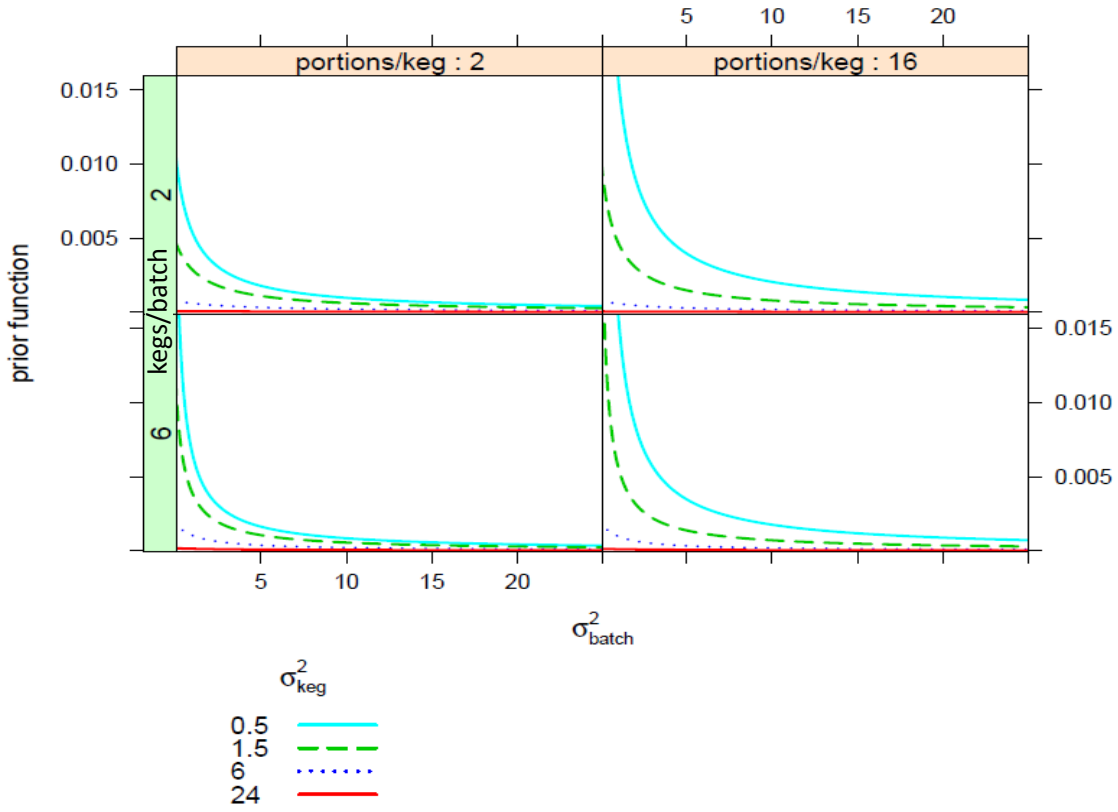


It is also seen that the design affects the prior. Increasing the number of portions/keg from 2 to 16 results in the curve on the variance metric becoming steeper for both σ^2_{batch} and σ^2_{keg} and thus the impact of the pull of the estimates towards zero will be greater for designs with larger numbers of portions/keg. Similarly the impact of the pull of the σ^2_{batch} and σ^2_{keg} estimates towards zero is stronger with designs with 6 kegs/batch compared to those with 2 kegs/batch. Whilst the density was based on $\sigma^2_b \leq 10000$) similar patterns were seen if U_b was varied.

The distribution of $\pi(\sigma^2_b, \sigma^2_k | \sigma^2_p = M_p, \sigma^2_b \leq U_b)$ for $M_p = 3$ and 12 is plotted on the variance scale in Figures 5.16 and 5.17 respectively for $U_b = 10,000$.

Similarly to the one-way variance components design it is seen from Figures 5.14, 5.16 and 5.17 that as $\sigma^2_{portion}$ decreases the curve becomes steeper and thus the impact is to pull the estimate of σ^2_{batch} or σ^2_{keg} even closer to zero. This is likely to be reasonable with the variation due to higher level sources being expected to be smaller if the bottom level source of variation is smaller and that bottom level source is intrinsic to the variation. Jeffreys' prior thus provides a way of scaling the prior for the higher level source of variance to the bottom level source.

FIGURE 5.16: **Jeffreys' joint prior for $\sigma^2_{batch}, \sigma^2_{keg}$, for $\sigma^2_{portion} = 3$, given $\sigma^2_{batch} <$ 10,000, for various kegs per batch and portions per keg**

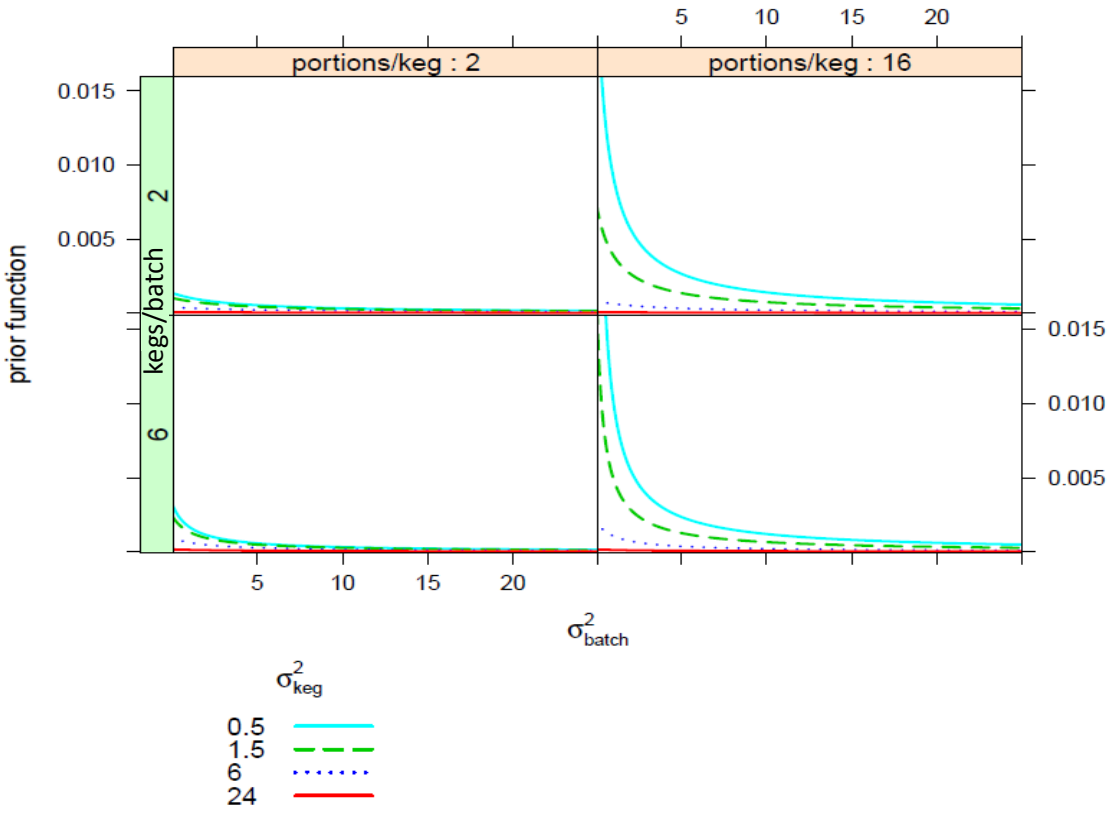


In summary, the plots of the prior function for Jeffreys' prior show its overall tendency to pull σ^2_{batch} and σ^2_{keg} estimates towards zero. Whether and how undesirable this behaviour is when estimating the uncertainty of variance components estimates will be evaluated in the studies to be performed. However, this behaviour being dependent on the design does not seem desirable for a prior intended to be vague. The behaviour for a variance component is also dependent on the size of other variance components. To be dependent on the bottom-level variance if that is an intrinsic source of variance may be reasonable for many applications. However, the dependence between higher-level variance components also seen may not be reasonable in many application areas.

5.2.6 Flat Prior (FLAT)

According to the SAS manual (SAS Institute Inc.: The MIXED Procedure, 2013) the "flat (equal to 1) prior for the variance components" is an option which "specifies a prior density equal to 1 everywhere, making the likelihood function the posterior". Consider the two-way variance components model with measurements taken from P portions

FIGURE 5.17: **Jeffreys' joint prior for $\sigma^2_{batch}, \sigma^2_{keg}$, for $\sigma^2_{portion} = 12$, given $\sigma^2_{batch} < 10{,}000$, for various kegs per batch and portions per keg**



from each of K kegs from each of B batches as described in Section 4.2. Though $L(y|\sigma_b^2, \sigma_k^2, \sigma_p^2) = L(y|\sigma_b, \sigma_k, \sigma_p)$, because the SAS syntax uses the variance components I assume that the flat prior is on σ^2 parameters rather than σ parameters. Then a flat prior $\pi(\sigma_b^2, \sigma_k^2, \sigma_p^2)$ is given by $f(\sigma_b^2) = f(\sigma_k^2) = f(\sigma_p^2) = constant$. The posterior is given by $p(\sigma^2|y, \mu) = \frac{p(y|\sigma^2,\mu)\pi(\sigma^2)}{\int_0^\infty p(y|\sigma^2,\mu)\pi(\sigma^2)\, d\sigma^2}$, where $p(y|\sigma^2, \mu)$ is the likelihood. So if the posterior was truly the likelihood this would mean that instead of $\pi(\sigma^2) = 1$ a normalising constant would be required but that would depend on y. Thus I believe that provided the posterior is proper the statement should be that the posterior is proportional to the likelihood. The above places a flat prior on σ^2 but Gelman (2006) did not advise this "as it seems to have the miscalibration toward higher values as described for σ, but more so, and also requires ≥ 4 groups for a proper posterior distribution." Also note that a uniform prior on σ^2 implies $\pi(\sigma) = 2\sigma$ (from Equation (5.1)) and thus the effective prior on σ parameters are not flat.

5.3 General Aspects related to Choice of Priors for Variance Estimation Studies

A useful list of aspects to be considered when choosing priors is given in Section 5.3.1 and these aspects are discussed further in Section 5.3.2. The specific priors selected for evaluation are discussed in Section 5.3.3 with respect to these aspects and the literature.

5.3.1 Summary

In considering the choice of priors relevant aspects include,

- how informative/ vague the prior is (note this will be influenced by the defining metric)
- use of existing/ historical information
- conjugacy
- calibration
- whether the priors are proper or improper
- the metric on which the prior is defined
- whether the priors are independently placed on the variance components
- the choice of the parameter values for the distribution family
- the support of the prior distribution and/or regions of low and high density
- whether the variance parameter is a "core" level variance or not.
- software

These aspects include some related to a list of concepts given in Gelman (2006) relating to the choice of prior distribution: conditionally-conjugate families; the improper limit of a prior distribution; weakly-informative prior distribution; and calibration. When discussing a uniform prior distribution Gelman stated that it was necessary to be explicit about the scale (referred to as "metric" below) on which the distribution is defined. The list includes the "whether the variance parameter is a "core" level variance" which I also thought to be important in this application area, though, aside from the bottom-level effect versus higher-level effects, it is not considered in the literature. Whether priors are independently placed on variance components also does not seem considered widely in the literature (except for random regression models) though Gelman (2006) suggests the use of hierarchical priors for variance components. Note general principles on choosing priors and advice for specific modelling situations was provided in 2015 to accompany the software Stan (Stan, 2017). This provides additional aspects to consider e.g. whether priors should avoid the boundary as well as not considering all the above aspects to be important e.g. referring to conjugacy the principles say "we don't care about that".

5.3.2 Discussion

The desire is to choose priors which are at most weakly informative, Gelman (2006) characterises a prior distribution as weakly informative if it is proper but is set up so that the information it does provide is intentionally weaker than whatever actual prior knowledge is available and Gelman (2009) describes them as priors, "which attempt to let the data speak while being strong enough to exclude various "unphysical" possibilities which, if not blocked, can take over a posterior distribution in settings with sparse data". Whilst the application area of the variation study will rule out certain values as being infeasible, it was desired to use priors which contain no more information than that which might be applied to any study of that type. For example in the pharmaceutical industry the prior should apply to that type of measurement across any similar product.

A form of prior that could be easily modified to include existing/historical data would be useful. Gelman (2006) indicates that a family of prior distributions $\mathrm{p}(\phi)$ is conditionally conjugate for ϕ if the conditional posterior distribution, $\mathrm{p}(\phi|y)$ is also in that class. This is a nice property since conditional conjugacy allows a prior distribution to be interpreted in terms of equivalent data and thus the prior distribution is easily used for informative priors as well as, at most weakly, informative priors. Conjugacy can also aid computational efficiency. For the variance components model an inverse gamma prior on the variance components is conditionally conjugate for the variances but this prior is not recommended.

The metric or scale on which the prior distribution is defined is important e.g. τ (precision), σ^2 $(= \frac{1}{\tau})$, σ, $\log(\sigma)$. In Figure 5.1 priors were plotted on the SD and variance metrics. It is seen that a prior which looked very vague on one metric looks more informative on the other. For variation studies both σ^2 and σ metrics seem relevant. When evaluating individual components of variation, σ^2 is often referred to, perhaps because the components are additive on this metric and components can be represented as a % of the total variation. However, σ is also referred to, for example when describing the variability of an analytical measurement e.g. requiring %RSD$\leq x$% where %RSD$=\frac{\text{Std Dev (SD)}}{\text{mean}} \times 100\%$. This is likely to be because the σ metric is used to place likely bounds on the range of variation of individual measurements e.g. 95% of values from a Normal distribution will lie within approximately 2σ of true value. Gelman (2006) said he "prefers to work on the scale of the standard deviation parameter σ_α, which is typically directly interpretable in the original model".

Whether the priors are proper or improper is an important aspect as improper priors may lead to improper posterior distributions. Improper priors also cannot be used in some software e.g. WinBUGS.

Gelman refers to calibration of the posterior mean, the Bayesian analogue to the classical notion of "bias" and that some positive bias is expected for improper priors. The

posterior distribution for small variance components is likely to be skewed, particularly when there are a small number of levels of the random effect in the design. Therefore the median or mode will usually be a more appropriate summary statistic than the mean and thus the concept of calibration would need modification to apply. This is not of primary interest in this thesis as it is concerned with estimating the uncertainty of the variance component estimators and will be assessing the quantiles of the distribution through the concept of coverage. However, for information it is planned to examine the mean and medians of the posterior distribution.

For variation studies I propose the concept of a "core" or intrinsic source of variability in the design to which other random effects could be compared is useful when choosing priors. The "core" variability represents the intrinsic measurement error which includes unavoidable sources of variation but excludes those which could be avoided, i.e. they could be intentionally controlled, at least in the short term. Considering analytical method precision studies:

- It is generally expected that there will be meaningful variation at the core level – in pharmaceutical analyses it is very unusual to have sufficiently precise analytical measurements that the analytical variation is of no practical importance.
- The core variation provides a reference against which other components of variation may be compared e.g. in analytical studies the preparation variance (including lower level sources of variation e.g. injection variance) generally cannot be avoided and the magnitude of this provides a reference point for other sources of variation.
- Often there is some prior knowledge of this core variation and it might be hoped (though often not achieved) that in comparison the higher- level components could be practically zero. The higher-level sources of variation can be avoided, at least in the short term, e.g. by keeping the same analyst, equipment etc. when making measurements.
- Often in designs, the core variance will be the lowest level variance studied. However it is possible that designs may separate the actual variance due to preparation from other sources of variation e.g. injection variance, introducing a further complication. The variance which is intrinsic to measuring different samples represents the core variability rather than necessarily the lowest level variance component assessed.
- It is usually easier to replicate at the core variance level than at higher levels.

This concept is useful in the choice of priors for the intrinsic variance in that, since it is not anticipated that it could be zero, it is not necessary to avoid priors such as inverse gamma which place infinite density at zero. Also, as it is usually easier to replicate at this level, the choice of prior will be less important. There may also be information at this level which allows more informative priors to be used. However an additional benefit is that other random effects can be expected to have sizes relative to the intrinsic

variance. Thus the parameters of priors for other random effects can be selected based on knowledge of typical values for the intrinsic variance in order to have low density for the standard deviation or variance values that are extremely unlikely for that type of study. Alternatively priors can be placed on some function which compares the random effect to the intrinsic variance rather than needing to explicitly set parameter values for a prior.

The scenarios chosen in Section 4.3 and most of the priors chosen in this chapter are based around values for the original batch sampling study. However, it is important to note that they will translate to similar designs with alternative values for the variance components. Provided the variance components for which it is wished to estimate the uncertainty of their estimators, are within $\frac{1}{12}$ and 4 times that of the intrinsic variance (or most likely value) the findings should apply for most of the priors to be studied. For example, if the intrinsic variance is twice that assumed in this study, the scale factor for the HT3 priors can be doubled. This is not necessarily the case for priors such as the vague IG priors which do not directly relate to the values in the batch sampling study. However, even in these cases, for example if the variances were smaller than those evaluated here, it is possible to rescale the data to achieve an intrinsic variance with similar values to those used here.

5.3.3 Further Discussion and Literature Review on Choice of Priors

5.3.3.1 Priors Selected for Evaluation

Gelman (2009) suggested using "weakly informative" priors as described in Section 5.3.2. The choice of the priors in this thesis was aligned with this philosophy. Whilst the application area of the variation study will rule out certain values as being infeasible, it was desired to use priors which contain no more information than that which might be applied to any study of that type. For example in the pharmaceutical industry the prior should apply to that type of measurement across any similar product.

For many of the chosen priors "weakly informative" priors were achieved by selecting the prior distribution parameter values to give higher density to values thought likely, given knowledge of the intrinsic source of variance (see Section 5.3.2 for discussion of intrinsic variances). Choice of the parameter also allows these priors to be re-scaled to other situations where the intrinsic variance is not assumed to be around 6. Other chosen priors do not have the necessity or flexibility of choosing a parameter for the prior. In particular, one prior (ICCU) was chosen where a Beta[α,β] prior distribution was placed on a function of the random effect of interest and the intrinsic source of variability. In the batch sampling study it is considered that the portion is the core source of variability. The beta prior on the ICCU ratio allows for incorporating a variety of shapes for the prior, though in this thesis only Beta[1,1] is evaluated. Given that the original study

does not split the portion variance into further components, the priors that might be used in that case were not considered further.

In Section 5.3.2 the metric on which the prior distribution was shown to be important and for variation studies both σ^2 and σ metrics seemed relevant. Given that it was illustrated in Figure 5.1 a prior that is non-informative (flattish) on one metric it will not be on the other and both metrics seem relevant, priors on σ^2 and σ metrics were chosen. The first three priors listed in Table 5.1 (UNI, HCY, HT3) were placed on the SD of the higher-level random effect whilst the other priors were placed on a function of the variance of the random effect(s). A uniform prior on the SD related to the bottom-level random effect, $\sigma_{portion}$, is used for UNI, HCY, HT3 priors. This aligns with guidance from Gelman (2006) that states for the lowest level variances typically, enough data will be available to estimate $\sigma_{portion}$ that one can use any reasonable non-informative prior distribution e.g. uniform on $\sigma_{portion}$ or $\log(\sigma_{portion})$.

Priors from the uniform and half-t families on σ were recommended by Gelman (2006), a paper often cited in the literature in relation to the choice of priors for variance components (for two examples see Weaver et al. (2012) and Hofer and Rauk (2017)). Gelman suggested using a uniform prior on the group level (Gelman uses the terminology hierarchical) standard deviation when the number of groups is ≥ 5, and using the half-t family (half-Cauchy) when the number of groups is small (below 5) and also in other settings where a weakly informative prior is desired. Polson and Scott (2012) also recommend the half-Cauchy prior saying "The half-Cauchy occupies a sensible middle ground within this class [hypergeometric inverted-beta priors]: it performs well near the origin, but does not lead to drastic compromises in other parts of the parameter space." However, it is noted that this may not directly translate to the half-Cauchy I am using, as Polson and Scott applied the half-Cauchy to a global scale parameter (λ) where λ^2 multiplies the residual variance σ^2 for the variance of the top-level effects. In this research, as well as the uniform prior, a half-Cauchy prior and a half-t with 3 degrees of freedom prior are investigated.

For many users a consideration is the software to be used, with SAS being commonly used in the pharmaceutical industry. PROC MCMC in SAS software allows a wide variety of proper priors to be used. However, a frequentist analysis for a mixed models is often performed using the procedure PROC MIXED, which also offers the facility to extend the modelling to perform a Bayesian analysis for variance components models (Wolfinger and Kass, 2000). In addition the Bayesian analysis is more efficient than PROC MCMC as it uses knowledge of the distribution of the posterior ($p(\theta|y)$ is equal to or well approximated by products of the inverted gamma density - see Box and Tiao (1973), Section 5.2.2). Hence the priors implemented in PROC MIXED using the PRIOR statement were also of interest (Inverse Gamma, Jeffreys' and FLAT).

PROC MIXED allows inverse gamma priors to be placed on the stratum variances - see Section 5.2.4. Note these are placed on linear combinations of the variance components, not the variance components themselves. For the latter, though the IG(ϵ, ϵ) distribution had originally been suggested in the literature as a vague prior for σ_h^2, with ϵ set to a low value such as 1 or 0.01 or 0.0001 (e.g. in examples in WinBUGS - Lunn et al. (2000)) it is too informative when variance components may be close to zero and thus not recommended by Gelman (2006). Placing the priors on the stratum variances results in approximately independent transformed parameters so that sampling can be performed from univariate inverse gamma densities (SAS Institute Inc.: The MIXED Procedure (2013), P517). However there are two aspects to be considered for the resultant priors on σ_h^2. Firstly, the priors will be dependent on the design of the study. Secondly, without further adjustment they provide support for negative values as is seen in Figure 5.11 which showed the marginal distributions for σ_{keg}^2 for the various informative inverse gamma distributions. In addition, when considering the effective prior distribution for σ_{keg}^2 given $\sigma_{portion}^2$, the proportion of density which corresponds to negative values of σ_{keg}^2 depends on the value of $\sigma_{portion}^2$. Rather than explicitly restricting/modifying the prior to a distribution which will give non-negative variance estimates, SAS rejects samples from the proposal distribution with negative values (the algorithm is described in Section 6.4). At first interpretation of the above, this may seem like the effective prior on σ_{keg}^2 is data dependent. However, conceptually the effective prior on σ_{keg}^2 could be considered a hierarchical prior for the σ_{keg}^2 with a parameter for $\sigma_{portion}^2$. Then a proper prior for σ_{keg}^2 could be defined with support only for non-negative σ_{keg}^2 values and with a normalising constant dependent on $\sigma_{portion}^2$. In practice the process used in SAS of rejecting proposal samples achieves the same result. Though the SAS documentation indicates that the inverse gamma prior distributions provide an opportunity to specify an informative prior, the option of using the functionality to set vague priors was also explored.

A key feature of Jeffreys' prior is that it is invariant under reparameterization of the parameter. Jeffreys' prior was designed by Jeffreys (1946) to be invariant under one to one parameter transformations. Given both the σ and σ^2 metric seem relevant this is an advantage. Jeffreys' priors satisfy the local uniformity property: a prior that does not change much over the region in which the likelihood is sizeable and does not assume large values outside that range (Lesaffre and Lawson, 2012). Jeffreys' prior violates the strong likelihood principle that all the information about the parameter is in the likelihood and this principle applies to the actual data collected irrespective of the original plans. Jeffreys' prior violates that principle because it uses the Fisher Information which is defined as an expectation (of the 2nd derivatives of the log likelihood) and thus an integration over a set of values which represent the universe of all possible experimental outcomes, as determined by the experimental design. The strong likelihood principle is not accepted by all statisticians so this may not universally be considered a problem.

However, the reliance of Jeffreys' prior on the actual design of the experiment may well be of practical concern.

Jeffreys' prior and the FLAT prior, are improper and thus could lead to improper posterior distributions. The implementation of Jeffreys' prior in PROC MIXED is a prior on the random effects in the restricted likelihood together with a separate prior on the fixed effects. For Jeffreys' prior Wolfinger and Kass (2000)) conjectured that under fairly mild conditions on dimensionality and configuration of the data, the posterior would be proper. Datta and Smith (2003) showed this for a design with one higher-level random effect. An improper posterior distribution could be an issue for the FLAT prior (a uniform distribution on the variance components) as Gelman (2006) notes that for a two-level design a uniform distribution on σ^2 requires at least four groups for a proper posterior distribution.

5.3.3.2 Further Discussion of the Literature

Gelman (2006) is a key literature reference for choosing priors for variance components. The basis for, and more detail concerning, his recommendations for the higher level variance sources are described further:

1. For the one-way nested variance model considered by Gelman the marginal likelihood for the group-level variance component σ_g approaches a finite nonzero value as $\sigma_g \to 0$. Thus in a hierarchical model the data can never rule out a group-level variance of zero, and so the prior distribution cannot put an infinite mass in this area.
2. From Point 1 a uniform prior on $\log(\sigma)$ should not be used as it would yield an improper posterior density with a spike at $\sigma_g = 0$.
3. Gelman recommends the uniform prior density on σ_g when the number of groups is reasonable ≥ 5. However, he described that if the number of groups is low, the uniform prior density on σ_g tends to lead to high estimates of σ_g and for 1 or 2 groups it actually results in an improper posterior density. This miscalibration is an unavoidable consequence of the asymmetry in the parameter space. Prior distributions for variance parameters in hierarchical models with variance parameters are restricted to be positive. Similarly, Gelman points out that there are no always-nonnegative classical unbiased estimators of σ or σ^2 in the hierarchical model. He did not recommend uniform prior density on σ_g^2 as it seemed to have the miscalibration toward higher values as described for σ_g, but more so, and also requires ≥ 4 groups for a proper posterior distribution.
4. The inverse-gamma(ϵ, ϵ) had been suggested in the literature as a vague prior for the variances with ϵ set to a low value such as 1 or 0.01 or 0.0001 (the latter values originally being used in examples in BUGS - Lunn et al. (2000) though inverse-gamma priors are now designated as not recommended). However this does not

have any proper limiting posterior distribution as $\epsilon \to 0$. As a result, posterior inferences are sensitive to choice of ϵ in cases where σ is estimated to be near zero, the resulting inferences will be sensitive to ϵ and the prior distribution doesn't look vague. The effect of ϵ on the inverse-gamma distribution can be seen in Figure 5.2, Section 5.2.4.1.

5. When the number of groups is small (below 5) Gelman recommends a prior from the folded noncentral t distribution which is conditionally-conjugate. In Gelman's school example he used the half-Cauchy - a special case of the conditionally-conjugate folded-noncentral-t family of prior distributions and has a broad peak at zero and a scale parameter A. The half-Cauchy distribution is $\frac{\text{Abs}(N(0,\tau))}{\sqrt{(\chi^2)}}$ where $\tau = \frac{1}{A^2}$ and N represents the normal distribution. In the limit $A \to \infty$ this becomes a uniform prior density. Gelman observes that large but finite values of A represent prior distributions which are "weakly informative" because, even in the tail, they have a gentle slope (unlike, for example, a half-normal distribution) and can let the data dominate if the likelihood is strong in that region. With reference to the 3 school, one-way variance component, example he said "we went to the trouble of using a weakly informative prior, a distribution that was not intended to represent our actual prior state of knowledge about but rather to constrain the posterior distribution, to an extent allowed by the data."

6. Gelman also discusses priors for a model with several higher-level variance components. He illustrates this with a split-plot latin square design where nine variance components were to be estimated. He suggests using a hierarchical prior where the variance parameters have a common distribution with hyperparameters estimated from the data. He proposed considering a half-Cauchy prior distribution with location 0 and scale A, and with a uniform prior distribution on A. He did not specify whether the uniform prior for A is improper or proper. For the latin square example Gelman showed that for most of the variance parameters the posterior medians were similar under the uniform prior or the hierarchical half-Cauchy prior but that the 75th and 97.5th percentiles were shrunk by the hierarchical half-Cauchy prior. He suggested that the hierarchical half-Cauchy model allows most of the variance parameters to be small but with the occasionally large which seems reasonable in the typical settings of analysis of variance, in which most sources of variation are small but some are large.

Gelman (2006) showed for the one-way nested variance model, the half-Cauchy prior was preferred to the uniform prior density on σ_g when there were 3 groups ("3 schools problem"). He also showed that for 8 groups ("8 schools problem") the uniform prior is reasonable though the half-Cauchy prior distribution would also perform well. It is not clear from the paper how the cut off between 4/5 groups was chosen, nor how generally applicable Gelman's recommendations are e.g. to changing the sample size at the bottom level and/or the relative size of the variances. Three groups is the smallest number of

levels I have seen a prior recommended for in the literature and is relevant to some of the small designs for variation studies. There are risks with choosing a slightly informative prior in a situation where there is a little data - the half-Cauchy prior requires a scale parameter to be chosen which, if chosen inappropriately, could affect the conclusions but, when chosen appropriately, enables infeasible values for the variance components to be ruled out and the variance components to be estimated more precisely. However, without that assumption it may be that nothing useful can be gleaned from the data. The provision of credible intervals, the aim of this research project, at least provides an estimate of the uncertainty around such estimators.

As described in Point 6 above, Gelman provides an interesting proposal regarding the use of hierarchical priors. He thought this could be the ultimate solution to the difficulties of estimating σ_g for models with many higher level variance components where each has a small number of levels. The hierarchical prior doesn't require direct specification of the scale factor for the half-Cauchy so it can be made less informative (assuming the improper uniform distribution for the scale factor). That may be good or bad dependent on whether we want it to be weakly informative or not, and whether we have any prior information which would suggest that different scale parameters should be used (though Gelman mentions batches of parameters so this aspect could be incorporated into the specification of the priors). Again there are pros and cons, with the hierarchical prior resulting in shrinkage between the estimates of variance components so the estimation of some may be biased by the values of others. However, the advantage is that the information from the other variance components is used to reduce uncertainty of the estimate of each individual variance component. Thus it is possible that a hierarchical prior set up this way may have the same marginal distribution as priors set up directly on the individual variance components but the hierarchical prior induces a correlation between them. Given that the design used for the evaluation in this thesis only has two higher-level variance components a hierarchical prior was not investigated.

As an alternative to Jeffreys' priors (based on the information matrix), reference priors have been proposed in the literature as vague priors. Ni and Sun (2003) note the concern about the "poor performance of the Jeffreys' prior in multi-parameter settings in that the parameter inter-dependence amplifies the effect of the prior on each parameter" and that "Bernardo (1979) proposed an approach for deriving a reference prior by breaking a single multi-parameter problem into a consecutive series of problems with fewer parameters. The reference prior is designed to extract the maximum amount of expected information from the data in the sense of maximizing the divergence (difference) between the posterior and the prior when the number of samples drawn goes to infinity." Bernardo used the Kullback-Leibler distance: $E[\log \frac{\pi(\theta|X)}{\pi(\theta)}]$, where expectation is taken over the joint distribution of X and θ to assess the divergence but other measures have been proposed. Ni and Sun (2003) comment that "the reference priors preserve desirable features of the Jeffreys' prior such as the invariance property, but they

often avoid paradoxical results produced by Jeffreys' prior in multi-parameter settings". Ghosh (2011) says: "For regular models where asymptotic normality holds, Jeffreys' general rule prior, the positive square root of the determinant of the Fisher information matrix, enjoys many optimality properties in the absence of nuisance parameters. In the presence of nuisance parameters, however, there are many other priors which emerge as optimal depending on the criterion selected.". van der Linde (2000) notes that the reference priors are often the same as Jeffreys' priors in particular circumstances (Berger and Bernardo (1992) applied the ordered group reference prior algorithm to balanced variance components and showed that Jeffreys' prior (based on $(\mu; \sigma^2; \tau^2)$) coincides with the reference prior if the ordered grouping of parameters involves one group only and Ye (1994) that one group prior is equal to Jeffreys' prior in the reparameterised problem).

Reference priors were not chosen for further exploration for the batch sampling design since:

- The form of Jeffreys' prior described in Section 5.2.5 excludes the location parameter, leaving three variance components of interest and thus only incorporates parameters of interest. More generally, breaking down of parameters into ordered groups is unlikely to be easy for a wide variety of situations in routine use.
- Jeffreys' prior is already implemented in SAS PROC MIXED, alternative reference priors would require derivation individually for different designs which may not be easy, particularly in unbalanced cases and thus will not be suitable for routine use.
- Given that for the one-way variance components model Jeffreys' prior is the same as the reference prior when there is only one group of parameters, it seems likely that they will be reasonably similar for the two-way balanced design to be investigated in this work.
- The reference priors can not be implemented in SAS PROC MIXED and if the reference priors are improper they could not be fitted in WinBUGS
- Both Jeffreys' and reference priors have the disadvantage that they take into account the design of the study.

Another principle used for producing priors is that of probability matching. Datta and Sweeting (2005) define a probability matching prior distribution as one "under which the posterior probabilities of certain regions coincide with their coverage probabilities, either exactly or approximately". There is much literature on proposals for probability matching priors in various design and modelling situations. Another complication is to be clear on which criteria it is desired to match on e.g. quantiles, distribution function, highest posterior density. Chang et al. (2003) developed a probability matching prior on the unobservable random effects for the one-way random ANOVA model and showed this to be different from the probability matching priors for interval estimation of the two variance parameters. Since, like the reference priors, they require development for specific situations the approach is not suitable for routine use where a wide variety of

situations will be encountered for variation studies. This view is aligned with the concluding remarks of Datta and Sweeting (2005): "The probability matching property is an appealing one for a proposed nonsubjective prior, since it provides some assurance that the resulting posterior statements make some inferential sense at least from a repeated sampling point of view. However, in view of the many alternative matching criteria, and the fact that in the multiparameter case there is usually an infinite number of solutions for any one of these criteria, in the authors' view it is inappropriate to use probability matching as a general paradigm for nonsubjective Bayesian inference. Instead, probability matching should be considered as one of a number of desirable properties ...". It is also worth noting that for some situations the probability matching prior is the same as a reference prior under certain ordering.

A third approach to the development of priors is based on shrinkage of the estimates of the random effects. Daniels (1999) proposed the uniform shrinkage prior $\pi(\sigma_h^2) = \frac{\sigma_p^2}{(\sigma_p^2+\sigma_h^2)^2}$. In the simulation study performed for the smallest number of levels (10), for the higher-level variance component coverage ranged from 0.85 (for $\frac{\sigma_p^2}{\sigma_h^2}$=0.2) to 0.97 (when $\frac{\sigma_p^2}{\sigma_h^2}$=1 or 5) for 90% nominal intervals. Daniels concluded "First, in very small samples, a truly noninformative prior does not exist. In this situation, the uniform shrinkage prior can be thought of as a proper prior which is a compromise between Jeffreys' prior which pulls the estimate strongly towards zero and the flat prior which pulls the estimate away from zero. Secondly, placing a uniform distribution on the shrinkage weight implies that the data variance, σ_p^2 [my notation] and the prior variance, σ_h^2 [my notation and I have presumed he means the higher-level variance] are of the same order of magnitude. Gustafson et al. (2006) extend these to a wider class and propose "conservative" priors for the higher-level variance component which deliberately give more weight to smaller values. Though the uniform shrinkage prior could be of interest, deliberately giving weight to smaller values was not considered appropriate in situations where the variance is being estimated to demonstrate it meets requirements.

Chapter 6

Details of Bayesian Analyses

Contents

This chapter explores various choices and issues in the implementation of the Bayesian analysis which will be used to evaluate the uncertainty of various variance component estimators under different design options for precision studies. The original study and various scenarios based on that study which will be used in the Bayesian analysis were described in Chapter 4 together with the model to be fitted. An introduction to the evaluation to be performed including how the number of simulation datasets was chosen is given in Section 6.1. Chapter 5 discussed the choice of priors likely to be applicable to precision studies. The priors are to be evaluated through examining the coverage of credible intervals. The credible intervals for the variance components and the total variance produced during the analyses are described in Section 6.2. Six families of priors were chosen to be screened as described in Section 5.1.1. The Bayesian analyses using three of the families of priors (UNI, HCY/HT3, ICCU) were performed using Gibbs sampling from the posterior and are further described in Section 6.3. The analyses for the other three families of priors (IG, JEFF, FLAT) were performed using sampling algorithms from a proposal distribution based on the inverse gamma distribution and are further described in Section 6.4. The analyses using Gibbs sampling were run using WinBUGS (accessed via R) whilst those using the inverse gamma proposal distributions were run using PROC MIXED(SAS Institute Inc.: SAS/STAT, 2011)[1] in SAS.

For the Bayesian analysis performed using Gibbs sampling, Sections 6.3.1, 6.3.2 and 6.3.4 describe the algorithms used, the assessment of convergence and approaches adopted to overcome issues identified. In Section 6.3.2 it is found that with the burn-in period employed, diagnostics typically used to assess convergence (Gelman-Rubin statistics), actually indicated a tendency for a chain of samples from the posterior to spend time in values in the tails of the distribution (usually very small values of variance components) rather than issues of non-convergence. The choice of number of posterior samples is discussed in Section 6.3.3. A summary of the findings and recommendations from implementing the Gibbs sampling for variance component estimation is provided in Section 6.3.5.

Sections 6.4.1 and 6.4.2 describe the algorithms used for sampling from the posterior in PROC MIXED and the initial choice of number of posterior samples. However, for some analyses the requested number of posterior samples was not provided. This occurs when acceptance rates drop too low. The causes of this and approaches to overcoming these issues are developed in Section 6.4.3, given that the SAS documentation and literature do not provide solutions. Section 6.4.4 summarises the final strategy employed to obtain full samples from the posterior and the success of this. The strategy employed results in samples from the posterior which are not independent and thus Section 6.4.5 discusses the number of posterior samples used in the final analyses.

[1] SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

The coverage of the credible intervals for the variance components in the range of interest is provided in Chapter 7 where the appropriateness of the six families of priors is evaluated. A summary of the credible intervals obtained from the priors for the various scenarios is provided in Chapter 8.

6.1 The Evaluation and Number of Simulation Datasets

Introduction to the Evaluation The performance of the priors will be evaluated using the frequentist concept of coverage of the credible intervals (CrIs). This is discussed further in Section 7.1. The estimated coverage is subject to simulation error from the number of datasets evaluated and the number of samples obtained from the posterior. Details of the analyses performed are given in Sections 6.3 (for the analyses using Gibbs sampling performed in WinBUGS) and 6.4 (for the analyses using inverse gamma proposal distributions performed in PROC MIXED) including the number of posterior samples taken. The choice of the number of simulation datasets is described in Section 6.1.

Number of Simulation Datasets At least 2000 datasets were required to be evaluated so that the estimated coverage $\hat{p}$ is likely to be within $\pm$ 0.01 of the true coverage p if the true coverage is approximately 0.95 (0.95 is the desired coverage). The rationale is as follows. For each dataset the true value is either within or outside the 95% CrI derived in a particular analysis. Thus the result of each simulation has a Bernoulli distribution and from a total number of simulated datasets (n) the number for which the CrI includes the true value will be Binomial(n,p) where p is the probability of inclusion. If $\hat{p}$ is the estimated probability of inclusion then approximately $\hat{p} \sim$ N(p,$\frac{p(1-p)}{n}$). A (1-α)% probability interval for $\hat{p}$ is approximately $p \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$. When p=0.95 and n=2000 this interval for $\hat{p}$ is 0.9404 – 0.9596 and thus $\hat{p}$ is likely to be within $\pm$ 0.01 of the true coverage p. Thus for n=2000, $\hat{p} < 0.94$ is unlikely if p is $\geq$ 0.95. For p=0.96 and n=2000 the 95% probability interval for $\hat{p}$ is 0.9514 – 0.9686 and thus $\hat{p} > 0.97$ is unlikely if p is $\leq$ 0.96. So to summarise if p is in range 0.95-0.96, $\hat{p}$ <0.94 or $\hat{p} > 0.97$ is unlikely. In Section 7.1 situations with $\hat{p}$ <0.94 or $\hat{p} > 0.97$ will be highlighted as they are unlikely to occur with true coverage in the range 0.95-0.96. Conversely a $1-\alpha$ Wald confidence interval (CI) for p is given by $\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and thus for $\hat{p}$=0.95, the half-width of the 95% confidence interval is approximately 0.1. So if $\hat{p}$ =0.95 a true coverage <0.94 is unlikely.

For those priors evaluated using Gibbs sampling (see Section 6.3) the same 2000 datasets were used across all the priors investigated. For those priors evaluated using PROC MIXED, a larger number of datasets (10000) was used (including the 2000 evaluated

using Gibbs sampling) since these analyses were much faster to run. For n=10,000 the 95% probability interval for $\hat{p}$ when p=0.95 is 0.9457 - 0.9543.

6.2 Credible Intervals

95% percentile equal tailed CrIs (pct) and 95% highest posterior density intervals were obtained for the variance (hpdv) and standard deviation (hpds) posterior distributions. Additionally for the priors assessed using Gibbs sampling highest posterior density intervals for the logged variance (hpdl) were also evaluated. By definition the logged variances will give same coverage as logged standard deviations and the pct intervals will give the same coverage for variance and standard deviation.

Percentile and highest posterior density credible intervals are defined as follows. For a parameter θ, a $100(1-\alpha)\%$ credible set $\mathbf{C}$ is a subset of the parameter space Θ such that $\int_{\mathbf{C}} \pi(\theta|X)\, d\theta = 1-\alpha$. The percentile credible interval is given by (L_p, U_p) where L_p is the $\frac{\alpha}{2}$ posterior quantile for θ and (L_p, U_p) where U_p is the $1-\frac{\alpha}{2}$ posterior quantile for θ. The highest posterior density region for θ is a subset $\mathbf{C} \in \Theta$ defined by $\mathbf{C} = \theta : \pi(\theta|x) \geq k$ where k is the largest number such that $\int_{\theta:\pi(\theta|x)\geq k} \pi(\theta|x)\, d\theta = 1-\alpha$. HPDinterval in R was used to calculate the interval with the default option allowSplit=FALSE. The interval is constructed from the empirical cdf of the sample as the shortest interval for which the difference in the ecdf values of the endpoints is the nominal probability. For a distribution with a single mode (and in practice distributions that are not severely multimodal) this will be the highest posterior density credible interval.

6.3 Gibbs Sampling Algorithm

Three families of priors were evaluated using a Bayesian analysis implemented in WinBUGS called via R2WinBUGS (Sturtz et al., 2005).

WinBUGS uses the Gibbs sampling algorithm which is described in Section 6.3.1. The package CODA (Plummer et al., 2006) was used to summarise the posterior and produce diagnostics from the output. For the chosen scenarios and models, convergence and autocorrelation of posterior samples are likely to be issues. A burn-in period was implemented where the initial samples from the posterior collected during the burn-in are not included in the final sample. For each Bayesian analysis 5 chains were run, each with a burn-in of 40,000 samples followed by a further 160,000 samples from the posterior. Given that reasonably high autocorrelation was expected, the samples were thinned to reduce running time when calculating CrIs and diagnostics, with only 1 in 10 being retained. This gave a total of 80,000 samples post burn-in. Given the number of simulations performed it is not practical to visually examine convergence of the chains.

TABLE 6.1: **Sampling methods used by WinBUGS for continuous target distributions**

	Continuous target distribution	Method
1	Conjugate	Direct sampling using standard algorithms
2	Log-concave	Derivative-free adaptive rejection sampling
3	Restricted range	Slice sampling
4	Unrestricted range	Current point Metropolis

After completion of the analysis of the first 1000 simulated datasets the convergence was evaluated using the Gelman-Rubin statistic and where this was high the analyses were repeated with a longer burn-in. The evaluation of convergence is described in Section 6.3.2 along with the chosen burn-in periods. The evaluation of any autocorrelation on the ability of the requested sample size to achieve the desired precision of the CrIs is discussed in Section 6.3.3. Occasionally the requested number of samples was not provided by WinBUGS as described in Section 6.3.4 and the analysis was re-run. A summary of the findings and recommendations is provided in Section 6.3.5.

6.3.1 Sampling Algorithm in WinBUGS

WinBUGS uses the Gibbs sampling algorithm which splits the parameters to be sampled into single or subgroups of parameters and samples from each one (single or subgroup) conditional on the most recent values of all the other parameters (Lunn et al., 2013). It successively samples from the conditional distribution of each parameter given all the other parameters. Where the conditional distribution is known, the algorithm directly samples from that distribution. If not known, it employs further sampling methods dependent on the target distribution. For continuous target distributions the sampling methods are used in the hierarchies shown in Table 6.1 (in each case a method is only used if no previous method in the hierarchy is appropriate). The default sampling algorithms in WinBUGS were used in the Bayesian analysis.

Given that the conditional distribution is not known for all parameters in the analysis, convergence and autocorrelation of posterior samples are likely to be issues.

6.3.2 Evaluation of Convergence

Convergence was mainly evaluated using the univariate Gelman-Rubin (G-R) convergence diagnostic statistics. Essentially these compare the within and between variances of chains which use different starting values. The estimate (denoted G-R Est.) and upper limit of the 95% confidence interval (denoted G-R CI) were examined. If convergence is achieved then the G-R statistics should be close to 1. Details of the statistics are given in Appendix E, Section E.1. Note that the upper limit of the 95% confidence

interval is not necessarily greater than the estimate. It is based on an F distribution which requires the assumption of normality. Brooks and Gelman (1998) suggest that G-R statistic should be less than 1.2 and ideally less than 1.1.

6.3.2.1 Original Burn-in

Table 6.2 shows for the first 1000 datasets the percentage where the G-R Est. or G-R CI are greater than 1.1 for each variance component and the total variance pooled over batch variances of 0.5, 6 and 24 for each scenario and prior. The G-R statistic is based on the assumption that the samples from the posterior distribution are normally distributed. The table shows the statistic for estimates on the variance, SD and logged variance (LVar) metrics.

It is seen that for the total variance the variance metric has a higher percentage of datasets with the G-R statistic greater than 1.1 and the log variance metric a lower percentage than the SD metric. However, for the variance components the results are more mixed. In most cases but not all, the SD metric had an equal or lower percentage of datasets with the G-R statistic greater than 1.1 than either the variance or log variance metric. However whether the proportion of datasets with high G-R statistic greater than 1.1 was higher on the variance or log variance metrics depended on the scenario and variance component. The portion variance had no datasets with G-R statistic greater than 1.1 except for the scenario k6p2_0.5 but even here the percentage was very low. The number of datasets with G-R statistic upper C.I. limit greater than 1.1 was the same or higher than the number of datasets with G-R statistic estimate greater than 1.1. The Gelman-Rubin statistic and upper CI limit values were poorest for scenario b3k2p2_0.5, especially for the HCY prior and for the total variance and batch variance component.

TABLE 6.2: **Percentage of datasets with estimate or upper C.I. limit for Gelman-Rubin statistic >1.1 pooled over σ_b^2 = 0.5, 6 and 24, for each variance component and total variance, by metric**

Scenario	Prior	G-R	Total			Batch			Keg			Portion		
			Var	SD	LVar	Var	SD	LVar	Var	SD	LVar	Var	SD	LVar
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	33.433	2.900	0.133	34.000	2.433	1.200	2.367	0.100	1.700	0.000	0.000	0.000
b3k2p2	**HCY**	**CI**	34.000	3.667	0.133	34.467	2.967	2.367	2.533	0.267	4.100	0.000	0.000	0.000
_0.5	**HT3**	**Est.**	7.667	0.233	0.000	8.200	0.167	1.267	1.900	0.000	2.233	0.000	0.000	0.000
	HT3	**CI**	8.067	0.300	0.033	8.633	0.300	2.533	2.000	0.033	4.400	0.000	0.000	0.000
	ICCU	**Est.**	10.900	0.400	0.000	13.167	0.300	0.000	0.467	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	10.933	0.433	0.000	13.367	0.367	0.000	0.467	0.000	0.000	0.000	0.000	0.000
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	1.100	0.033	0.000	1.367	0.067	2.033	0.033	0.100	4.267	0.000	0.000	0.000
k2p2	**HCY**	**CI**	1.167	0.033	0.000	1.367	0.167	3.633	0.167	0.367	8.500	0.000	0.000	0.000
_0.5	**HT3**	**Est.**	1.900	0.100	0.000	3.300	0.033	1.800	0.300	0.100	3.967	0.000	0.000	0.000
	HT3	**CI**	2.167	0.200	0.000	3.900	0.133	2.867	0.433	0.367	7.300	0.000	0.000	0.000
	ICCU	**Est.**	0.133	0.000	0.000	0.267	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	0.133	0.000	0.000	0.267	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.2 – *Continued from previous page*

Scenario	Prior	G-R	Total			Batch			Keg			Portion		
			Var	SD	LVar	Var	SD	LVar	Var	SD	LVar	Var	SD	LVar
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.433	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	3.167	0.433	0.100	3.500	0.367	0.900	0.667	0.133	3.000	0.000	0.000	0.000
k2p16	**HCY**	**CI**	3.633	0.633	0.233	4.033	0.633	1.800	0.900	0.433	5.067	0.000	0.000	0.000
_0p5	**HT3**	**Est.**	4.067	0.733	0.200	4.933	0.300	0.767	0.867	0.067	3.000	0.000	0.000	0.000
	HT3	**CI**	4.733	1.033	0.267	5.833	0.500	1.467	1.067	0.200	5.533	0.000	0.000	0.000
	ICCU	**Est.**	0.400	0.067	0.000	0.533	0.067	0.000	0.167	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	0.500	0.067	0.000	0.700	0.067	0.000	0.167	0.000	0.000	0.000	0.000	0.000
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	2.233	0.133	0.000	2.767	0.033	2.067	0.000	0.000	0.167	0.000	0.000	0.000
k2p16	**HCY**	**CI**	2.567	0.300	0.000	3.300	0.200	4.100	0.100	0.033	0.333	0.000	0.000	0.000
_6	**HT3**	**Est.**	1.200	0.167	0.000	1.867	0.100	2.367	0.067	0.067	0.233	0.000	0.000	0.000
	HT3	**CI**	1.467	0.200	0.000	2.267	0.200	4.767	0.133	0.133	0.233	0.000	0.000	0.000
	ICCU	**Est.**	0.533	0.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	0.633	0.033	0.000	0.767	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.2 – *Continued from previous page*

Scenario	Prior	G-R	Total			Batch			Keg			Portion		
			Var	SD	LVar	Var	SD	LVar	Var	SD	LVar	Var	SD	LVar
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	1.833	0.033	0.000	2.867	0.133	2.167	0.000	0.000	0.000	0.000	0.000	0.000
k2p16	**HCY**	**CI**	2.233	0.300	0.033	3.500	0.267	4.600	0.033	0.067	0.067	0.000	0.000	0.000
_24	**HT3**	**Est.**	0.700	0.067	0.000	1.400	0.133	2.800	0.033	0.000	0.067	0.000	0.000	0.000
	HT3	**CI**	0.900	0.100	0.000	1.800	0.333	5.500	0.100	0.067	0.067	0.000	0.000	0.000
	ICCU	**Est.**	0.333	0.000	0.000	0.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	0.433	0.000	0.000	0.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**Est.**	0.100	0.000	0.000	0.233	0.000	0.000	0.000	0.000	0.133	0.000	0.000	0.000
	UNI	**CI**	0.133	0.000	0.000	0.233	0.000	0.000	0.000	0.000	0.133	0.000	0.000	0.000
	HCY	**Est.**	3.333	0.467	0.000	3.467	0.267	0.267	0.033	0.100	1.933	0.000	0.000	0.000
k6p16	**HCY**	**CI**	3.933	0.800	0.067	4.100	0.500	0.767	0.067	0.333	3.600	0.000	0.000	0.000
_0.5	**HT3**	**Est.**	2.933	0.400	0.067	3.167	0.167	0.433	0.000	0.067	2.067	0.000	0.000	0.000
	HT3	**CI**	3.367	0.600	0.100	3.733	0.333	1.067	0.033	0.133	4.367	0.000	0.000	0.000
	ICCU	**Est.**	0.867	0.100	0.000	0.867	0.067	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	1.267	0.300	0.033	1.267	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.2 – *Continued from previous page*

Scenario	Prior	G-R	Total			Batch			Keg			Portion		
			Var	SD	LVar	Var	SD	LVar	Var	SD	LVar	Var	SD	LVar
	UNI	**Est.**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	UNI	**CI**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HCY	**Est.**	1.900	0.000	0.000	2.467	0.067	1.167	0.200	0.633	7.167	0.033	0.000	0.000
k6p2	**HCY**	**CI**	2.100	0.033	0.000	2.800	0.100	1.900	0.833	2.100	16.267	0.067	0.067	0.033
_0.5	**HT3**	**Est.**	2.400	0.300	0.000	3.267	0.167	0.400	0.100	0.400	8.533	0.033	0.033	0.033
	HT3	**CI**	2.667	0.400	0.033	3.733	0.200	0.633	0.467	1.900	17.700	0.067	0.067	0.067
	ICCU	**Est.**	0.267	0.000	0.000	0.367	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	ICCU	**CI**	0.267	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

6.3.2.2 Increasing Burn-in for Scenario b3k2p2_0.5

The Gelman-Rubin statistic and upper CI values were poorest for scenario b3k2p2_0.5, especially for the HCY prior and for the total variance and batch variance component. For scenario b3k2p2_0.5 for the HCY and HT3 priors, datasets (from all 2000 datasets analysed) with high G-R values were repeated with increasing burn-ins of 80,000, 160,000 and 320,000. The HCY prior was chosen as it had the poorest G-R statistics. The HT3 prior is also included for comparison. As initially the Gelman-Rubin statistic was only examined for the parameters on their main metrics of interest (σ^2_{total}, σ_{total}, σ^2_{batch}, σ^2_{keg}, $\sigma^2_{portion}$) the datasets repeated were those having G-R Estimate or CI greater than 1.2 for these variance components and metrics.

Tables 6.3, 6.4 and 6.5 provide a summary of the values of the Gelman-Rubin statistic over all σ^2_b values for the original (40,000 samples) and increased burn-in for the variance, SD and log variance metrics respectively. These are shown for 2000 datasets for the HCY prior for the total variance and the batch and keg variance components. The results for the portion variance component are not shown as there were no datasets with G-R statistic greater than 1.1.

Similar to the results in Table 6.2 it is seen from Tables 6.3, 6.4 and 6.5 that the number of datasets with G-R estimate or CI above 1.1 was less on the SD or log variance metrics than on the variance metric for the increased burn-in. However when examining the number of datasets with G-R estimate or CI above 1.2, it is seen that for the batch and keg variance there are a number of datasets for the log variance metric compared to none for the variance or SD metric for the increased burn-in.

Describing the results in more detail it is seen for the HCY prior, the total variance on the SD metric with the increased burn-in there were 8 datasets out of 6000 with the G-R CI above 1.1, though none greater than 1.2. For the total variance on the log variance metric there were none above 1.05. However, for the batch and keg variance components the reverse occurred where log variance metric had a greater number of datasets with the G-R estimate and CI above 1.1 compared to the SD metric. For the batch variance component on the SD metric with the increased burn-in there were 5 datasets out of 6000 with the G-R CI above 1.1, though none greater than 1.2. For the log variance metric there were 116 datasets out of 6000 with the G-R CI above 1.1, with 52 greater than 1.2. For the keg variance component on the SD metric with the increased burn-in there were no datasets with the G-R CI above 1.1, whereas for the log variance metric there were 247 datasets out of 6000 with the G-R CI above 1.1, with 94 greater than 1.2.

For the HT3 prior the total variance had no datasets with G-R statistics greater than 1.1 for the SD and log variance metrics but did have 321 datasets with G-R CI between 1.1 and 1.2 on the variance metric, but none greater than 1.2. Now examining the batch

TABLE 6.3: **Number of datasets categorised by Gelman-Rubin convergence statistic for original and increased burn-in for HCY and HT3 for scenario b3k2p2_0.5 - variance metric**

Prior	G-R	Value	Burn-in 40,000			Burn-in Increased		
			Total	Batch	Keg	Total	Batch	Keg
HCY	Est.	**0.95-1.05**	2520	2472	5635	2957	2900	5666
		1.05-1.1	1416	1427	218	1661	1675	221
		1.1-1.2	1286	1314	118	1380	1423	113
		1.2-1.3	769	778	29	2	2	0
		1.3-1.5	9	9	0	0	0	0
		>1.5	0	0	0	0	0	0
	CI	**0.95-1.05**	2474	2421	5620	2906	2844	5656
		1.05-1.1	1419	1434	223	1662	1683	223
		1.1-1.2	1248	1273	124	1430	1471	121
		1.2-1.3	634	645	31	2	2	0
		1.3-1.5	181	183	2	0	0	0
		>1.5	44	44	0	0	0	0
HT3	Est.	**0.95-1.05**	4988	4903	5710	5118	5032	5718
		1.05-1.1	571	619	186	587	635	189
		1.1-1.2	346	367	93	295	333	93
		1.2-1.3	93	109	11	0	0	0
		1.3-1.5	2	2	0	0	0	0
		>1.5	0	0	0	0	0	0
	CI	**0.95-1.05**	4954	4859	5693	5083	4987	5702
		1.05-1.1	578	640	194	596	658	196
		1.1-1.2	332	352	98	321	355	102
		1.2-1.3	112	118	12	0	0	0
		1.3-1.5	20	27	3	0	0	0
		>1.5	4	4	0	0	0	0

and keg variance components. The variance metric had 355 and 102 datasets with G-R CI between 1.1 and 1.2 (none above 1.2) for the batch and keg variance components respectively. The log variance had 147 and 242 datasets with G-R CI greater than 1.1 for the batch and keg variance components respectively but 68 and 103 of these had G-R CI greater than 1.2 respectively and 17 and 29 had G-R CI greater than 1.5 respectively. In contrast the SD metric had just one dataset with the G-R CI greater than 1.1 (this occurred for the keg variance component). It was originally thought that where a higher number of datasets with higher G-R statistics is seen for one metric compared to another this was an indication of non-normality affecting the statistic. This is investigated in Section 6.3.2.3.

The priors will be assessed by examining the coverage of the credible intervals obtained in the analysis as was described in Section 6.2. Hence, the effect of increasing the number of burn-in samples on the coverage was examined to see whether it made a practical

TABLE 6.4: **Number of datasets categorised by Gelman-Rubin convergence statistic for original and increased burn-in for HCY and HT3 for scenario b3k2p2_0.5 - SD metric**

Prior	G-R	Value	Burn-in					
			40,000			Increased		
			Total	Batch	Keg	Total	Batch	Keg
HCY	Est.	**0.95-1.05**	5587	5642	5989	5964	5975	6000
		1.05-1.1	234	204	5	33	22	0
		1.1-1.2	133	112	5	3	3	0
		1.2-1.3	37	33	1	0	0	0
		1.3-1.5	9	9	0	0	0	0
		>1.5	0	0	0	0	0	0
	CI	**0.95-1.05**	5534	5598	5976	5942	5962	5996
		1.05-1.1	234	213	12	50	33	4
		1.1-1.2	152	119	4	8	5	0
		1.2-1.3	29	26	3	0	0	0
		1.3-1.5	21	17	4	0	0	0
		>1.5	30	27	1	0	0	0
HT3	Est.	**0.95-1.05**	5964	5976	5995	5999	6000	6000
		1.05-1.1	24	16	5	1	0	0
		1.1-1.2	10	6	0	0	0	0
		1.2-1.3	2	2	0	0	0	0
		1.3-1.5	0	0	0	0	0	0
		>1.5	0	0	0	0	0	0
	CI	**0.95-1.05**	5954	5964	5977	5999	5997	5999
		1.05-1.1	32	22	17	1	3	0
		1.1-1.2	11	11	4	0	0	1
		1.2-1.3	1	0	2	0	0	0
		1.3-1.5	0	1	0	0	0	0
		>1.5	2	2	0	0	0	0

difference to the results. The results are shown in Table 6.6 for the HCY and HT3 priors for the b3k2p2_0.5 scenario. The minimum coverage over the three true values for σ^2_{batch} is provided. It is seen that the coverage achieved was very similar, differing by at most 0.001. Hence, though Section 6.3.2.3 contains the results of the investigation of the high G-R values, the burn-in was considered acceptable for the analyses to evaluate the priors.

6.3.2.3 Further Evaluation of Datasets with High G-R Statistics

In Section 6.3.2.2 the Gelman-Rubin statistics were examined for datasets from scenario b3k2p2_0.5 analysed with HCY prior. Figure 6.1 shows the G-R CI statistic on the various metrics plotted against each other for the HCY prior and the original burn-in of 40,000 samples. It is seen that datasets with high G-R CI on the SD metric also

TABLE 6.5: **Number of datasets categorised by Gelman-Rubin convergence statistic for original and increased burn-in for HCY and HT3 for scenario b3k2p2_0.5 - log variance metric**

Prior	G-R	Value	Burn-in 40,000			Burn-in Increased		
			Total	Batch	Keg	Total	Batch	Keg
HCY	Est.	**0.95-1.05**	5984	5849	5683	6000	5863	5708
		1.05-1.1	7	85	194	0	82	196
		1.1-1.2	6	44	83	0	45	75
		1.2-1.3	2	10	18	0	8	19
		1.3-1.5	1	9	11	0	2	2
		>1.5	0	3	11	0	0	0
	CI	**0.95-1.05**	5965	5752	5437	6000	5779	5469
		1.05-1.1	23	120	290	0	105	284
		1.1-1.2	5	64	155	0	64	153
		1.2-1.3	2	20	47	0	21	47
		1.3-1.5	2	22	31	0	21	26
		>1.5	3	22	40	0	10	21
HT3	Est.	**0.95-1.05**	5999	5793	5679	6000	5807	5704
		1.05-1.1	1	120	184	0	118	183
		1.1-1.2	0	60	85	0	57	84
		1.2-1.3	0	13	30	0	12	28
		1.3-1.5	0	12	13	0	6	1
		>1.5	0	2	9	0	0	0
	CI	**0.95-1.05**	5997	5684	5420	6000	5697	5445
		1.05-1.1	2	154	314	0	156	313
		1.1-1.2	1	82	138	0	79	139
		1.2-1.3	0	24	43	0	22	42
		1.3-1.5	0	30	33	0	29	32
		>1.5	0	26	52	0	17	29

typically have high values on the variance metric though the absolute value of the G-R CI statistic is smaller on the variance metric. However, there is not a similar relationship between the values on the SD and variance metrics with those on the log variance metric. Datasets with high G-R CI values on the SD and variance metrics have slightly higher G-R CI values than typical on the log variance metric. However, most datasets which have high G-R CI values on the log variance metric do not have high G-R CI values on the SD and variance metrics.

To investigate the reasons for the high G-R statistics and the differences between the G-R statistics on the different metrics datasets were identified where the Gelman-Rubin CI was greater than 2 on at least one metric for the original analysis and burn-in period. For the batch variance 21 datasets had Gelman-Rubin CI greater than 2 on at least one metric whereas the keg variance had no such datasets. Table 6.7 identifies the datasets

FIGURE 6.1: **Comparison of Gelman-Rubin CI for original burn-in on variance, SD and logged variance metrics**

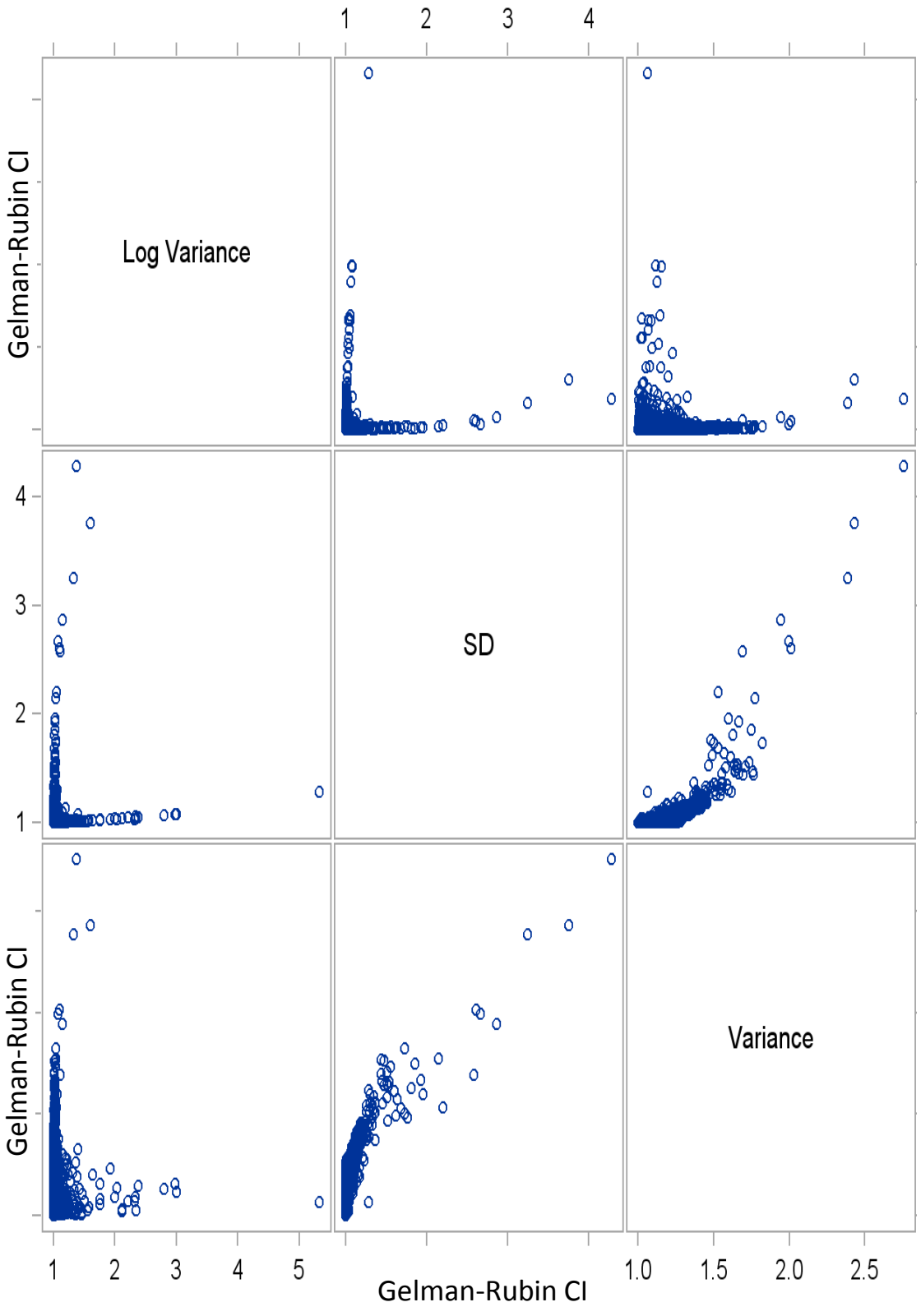


TABLE 6.6: **Comparison of minimum coverage (over σ^2_{batch} values 0.5, 6, 24) of CrIs for analyses with original and increased burn-in**

		Total		Batch		Keg		Portion	
Prior	**Burn-in: CrI Type***	**Orig.**	**Incr.**	**Orig.**	**Incr.**	**Orig.**	**Incr.**	**Orig.**	**Incr.**
HCY	**hpdv**	0.993	0.993	1.000	1.000	1.000	1.000	0.961	0.961
	hpds	0.969	0.969	1.000	1.000	0.995	0.995	0.959	0.958
	hpdl	0.897	0.897	0.998	0.998	0.966	0.967	0.958	0.958
	pct	0.806	0.806	0.996	0.996	0.980	0.980	0.957	0.957
HT3	**hpdv**	0.988	0.988	1.000	1.000	1.000	1.000	0.962	0.962
	hpds	0.963	0.963	1.000	1.000	0.995	0.995	0.961	0.961
	hpdl	0.883	0.882	0.997	0.997	0.965	0.964	0.958	0.958
	pct	0.801	0.801	0.998	0.997	0.981	0.981	0.957	0.957

* hpdv,hpds,hpdl = hpd intervals for variance, SD and log variance
pct = percentile

TABLE 6.7: **Number of datasets on each metric where G-R CI >2 for batch variance component for original analysis**

Batch	**All**	**Metric**		
Variance		**Variance**	**SD**	**Log Variance**
0.5	7	2	3	4
6	10	1	5	5
24	4	1	1	3

where the G-R CI is greater than 2 for the batch variance by the true batch variance and the metric.

To explore the reasons for the high GR CI the ten datasets with a true batch variance of 6 were re-run ten times with different initial values with the original burn-in of 40,000 samples. This was then repeated with a burn-in of 80,000 samples. Figure 6.2 shows the G-R CI for the original analyses and the repeats on the variance, SD and logged variance metrics. It is seen that though the original analysis had high G-R CI values on at least one metric for at least one variance component, when repeated in general, high values were not obtained. The repeated analyses where the G-R CI was greater than 1.5 are listed in Table 6.8. It is seen that only four of the ten datasets repeated had G-R CI greater than 1.5 but this occurred for datasets 408 and 727 more than once. These observations suggests that the convergence issue is predominantly likely to be related to the initial values and/or the random stream rather than an issue with the dataset itself. However, the dataset itself may have some influence.

The chains of samples from the posterior were then plotted for the repeated analyses

FIGURE 6.2: **Gelman-Rubin CI for the original and repeated analyses with burnin-in of 40,000 and 80,000 samples on variance, SD and log variance metrics for selected datasets**

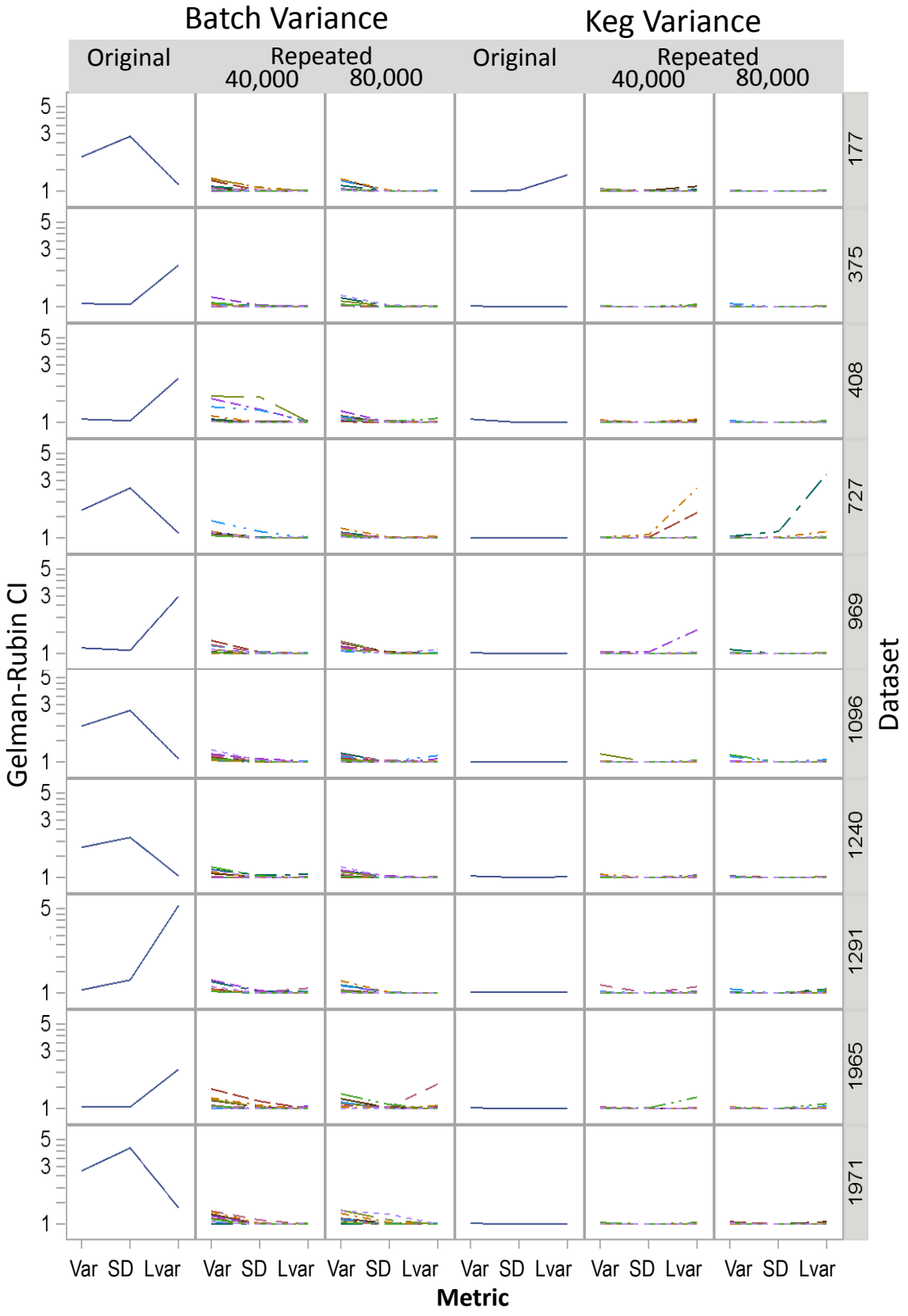


TABLE 6.8: **Repeated datasets where G-R CI >1.5**

Dataset	Variance	Burn-in	Repeat	Metric	G-R CI
408	Batch	40	4	Var	1.57
408	Batch	40	5	SD	1.63
408	Batch	40	5	Var	1.65
727	Keg	40	1	Lvar	1.62
727	Keg	40	8	Lvar	2.59
727	Keg	80	2	Lvar	3.39
969	Keg	40	4	Lvar	1.58
1965	Batch	80	7	Lvar	1.61

where the G-R CI was greater than 1.5 to examine the reasons for the high G-R CI values. They are shown in Figure 6.3 for the batch variance and Figure 6.4 for the keg variance.

In Figure 6.3 it is seen that two datasets with high G-R CI for the batch variance on the SD and/or variance metric were due to one chain having high batch variance sample values for an interval during the sampling. This was also visible in the plot on the log variance metric but did not result in a G-R CI value greater than 1.5. In the plot of the G-R CI statistic on the log variance metric against the SD metric in Figure 6.1 (1st row, 2nd column) these are examples of the datasets seen with higher values on the SD metric but lower on the log variance metric (points stretching to the right). The third dataset had a high G-R CI for the batch variance on the log variance metric which was due to having low batch variance sample values for an interval during the sampling. The G-R CI values were not greater than 1.5 on either the SD or variance metric nor was the excursion visible on the plots. In the plot of the G-R CI statistic on the log variance metric against the SD metric in Figure 6.1 (1st row, 2nd column) these are examples of the datasets seen with higher values on the log variance metric but lower on the SD metric (points on the left hand side stretching upwards). In Figure 6.4 for all four datasets the high G-R CI was observed on the log variance metric for the keg variance. The G-R CI values were not greater than 1.5 on the SD and variance metrics and the chain did not visibly protrude below the rest of the values - though the drop to near zero values is visible for dataset 969.

A close up of the chains for two of the datasets/analyses with high G-R CI values are shown in Figure 6.5 and for two datasets without high G-R values in Figure 6.6 on the log variance metric. The main excursion responsible for the high G-R CI value is seen in each plot in Figure 6.5 but there are also smaller intervals where the values are very small ($< 10^{-5}$) in both figures. It appears that the chain gets temporarily stuck at very small values and very occasionally this is for a not minimal proportion of the chain which is then detected by the G-R statistics.

FIGURE 6.3: **Posterior sample chains for datasets with G-R CI > 1.5 for batch variance**

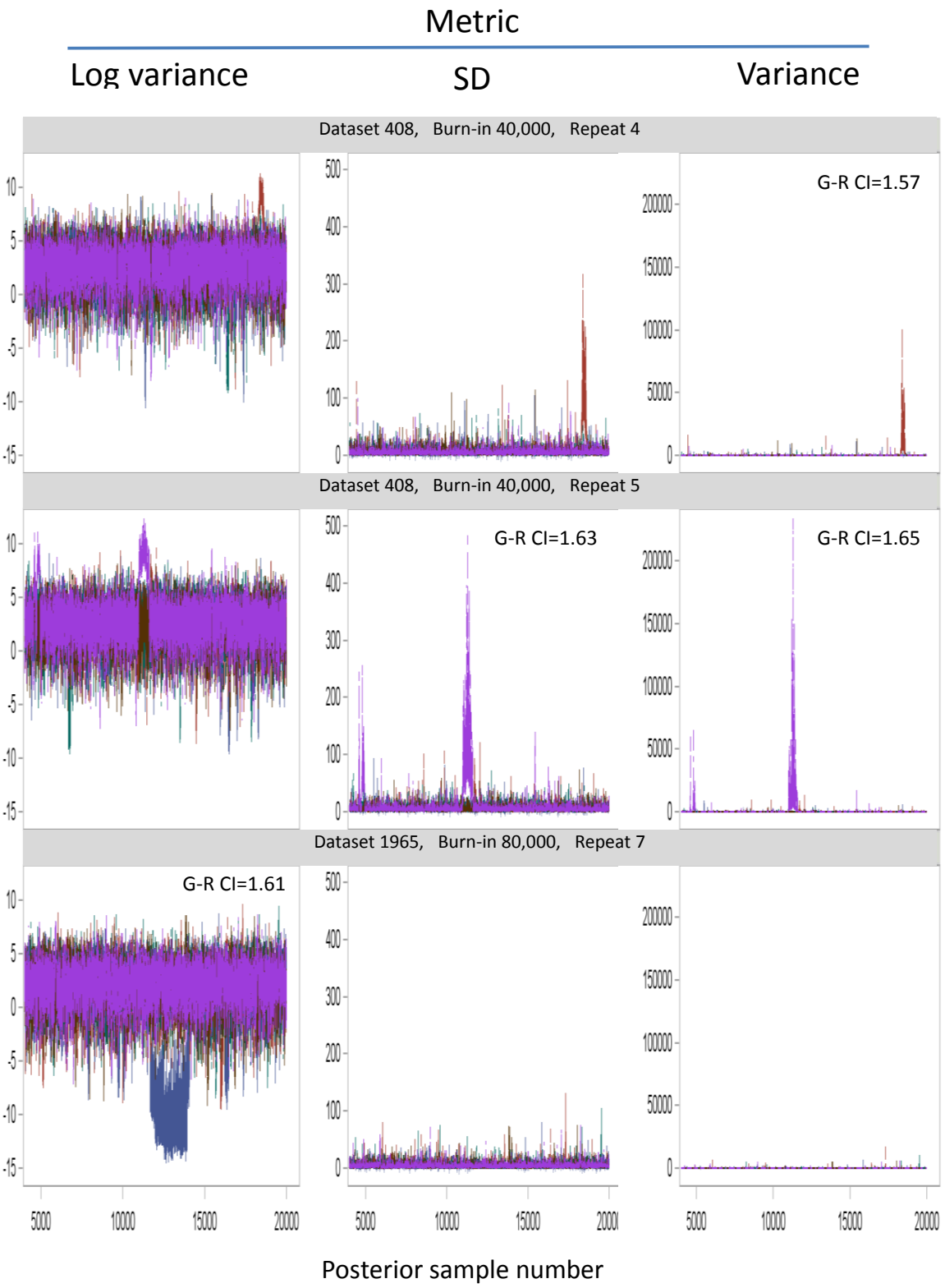


FIGURE 6.4: **Posterior sample chains for datasets with G-R CI > 1.5 for keg variance**

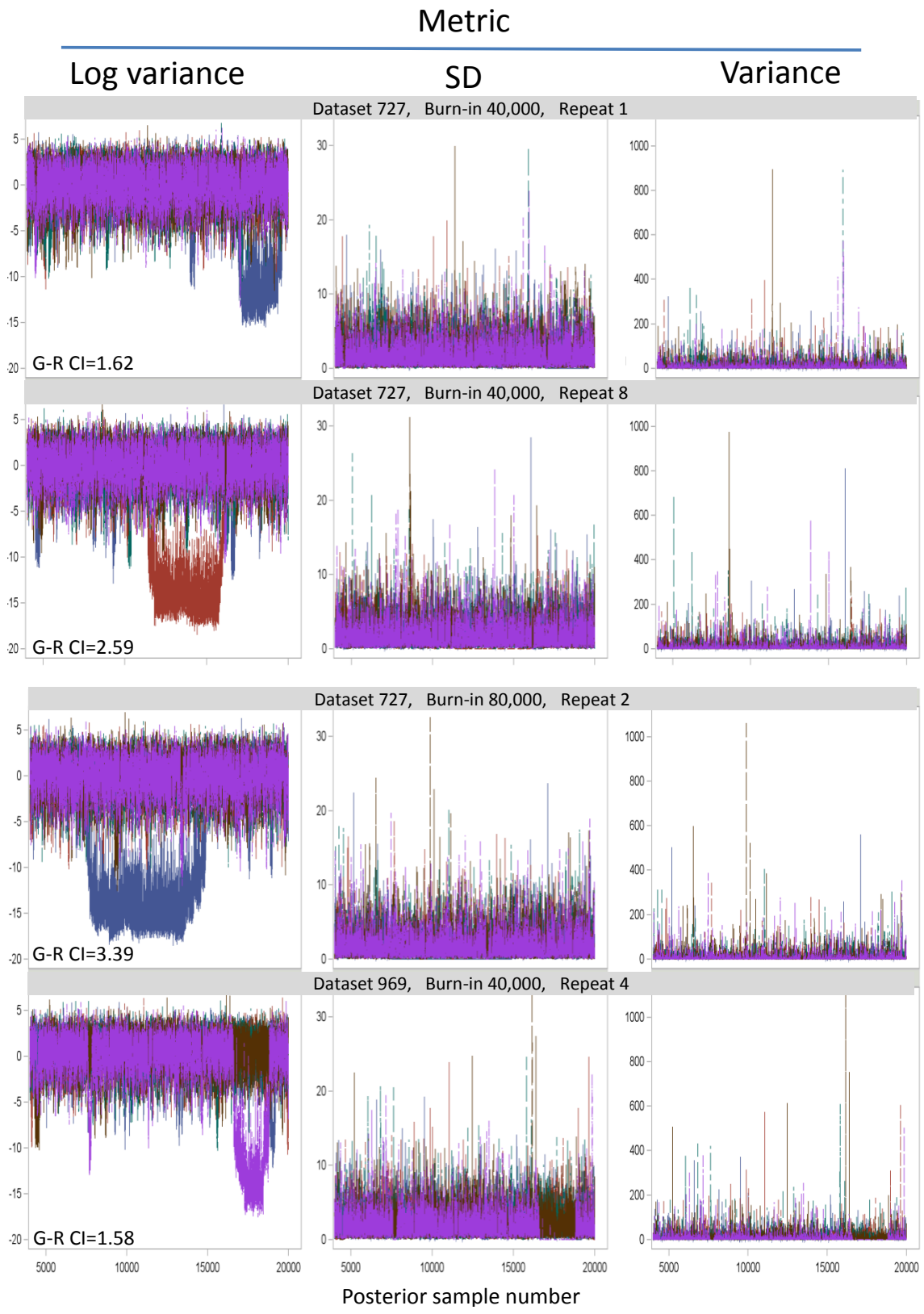
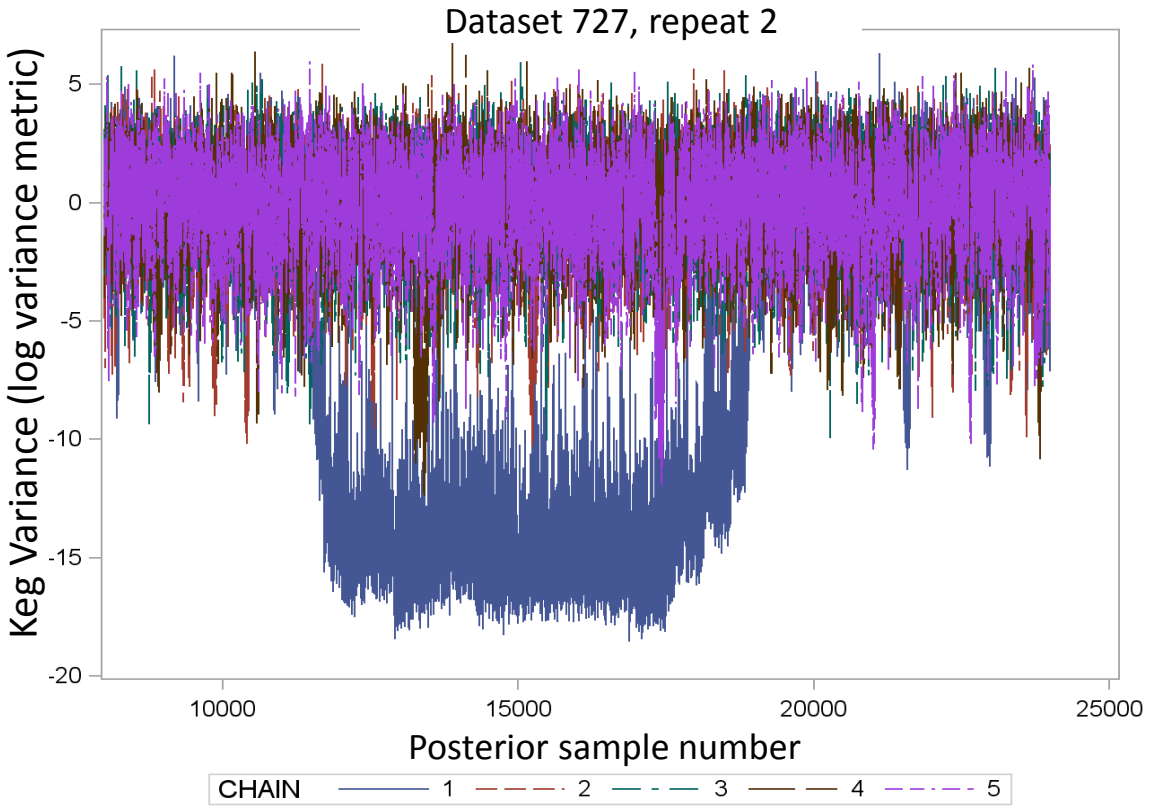


FIGURE 6.5: **Posterior sample chains for two examples with high G-R CI - log variance metric)**



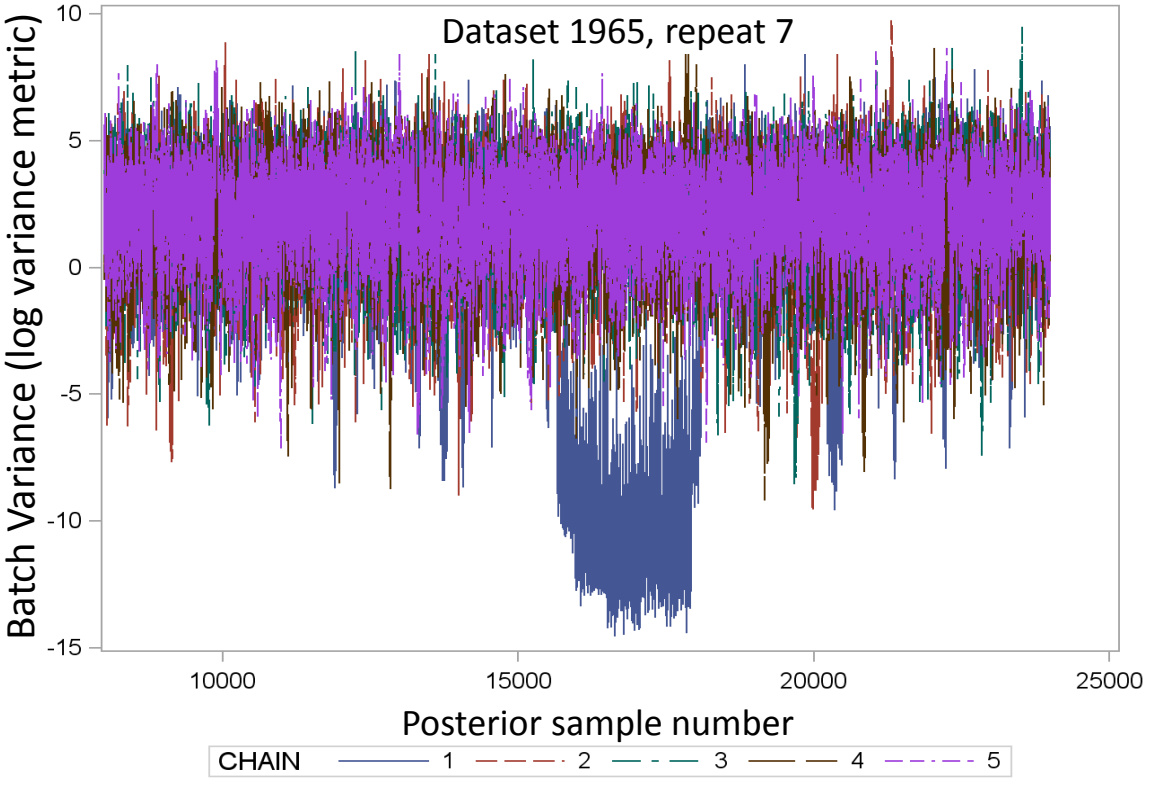
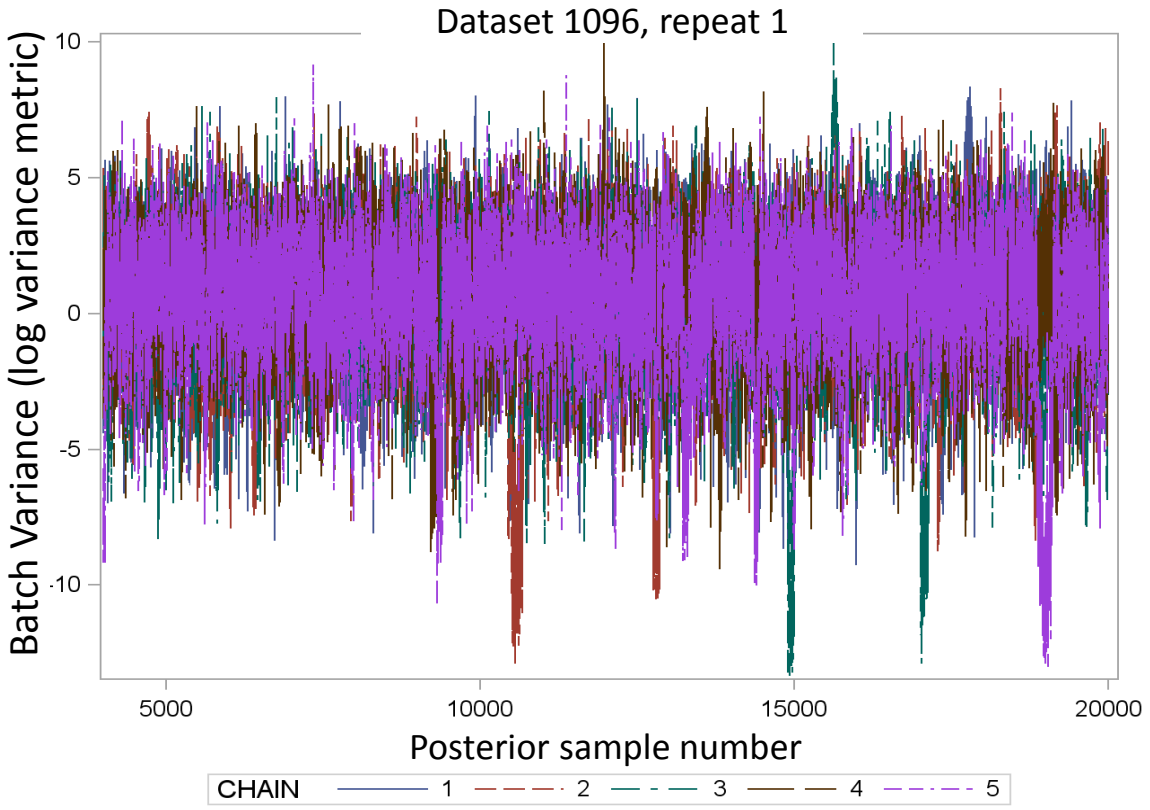
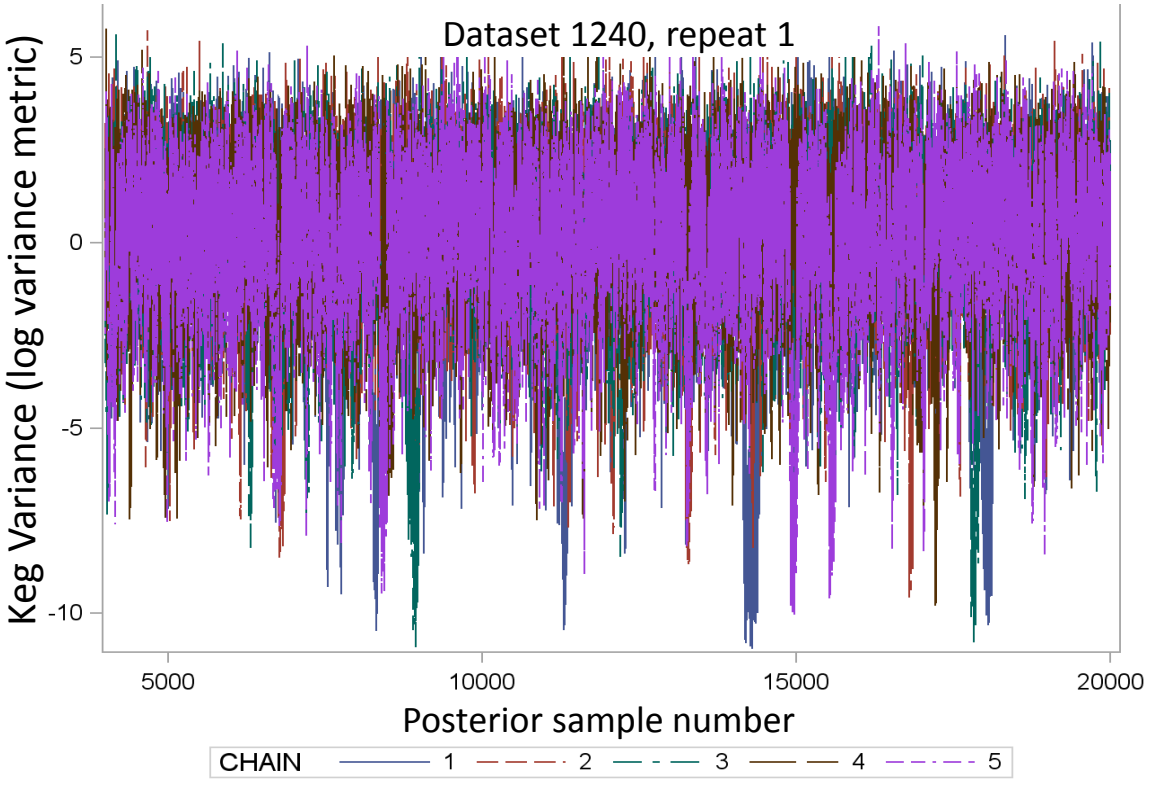


FIGURE 6.6: **Posterior sample chains for two examples with acceptable G-R CI - log variance metric**





An investigation did not identify any particular cause of this behaviour - either as a particular multivariate combination of initial values, nor being due to a local minimum of the deviance. Whilst datasets were re-run with a longer burn-in and in most cases this solved the issue, it seems likely that this was due to random chance that the chain did not enter the region of very low values in the repeated analysis rather than due to lack of convergence from the initial values. Hence it was considered that the burn-in period used for the study was sufficient despite the high G-R CI values on the log variance metric for a few analyses.

The high G-R statistics thus appear to be a result of poor mixing / a chain spending considerable time at these low values rather than a lack of convergence. This was referred to as "stickiness" by Gilks et al. (1996), and Besag and Green (1993) refer to hierarchical Bayes models when the prior information is sufficiently diffuse to create computational black holes from which no single-component updating scheme can escape. In the investigations here, whilst a few analyses had very low values for a period of time, there were none observed that did not "escape".

Excursions to high values are more detectable on the variance and SD metric than the log variance metric using the G-R statistic (see dataset 408 in Figure 6.3). Thus it was concluded that provided the G-R statistics were acceptable on the SD and variance metrics, there would not be substantial excursions/mixing involving high variance component posterior sample values. For excursions to low values, typically between $< 10^{-5}$ and $< 10^{-15}$, the values are very close on the SD and variance metrics compared to the range of the data and thus it is not surprising that the G-R statistics on those metrics are not affected by them.

It is also seen that aside from the very low values for one chain the posterior samples have a more symmetrical distribution than those on the SD or variance scale, suggesting that this metric is useful in the examination of convergence but that high G-R values should not necessarily be taken as an indication of non-convergence.

Figures 6.7 and 6.8 show the lower and upper limits for the highest posterior density intervals for the ten repeated analyses for datasets 727 and 1965 respectively. The analysis repeat which had a high G-R statistic is circled on the x-axis. It is seen that, though the lower hpd limit is clearly smaller on the log variance metric, on the SD and variance metrics they don't appear outlying and in the context of the true variances investigated in the study being 0.5, 6 or 24 any differences are practically unimportant.

FIGURE 6.7: **Highest posterior density intervals for keg variance component for ten repeat analyses of dataset 727 with 80,000 burn-in samples (repeat with high G-R on log variance metric circled)**

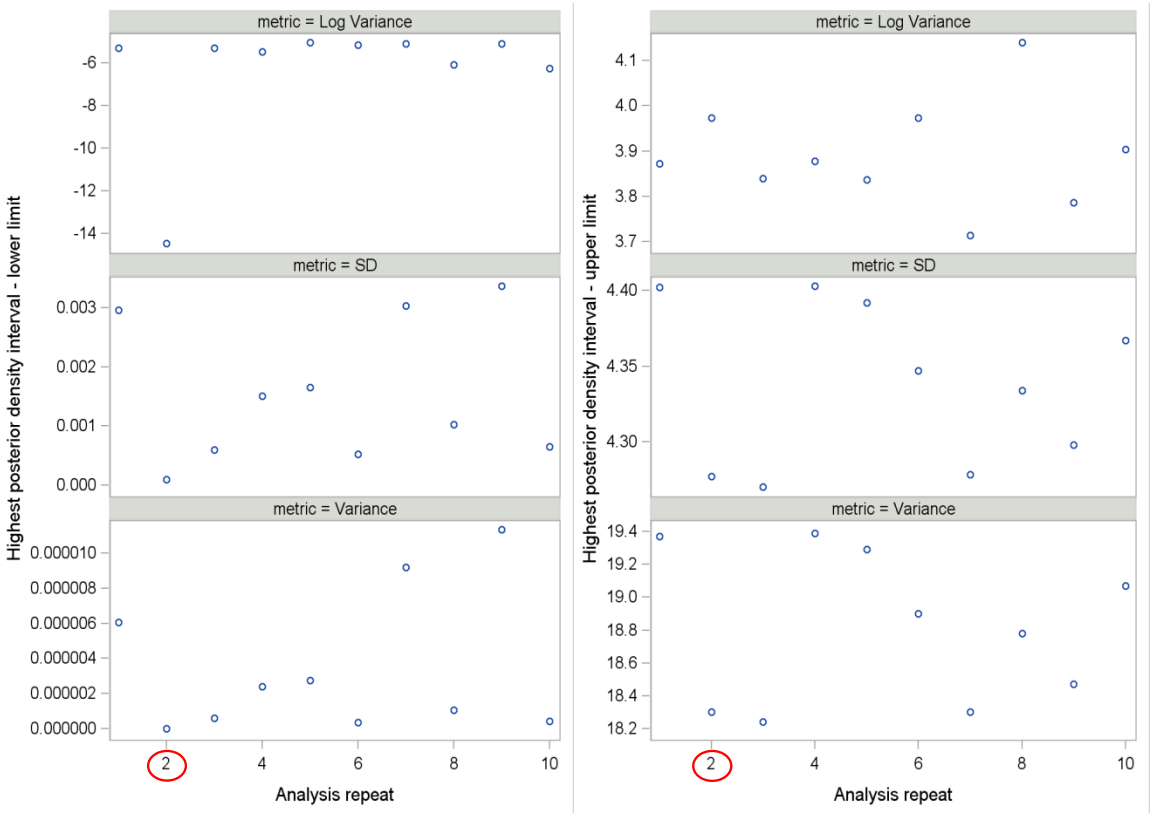
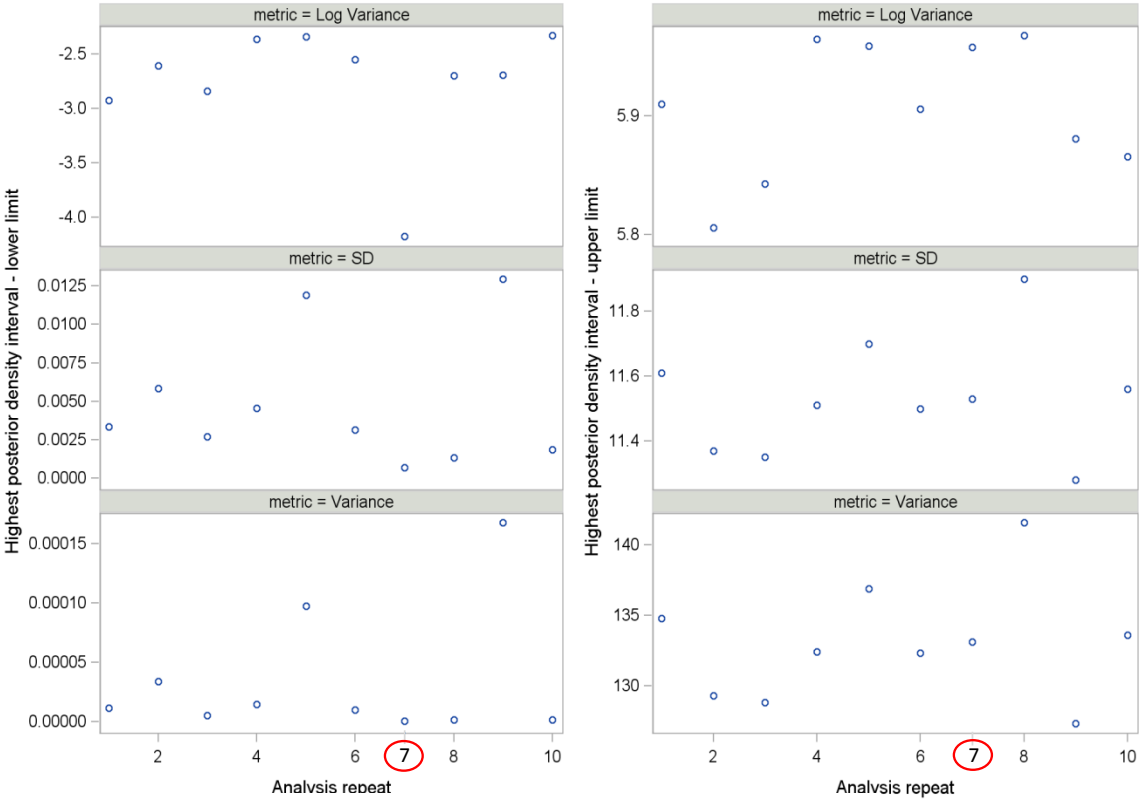


FIGURE 6.8: **Highest posterior density intervals for batch variance component for ten repeat analyses of dataset 1965 with 80,000 burn-in samples (repeat with high G-R on log variance metric circled)**



6.3.2.4 Final Burn-in

The number of datasets with Gelman-Rubin estimate or upper confidence interval (CI) value over all scenarios, σ_b^2 values and priors for the final burn-in used in the first 1000 datasets are shown in Table 6.9. The results are shown for all three metrics variance, SD and log variance. However, the log variance metric is considered less relevant for the variance components as a high G-R value did not necessarily indicate a lack of convergence as discussed in Section 6.3.2.3. The G-R statistics are based on the assumption of normality. Posterior variance samples on the SD metric are less skewed than those on the variance metric and thus this was considered the most relevant metric to assess convergence. It is seen that the keg variance had most datasets with G-R statistics greater than 1.2. However, only 0.2% of datasets had G-R CI greater than 1.1 and only 0.08% of datasets had G-R CI greater than 1.2. Thus given these results and the minimal effect (≤ 0.001) seen on the coverage when the burn-in was extended it seems reasonable to only perform the extended burn-in for the HCY prior for the b3k2p2_0.5 scenario.

The CODA package also provides a multivariate version of the G-R statistic. This provides a value that bounds above the potential scale reduction factor for any linear combination of the variables. Samples from the posterior on all three metrics for each variance of interest were submitted to CODA. Over all scenarios, priors and batch variance components only 0.2% of datasets had the G-R multivariate statistic greater than 1.2. The combination having the greatest percentage was scenario k6s2_0.5, prior HCY and batch variance of 0.5 with 1.8% greater than 1.2. Given the effect of the metrics on the univariate G-R statistics seen above, in particular that some datasets had a high G-R CI on the log variance metric, it does not seem surprising that some high multivariate G-R values might be seen when the statistic includes all three metrics. Given the low percentage of datasets with the multivariate statistic greater than 1.2, and that the high values of the univariate G-R statistics did not indicate non-convergence, the results obtained for the multivariate G-R statistic were considered acceptable.

TABLE 6.9: **Number of datasets categorised by Gelman-Rubin convergence statistic for various metrics for the final burn-in used in the analysis**

G-R	Value	Variance				SD				Log Variance			
		Total	**Batch**	**Keg**	**Portion**	**Total**	**Batch**	**Keg**	**Portion**	**Total**	**Batch**	**Keg**	**Portion**
	0.95-1.05	79277	78403	83340	83998	83702	83814	83861	83998	83973	82575	81375	83998
	1.05-1.1	2742	3254	455	0	215	125	89	1	22	849	1499	1
Est.	**1.1-1.2**	1716	2018	194	2	72	53	34	1	5	408	743	1
	1.2-1.3	264	323	11	0	11	6	11	0	0	96	211	0
	1.3-1.5	1	2	0	0	0	2	3	0	0	44	114	0
	>1.5	0	0	0	0	0	0	2	0	0	28	58	0
	0.95-1.05	78976	78014	83171	83981	83575	83675	83533	83981	83944	81665	79298	83981
	1.05-1.1	2864	3400	556	15	288	210	284	15	41	1212	2401	16
CI	**1.1-1.2**	1779	2120	235	2	102	79	112	2	8	615	1216	1
	1.2-1.3	254	320	27	1	22	16	34	1	4	200	404	1
	1.3-1.5	105	118	9	1	11	12	22	1	3	142	295	1
	>1.5	22	28	2	0	2	8	15	0	0	166	386	0

6.3.3 Autocorrelation and Number of Samples from Posterior

As described at the beginning of Section 6.3 a total of 80,000 posterior samples were obtained post burn-in and after thinning of 1 in 10 samples. The thinning was performed to reduce the computational time in calculating credible intervals and diagnostics. This section evaluates the adequacy of the thinned sample size. Despite the thinning, the posterior samples still exhibited autocorrelation and thus it is necessary to take this into account when assessing the adequacy.

It was considered desirable to achieve the equivalent of at least 2000 independent samples and preferably 4000 samples from the posterior. The rationale for this is now described. For the highest posterior density CrIs the assessment will be made comparing the coverage achieved by the CrIs with the desired probability (0.95). If the CrI estimator has a true coverage of 95% then a posterior sample is either inside or outside the CrI with a probability of 0.95. Hence a similar rationale may be made to that in Section 6.1 and 2000 independent posterior samples are recommended. It is noted that it is not clear what effect the uncertainty in the credible intervals due to posterior MCMC sampling error will have on the evaluation of the coverage of the priors. However it is thought that this would only be important when differentiating between estimates of coverage which are close and in those cases it may also be important to consider the number of datasets evaluated. For percentile CrIs a similar justification applies. However, the individual limits for the percentile CrIs may also be of interest. Given samples θ from the posterior, for the lower limit of a $(1-2Q)$ percentile CrI the estimate of the quantile is denoted $\hat{\theta}_Q$. Let $\hat{P}_Q = Pr(\theta \leq \hat{\theta}_Q)$. Then it is desired that $Pr(Q - R \leq \hat{P}_Q \leq Q + R) = 1 - \alpha$, where R is the allowable error in $\hat{P}_Q$ and $1-\alpha$ is the confidence. Then as described in Section 6.1, applying a Bernoulli distribution to n independent samples and using the Wald confidence interval for an estimated probability gives a $(1-\alpha)$ confidence interval $\hat{P}_Q \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_Q(1-\hat{P}_Q)}{n}}$. Thus for 2000 independent samples the 95% CI for $\hat{P}_{0.025}$ is 0.018-0.032 and for 4000 independent samples the CI is 0.020-0.030. Thus whilst it was considered desirable to have no more than the uncertainty corresponding to 2000 independent samples based on the 95% coverage of the intervals, 4000 is to be preferred to provide less uncertainty with regard to the lower and upper percentile CrI limit estimation.

To take the autocorrelation of the thinned posterior samples into account when assessing the adequacy the effective sample size (ESS) is calculated. The formula is given by

$$\frac{n}{1 + 2\Sigma_{k=1}^{\infty}\rho(k)},$$

where n is the number of samples from the posterior and $\rho(k)$ is the correlation at lag k. The denominator is related to the spectral density at zero frequency of a stationary time series ($1 + 2\Sigma_{k=1}^{\infty}\rho(k) = 2\pi S_\theta(0)$ where $S_\theta(0)$ is the spectral density at zero frequency).

TABLE 6.10: **Number of datasets categorised by ESS for final burn-in used in the analysis**

ESS	Variance			SD			Log Variance		
	Total	**Batch**	**Keg**	**Total**	**Batch**	**Keg**	**Total**	**Batch**	**Keg**
0-1000	0	1	2	0	5	3	0	35	2304
1000-1500	2	5	7	2	46	53	1	537	3509
1500-2000	6	25	18	8	276	1106	3	1801	2790
2000-4000	487	2286	2671	445	5068	5465	202	12938	14405
>4000	83505	81683	81302	83545	78605	77373	83794	68689	60992

CODA fits an autoregressive model to the posterior samples and estimates the spectral density at frequency zero to give an estimate of the effective sample size.

Table 6.10 shows the level of effective sample sizes achieved over the final analyses for posterior samples of the total variance and the batch and keg variance components. These are shown for the three metrics variance, SD and log variance. The portion variance is not shown as on all three metrics all datasets had an effective sample size greater than 4000.

It is seen that the metric has a considerable effect on the ESS.

- For the total variance the variance, SD and log variance metrics have just 8, 10 and 3 datasets with an ESS less than 2000 respectively.
- For the batch variance the variance metric has 31 datasets with an ESS less than 2000, the SD metric has 327 and the log variance metric has 2373.
- For the keg variance the variance metric has 27 datasets with an ESS less than 2000, the SD metric has 1162 and the log variance has 8603.

Where the effective sample size was smaller this typically occurred for the HCY and HT3 priors. The batch variance on the SD metric and log variance metrics tended to have smaller ESS for the scenario k2s16_24 for true batch variances of 6 or 24 and additionally 0.5 for the log variance metric. The keg variance on the SD metric typically had smaller ESS for the k6s2_0.5 scenario. The keg variance on the log variance metric generally had smaller ESS for all scenarios where the true keg variance was 0.5 but not those with true values of 6 or 24.

Raftery (1992) also provides a method of sample size estimation as now briefly described. The estimated sample size for variable U is based on the process $Z_t = d(U_t \leq u)$ where d is the indicator function and u is the qth quantile of U. The process Z_t is derived from the Markov chain data by marginalization and truncation, but is not itself a Markov chain. However, Z_t may behave as a Markov chain if it is sufficiently thinned out. The smallest value of thinning interval k which makes the thinned chain Z_t^k behave as a Markov chain is calculated. The required sample size is calculated from this thinned sequence. Since some data is 'thrown away' the sample size estimates are conservative. Raftery gives

TABLE 6.11: **Number of datasets categorised by RESS for final burn-in used in the analysis**

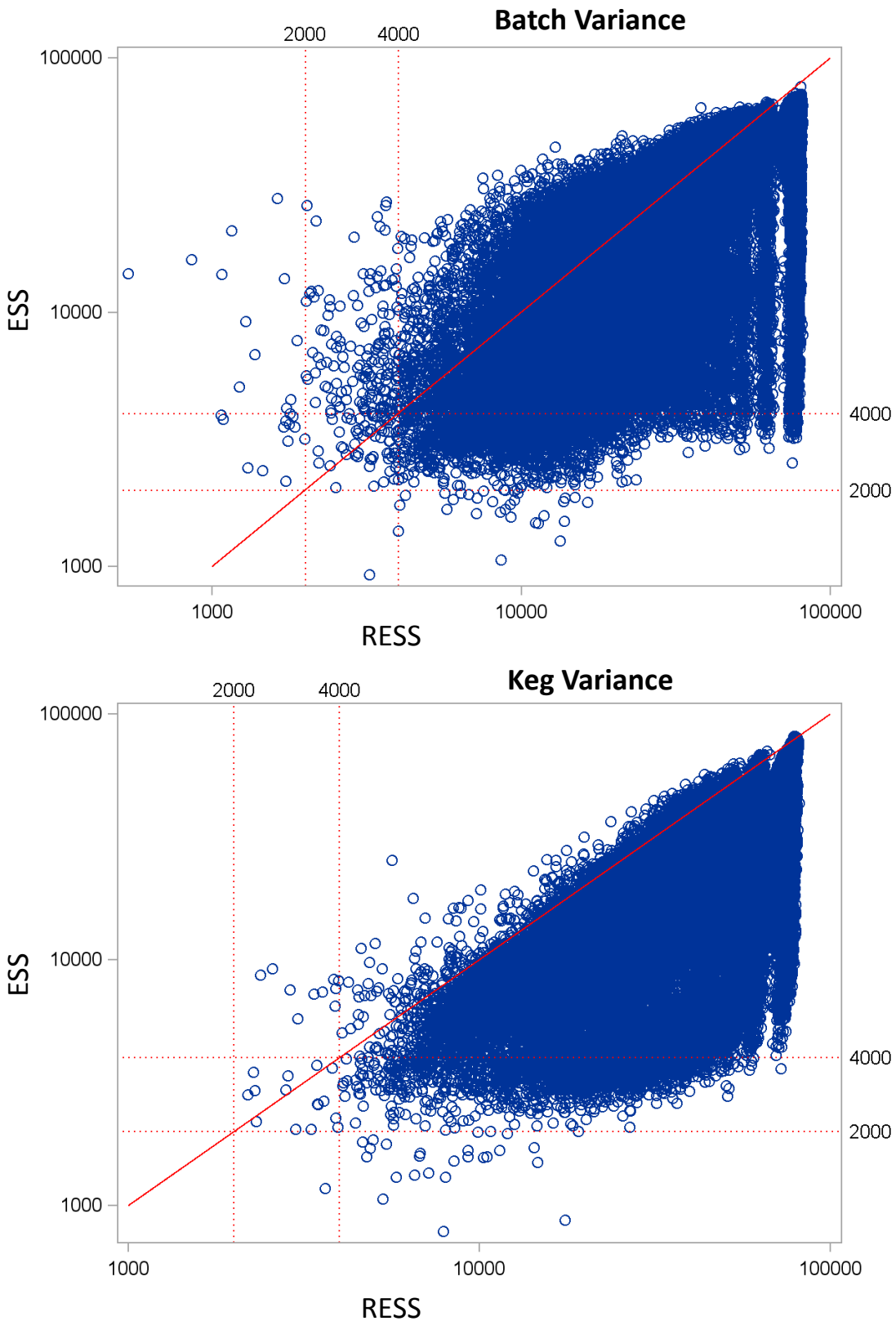
RESS	Variance			SD			Log Variance		
	Total	**Batch**	**Keg**	**Total**	**Batch**	**Keg**	**Total**	**Batch**	**Keg**
0-1000	3	2	0	3	2	0	3	2	0
1000-1500	7	9	0	6	9	0	7	9	0
1500-2000	11	14	0	12	14	0	12	14	0
2000-4000	190	219	26	191	220	26	189	220	26
>4000	83789	83756	83974	83788	83755	83974	83789	83755	83974

the sample size from the posterior to give the equivalent of 3746 independent samples the number required to give the 2.5 or 97.5 percentiles to an accuracy of 0.5 with 95% confidence. In order that the numbers are comparable to the effective sample sizes given in Table 6.10, the required number of samples from the posterior obtained using the Raftery method in CODA are transformed to a Raftery effective sample size (RESS) by the formula RESS= 80000*3746/required number, where 80,000 is the total number of samples from the posterior. The RESS achieved over the final analyses for posterior samples of the total variance and the batch and keg variance components are given in Table 6.11. These are shown for the three metrics variance, SD and log variance. The portion variance is not shown as on all three metrics all datasets had a RESS greater than 4000.

The ESS is plotted against the RESS in Figure 6.9. Reference lines at 2000 and 4000 samples are drawn. In addition the line of equality is plotted. It is seen that whilst there is a correlation between the two, one is not strongly predictive of the other. It is also interesting that despite the RESS being known to be conservative as it 'throws' away samples it is actually more frequently giving larger effective sample sizes than those based on the spectral density, especially for the keg variance. There were no datasets with both ESS and RESS less than 2000. In almost all cases where the ESS was less than 2000, the RESS was greater than 4000.

The metrics of interest for the parameters were variance and SD for the total variance and variance for the variance components. On these metrics, over all the analyses for varying priors, scenarios and true batch variances a few analyses did not reach the target effective sample size of 2000. The batch variance parameter had most analyses with effective posterior sample sizes <2000 but they only represented 0.04% of the analyses having an effective sample size <2000, 0.007% <1500 and 0.001% < 1000. Similar results were seen for the Raftery sample sizes. Thus the sample sizes used for sampling from the posterior for the analyses were considered satisfactory.

FIGURE 6.9: **Comparison of effective sample size with sample sizes from the Raftery method for batch and keg variance components (red solid line is line of equality)**



6.3.4 Sampling Stopping

Very occasionally the analysis in WinBUGS stopped partly through sampling from the posterior. The cause(s) of this was not fully identified as re-running the analysis almost always completed successfully. It is thought that on the rare occasion this occurred, the particular random sample obtained caused numerical problems with the algorithm. An example error message obtained is: "Error in chol.default(W) : the leading minor of order m is not positive definite.", where m is a value e.g. 6,7,11. Having identified that the analysis might be stopped by WinBUGS, the R program/procedure was set up to automatically check the number of samples obtained and if fewer than requested the dataset was resubmitted for analysis.

6.3.5 Summary

Gibbs sampling algorithm in Winbugs was used for the analyses to evaluate the UNI, HCY, HT3 and ICCU priors. The initial analyses were run with a burn-in of 40,000 (unthinned) samples for each of 5 chains. Given that visual evaluation of each dataset is impractical, Gelman-Rubin statistics are used to check for convergence of the chains. For the b3k2s2_0.5 scenario and HCY and HT3 priors the Gelman-Rubin statistics had higher than desired values and thus analyses for the datasets with higher values were repeated with larger burn-in periods. This reduced the Gelman-Rubin values. However investigations identified that the higher values were due to excursions of the chains to values in the tails of the distribution (usually very low values) rather than non-convergence from the starting values - referred to as "stickiness" by Gilks et al. (1996). The coverage of the highest posterior density intervals changed at most by 0.001 when the final analyses (including the additional burn-in) were compared with the original for the b3k2s2_0.5 scenario and HCY and HT3 priors. Thus the burn-in period was considered adequate. The work also identified that the metric employed has considerable effect on the Gelman-Rubin statistic - perhaps to be expected as the Gelman-Rubin statistic is based on the assumption of normality. The samples on the log variance metric were closer to a normal distribution than those on the SD or variance metric. However, the work showed that results from this metric cannot be used as an automatic default. The SD and variance metrics better identify excursions to higher values. On the log variance metric the Gelman-Rubin statistic might be high due to excursions to very low values (typically between $< 10^{-5}$ and $< 10^{-15}$) but not practically important in terms of the estimation of the highest posterior density interval.

Due to the autocorrelation of the posterior samples the effective sample size for estimating the highest posterior intervals is less than the actual sample size. The aim was for a posterior sample size equivalent to at least 2000 independent samples and preferably more than 4000. Post burn-in, 160,000 posterior samples were obtained from each of the

5 chains and then thinned to retain 1 in 10 posterior samples giving a total of 80,000 posterior samples for each analysis. Effective samples sizes were estimated through spectral density and Raftery's methods. On the variance metric (thought to be the most relevant) the batch variance component had most datasets with effective sample sizes less than 2000. As only 0.04% of the analyses had an effective sample size (spectral method) less than 2000 (0.03% for Raftery's method), and no dataset had an effective sample size less than 2000 on both methods, the thinned sample size of 80,000 was considered adequate. It is interesting to note that though the design of the Raftery's sample size estimate is conservative in general, it more often gave higher sample sizes than the spectral method. However for datasets with low sample sizes (< 2000) each method gave approximately the same percentage of datasets, though not the same datasets.

6.4 Sampling Algorithms with Inverse Gamma related Proposal Distributions

Bayesian analyses with three families of priors (IG, JEFF, FLAT) were performed using sampling algorithms from a proposal distribution based on the inverse gamma distribution which were available in PROC MIXED in SAS/STAT® software. The Bayesian analysis requires a proposal distribution for the posterior sampling (usually taken from the inverse gamma family), a sampling algorithm (PROC MIXED provides four) and the choice of the number of samples to take from the posterior distribution. This section describes these three aspects and the investigations and actions taken when sampling issues occurred. An overview is given below and Sections 6.4.1 - 6.4.5 describe these in more detail.

Initially the analyses for this thesis were generated using SAS software, Version 9.2 of the SAS System for PC (the version recorded in the SAS program log is noted as v9.02.02M3). However a bug was reported in SAS software Version 9.2, "LOGG values computed in the OUT= data set of the PRIOR statement in PROC MIXED are incorrect" (SAS Institute Inc.: PN41931, 2010), logG being the logged density values for the proposal distribution. This required either the hotfix B80011 (SAS Institute Inc.: B80011, 2011) to be applied to Version 9.2 or Version 9.3 to be used in which the issue was fixed. In the examples I investigated the logged density values output to the dataset were found to differ by a constant between Version 9.2 (no hotfix) and Version 9.3. Thus I did not expect the issue to affect the posterior samples obtained using the IC algorithm. However my investigations found that in some situations the posterior function value also differed between versions and also that the chain of numbers in the posterior samples differed between versions. Thus all the analyses were repeated in SAS software Version 9.3 (the version recorded was v9.03.01M2) and those are reported in the thesis. Despite the changes between versions the results were very similar.

Littell et al. (2006) describe how the proposal (base) distribution is selected. PROC MIXED makes use of the fact that the fixed-effects parameters can be analytically integrated out of the joint posterior, leaving the marginal posterior density of the variance components. For three of the four sampling algorithms in order to better approximate the marginal posterior density of the variance components, PROC MIXED transforms them in a way similar to that used to create expected means. It uses minimum variance quadratic unbiased estimates MIVQUE(0) equations (Rao, 1972; Giesbrecht, 1989) to perform the transformation with the purpose to create approximately independent parameters whose density can then approximated by a product of univariate inverted gamma densities (see Gelfand et al. (1990)). PROC MIXED uses the MIVQUE(0) method to compute initial values. To determine the parameters for the inverted gamma densities, PROC MIXED evaluates the logarithm of the posterior density over a grid of points in each of the transformed parameters. PROC MIXED then performs a linear regression of these values on the logarithm of the inverted gamma density. Alternatively the SAS software allows the user to input their choice of inverse gamma parameters to the proposal distribution.

The four sampling algorithms available in PROC MIXED are described in Section 6.4.1 and the independence chain is recommended as the main algorithm to be used for the evaluation of the priors. The initial choice of sample size from the posterior is given in Section 6.4.2.

At the end of the sampling, the SAS software provides information on the acceptance rate for the proposals from the proposal distribution. The acceptance rate reported is described further in Section 6.4.1. If the acceptance rate is too low the SAS software stops sampling and reports a warning message in the program log. This occurred for a number of analyses and thus for these analyses a good estimate of the credible interval is unlikely to be obtained. This issue of posterior sampling stopping is explored in Section 6.4.3. It is found that it typically occurred when the inverse gamma distribution proposed by SAS had minimal density for positive variance parameters. The SAS software documentation and other literature does not provide advice on the action to be taken when sampling stops and thus an approach was developed to overcome this problem whilst still using PROC MIXED in the SAS software. The parameters of the inverse gamma proposal distribution were modified to give a proposal distribution with adequate density for positive variances. The development of the approach is described in Section 6.4.3 and the final procedure used is presented in Section 6.4.4. For the analyses with modified proposal distributions the number of posterior samples requested was increased to 100,000 to take into account the sampling algorithm rejecting some proposals and repeating the previous proposal. In addition those datasets with a sampling acceptance rate less than 0.8 for the default proposal distributions were repeated with 100,000 posterior samples requested. The autocorrelation of the samples and the number of samples from the posterior are discussed in Section 6.4.5.

6.4.1 MCMC Algorithms in SAS Software PROC MIXED

SAS software provides a choice of four sampling algorithms for sampling from the posterior using the PROC MIXED procedure. Section 6.4.1.1 describes the independence chain algorithm which is the default in PROC MIXED and Section 6.4.1.2 describes the other three algorithms. If the proposal does not meet constraints on the parameters (e.g. variances should not be negative unless the nobound option is used) then it is rejected. It doesn't appear to be documented but it is assumed this is the same for all four algorithms. Section 6.4.1.3 explores the algorithms further and Section 6.4.1.4 discusses the choice of the algorithm to be used for the evaluation of the priors.

The SAS software documentation for PROC MIXED and the literature referred to use different notation for the posterior and proposal distributions. To avoid confusion in this section the posterior function is denoted $f(.)$ (note this is proportional to the posterior density function but may not be normalised) and the proposal distribution is denoted $g(.)$ which is the notation used by SAS. All formulae taken from the literature are translated into this notation. In addition some of the literature uses x or y to denote parameter proposals or samples from the posterior. To align with notation elsewhere in the thesis these are replaced by θ_x and θ_y.

6.4.1.1 Independence Chain Algorithm and Acceptance Rates

This section describes the Independence Chain (IC) algorithm, how the proposal distribution is chosen in SAS software, the information SAS software gives on acceptance rates, and the output it provides on the sampling from the posterior.

Littell et al. (2006) reference Tierney (1994) for the IC algorithm used in SAS software, who reference discussion of the idea in Hastings (1970). Tierney (1994) describes that it works by generating a pseudo-random proposal from a convenient proposal (base) distribution $g(.)$, chosen to be as close as possible to the posterior $f(.)$. Note in his paper Tierney uses the notation h for the proposal distribution g, and π for the posterior f. The proposal is then retained in the sample with probability proportional to the ratio of weights constructed by taking the ratio of the true posterior to the proposal density. If a proposal is not accepted, then a duplicate of the previous observation is added to the chain. The acceptance probability is given by:

$$\alpha(\theta_x, \theta_y) = \min\left(\frac{w(\theta_y)}{w(\theta_x)}, 1\right), \tag{6.1}$$

where $w(\theta_x) = \frac{f(\theta_x)}{g(\theta_x)}$, θ_x is the existing point in the chain and θ_y is the candidate point.

Linear combinations of variance components are created whose posterior density can then be approximated by a product of univariate inverted gamma densities. Inverse gamma

distributions are placed on these linear combinations for the proposal distribution (and for the inverse gamma priors). The introductory description to Section 6.4 described in more detail how the proposal distribution is selected. SAS software allows the user to input alternative inverse gamma parameters for the proposal distribution. For the batch sampling study the default linear transformations are σ_p^2, $P\sigma_k^2+\sigma_p^2$ and $KP\sigma_b^2+P\sigma_k^2+\sigma_p^2$.

At the end of the sampling, SAS software provides information on the acceptance rate for the proposals from the proposal(base) distribution. In the documentation details for the PRIOR statement syntax, SAS Institute Inc.: The MIXED Procedure (2013) P5290 refers to the acceptance rate as the number of accepted samples divided by the total number of samples generated. However, the output from the SAS software provides two acceptance rates, identifying them by "Boundary Constraints" and "Sampling". Proposals may be rejected due to:

- the proposal not meeting constraints on the parameters e.g. variances should not be negative. The proportion accepted is assumed to be the Boundary Constraint acceptance rate reported in the output.
- the proposal density not being the same as the true posterior distribution and thus the probability the proposal is retained, which is equal to the ratio of the weights constructed by taking the ratio of the true posterior to the base density (see Equation (6.1)), can be less than 1. The proportion accepted is assumed to be the Sampling acceptance rate provided in the output.

As mentioned above, the SAS software documentation indicates that the acceptance rate is computed as the number of accepted samples divided by the total number of samples generated. However, I do not think this is always correct when the analysis is bounded. I also note that for an example bounded analysis, Littell et al. (2006) P517 describe the "Acceptance rates" table in the SAS software output saying that for their example with boundary acceptance rate of 0.99 and sampling acceptance rate of 1.00 that 100% of the initial samples were accepted but 1% of these were dropped because they violated the non-negativity constraint on the original variance components. This reads that the boundary acceptance rate has as its denominator the number passing the sampling criteria (though since the sampling acceptance rate was 1.00 does not actually exclude the denominator being the total). This would suggest that the algorithm operates by evaluating the ratio of the densities before the constraints on the parameters. I do not think this is correct and that the software firstly applies the boundary constraint to proposals, followed by applying the sampling algorithm to those proposals that have passed the boundary constraint. The sampling acceptance rate is thus expressed not as those accepted divided by the total number of samples generated but divided by the number passing the boundary condition. Note that in the unbounded case there is no boundary and thus the denominator of the total number of samples or number of samples passing the boundary criteria is the same.

The reasoning for thinking that the boundary constraint is applied first to proposals, followed by applying the sampling algorithm to those proposals that have passed the boundary constraint is as follows. The default IG proposal distribution was found to be the same for both the bounded and unbounded analyses. If the resultant proposal distribution for the variance components has support for negative values, proposals with negative variance components are generated even if the analysis is specified to be bounded (else there would be no need to apply a boundary constraint to the proposals). In Section 6.4.3.4 Table 6.26 shows the acceptance rates for some bounded and unbounded analyses. For the proposal distribution identified as P10 it is seen that the proposal distribution has a sampling acceptance rate of 0.25 when the analysis is unbounded (with a boundary acceptance rate of 1) and it increases to 0.82 (with a boundary acceptance rate of 0.05) if the analysis is bounded. If the sampling acceptance rate had a denominator of all samples the rate would not be expected to increase when the analysis is bounded. Thus this suggests that the sampling acceptance rate is based on non-negative samples i.e. effectively the sampling acceptance rate only considers samples which have passed the boundary criteria. Another example is seen in the results in Table 6.24 in Section 6.4.3.4. A number of datasets are analysed with the same proposal distribution. If the sampling algorithm were applied first it would be expected that as the dataset varied so would the posterior and thus the accepted posterior samples would have varying proportions of negative proposal samples. However, the boundary acceptance rates are very consistent for a given proposal distribution suggesting that the boundary criteria are applied first.

If the acceptance rate is too low the SAS software stops sampling - see Section 6.4.3.

The output SAS dataset (see SAS Institute Inc.: The MIXED Procedure (2013) PROC MIXED PRIOR statement for details) contains the samples from the posterior together with log values of the posterior (denoted LOGF), the log of the base sampling density (denoted LOGG) and the log of their ratio (denoted LOGRATIO).

6.4.1.2 Other Algorithms

Three other sampling algorithms are also provided by SAS software in the PROC MIXED procedure. The option ALG=IS requests importance sampling, ALG=RS requests rejection sampling, and ALG=RWC requests a random walk chain. For more information on these techniques, the SAS documentation refers to Ripley (1987), Smith and Gelfand (1992), and Tierney (1994). These are now described.

Importance sampling Importance sampling has been used since the 40's and 50's and, for example, is described in Hammersley and Handscomb (1964), Section 5.4. Tierney (1994) (referenced by the SAS documentation) describes importance sampling as

drawing independent samples from a distribution similar to f. The sample is then weighted to make up for the difference between f and the distribution used to generate the sample. Tierney indicates that the function w (see Equation (6.1)) is the importance weight function that is used in importance sampling. Importance sampling is also described in Liu (1996) where the samples and their weights are then used to estimate the expected value of a quantity of interest. The dataset output from SAS software includes the values of the posterior, the log of the base sampling density and the log of their ratio (denoted LOGRATIO) as described in Section 6.4.1.1.

Rejection sampling Tierney (1994) describes rejection sampling chains, the idea originating from von Neumann (Eckhardt, 1987). In essence, rejection sampling usually takes a density g and constant c such that $f(\theta_x) \leq c\, g(\theta_x)$ for all θ_x. Pairs (θ_Z,U) are generated by generating θ_Z from g and U uniformly on the interval [0,$c\, g(\theta_Z)$] until a pair satisfying $U < f(\theta_Z)$ is obtained. Then an i.i.d. sample will have been obtained from f. Smith and Gelfand (1992) described how to generate a random sample from a distribution via the rejection method.

Tierney (1994) describes an extension which overcomes the difficulty in ensuring that c is large enough for $c\, g(\theta_x)$ to dominate $f(\theta_x)\ \forall \theta_x$ by adopting the following chain. Define $\mathrm{C} = \{\theta_x : f(\theta_x) \leq c\, g(\theta_x)\}$, then the Metropolis acceptance probability is:

$$\alpha(\theta_x, \theta_y) = \begin{cases} 1, & \text{for } \theta_x \in C, \\ \frac{c\, g(\theta_x)}{f(\theta_x)}, & \text{for } \theta_x \notin C, \theta_y \in C, \\ \min\left\{\frac{f(\theta_y) g(\theta_x)}{f(\theta_x) g(\theta_y)}, 1\right\}, & \text{for } \theta_x \notin C, \theta_y \notin C. \end{cases} \tag{6.2}$$

Thus if $c\, g(\theta_x)$ does not dominate $f(\theta_x)\ \forall \theta_x$ the algorithm occasionally rejects candidate steps when the chain is at a point $\theta_x \notin C$. This repeats the point θ_x within the sample path and thus compensates for the deficiency in the envelope at θ_x and will induce autocorrelation. However the following discussion of the SAS documentation suggests that this extension is not actually implemented and SAS restarts the sampling if the envelope is breached. In SAS software there is a LOGRBOUND=number option on the PRIOR statement. The documentation SAS Institute Inc.: The MIXED Procedure (2013) P5292 states "The value of number equals the maximum of $\log(f/g)$ over the variance component parameter space, where f is the posterior density and g is the product inverted gamma densities used to perform rejection sampling." Though the documentation does not explicitly state the density g used under the choice of algorithm, inverse gamma densities are referred to in the description of the LOGRBOUND option and parameters for an inverse gamma proposal distribution are output. Thus I assume this is used as density g. The LOGRBOUND number is $\log(c)$ as shown in Section 6.4.1.3 and by default is set so $\frac{c\, g(\theta_x)}{f(\theta_x)} \leq 1$ and $\frac{c\, g(\theta_x)}{f(\theta_x)} = 1$ for some θ_x. The output dataset contains values for LOGF, LOGG and LOGRATIO as described in Section 6.4.1.1.

The SAS software documentation also notes that "When performing the rejection sampling, you might encounter the following message: WARNING: The log ratio bound of LL was violated at sample XX. When this occurs, PROC MIXED reruns an optimisation algorithm to determine a new log upper bound and then restarts the rejection sampling. The resulting OUT= data set contains all observations that have been generated; therefore, assuming that you have requested N samples, you should retain only the final N observations in this data set for analysis purposes."

Random Walk Chain Tierney (1994) also describes random walk chains where θ_Y is generated by drawing Z independently from an "incremental distribution" and setting $\theta_Y = \theta_x + Z$. These ideas originated from Metropolis et al. (1953) and were generalised by Hastings (1970). Tierney (1994) refers to natural choices for the incremental distributions being uniform, normal or multivariate t distributions. The SAS software documentation does not seem to give further information on the implementation in PROC MIXED. However outputting the base distribution seems to imply it is a multivariate t distribution (see Section 6.4.1.3).

6.4.1.3 Investigations of Algorithms

The documentation provided by SAS was limited. Thus in order to further understand the implementation and output provided by SAS some investigations were performed.

The algorithms are investigated for a dataset (listed in Table E.1) where the variance components are not close to zero and thus boundary acceptance rates should not be an issue. The two-way variance components model in Section 4.2 was fitted to the response variable in PROC MIXED and 2,000 posterior samples using Jeffreys' prior were requested via sampling algorithms IC, IS, RS and RWC (using a seed of 198631).

For the IC and RS algorithms (with default LOGRBOUND) boundary and sampling acceptance rates of 1.00 were obtained for the example dataset. SAS software did not provide acceptance rates for the IS algorithm, though it might be expected the boundary acceptance rate would be relevant. The RWC algorithm gave boundary and acceptance sampling rates of 0.00 - this is further discussed below. For the RS algorithm the SAS software documentation indicates that LOGRBOUND=*number* specifies the bounding constant for rejection sampling where the value of *number* equals the maximum of $\log(f/g)$ over the variance component parameter space. By default this is the case but SAS software allows LOGRBOUND number to be specified in which case it is the log of the bounding constant c as given in the rejection sampling algorithm described in Section 6.4.1.2 and can be greater than the maximum of $\log(f/g)$ as is now discussed. For the example dataset the default LogBound used by the SAS software is -452.95 i.e. the default bounding constant $c_d = \exp(-452.95)$. Note these are values from SAS9.3

- the values reported for LOGG in the output dataset from SAS9.2 are incorrect (see SAS Institute Inc.: PN41931 (2010)). This is also the LOGRATIO value given for all proposals for the IC, IS and RS algorithms for this dataset as the posterior function is proportional to the proposal distribution. If the LogBound was reduced slightly (option LOGRBOUND=-455 was used) the warning message "WARNING: The log ratio bound of -455 was violated at sample 1" appeared as expected since $f(\theta_x) \leq cg(\theta_x)$ is not satisfied. If the LogBound was increased slightly (option LOGRBOUND=-451 was used) the procedure took longer to run than when using a value of -452.95, all 2000 samples were obtained and the sampling acceptance rate was 0.14. Using a value of -451 implies the ratio of this bounding constant to the default is $\frac{c}{c_d} = \exp(-451 - (-452.95)) = \exp(1.95) = 7.03$ which is consistent with the sampling acceptance rate of 0.14 ($\frac{1}{7.03}$) which was obtained, since the rejection sampling algorithm described in Section 6.4.1.2 requires $U < f(\theta_Z))$ when U has uniform distribution on the interval $[0, c\,g(\theta_Z)]$.

The values sampled from the posterior were then examined for the IC, IS and RS algorithms. Given that the dataset was chosen to be balanced and with variance components not close to 0, the proposal distribution was expected to be proportional to the posterior, and thus the IC algorithm would accept all samples and that the IS algorithm would give the same samples. Also if the default bound for the RS algorithm was used so that $f(\theta_x) = cg(\theta_x)$ then no samples should be rejected and the samples would be the same as the IC and IS algorithms. However, this was not the case. Table 6.12 shows the first 20 samples obtained via each algorithm with a seed of 198631. The values which are common between the algorithms are highlighted. It is seen that though both boundary and sampling acceptance rates were 1.00 for the IC and RS algorithms, the samples from the posterior are not the same for each algorithm. Some sample values are the same but some are missed from some algorithms and some additional ones interspersed. It is thus not entirely clear how the algorithms operate with respect to the stream of random numbers.

For the IC, IS and RS algorithms SAS software outputs that the base density (proposal) is type "ig" and all three produce the same values of the inverse gamma parameters for the proposal or "base" density. For the RWC algorithm SAS software outputs that the base density is of type "t" and the default base for the example dataset has 10 parameters with values: Parm1=0; Parm2=0; Parm3=0; Parm4=742.5; Parm5=-2.48; Parm6=4.9634; Parm7=0.0025; Parm8=-0.047; Parm9=0.7497; Parm10=4. There are three variance components, so it seems that Parm1-3 refer to the mean of the multivariate t, Parm4-Parm9 refer to the parameters of the variance covariance matrix (with Parm4,Parm6 and Parm9 being the variances) and Parm10 is the degrees of freedom (note no documentation was found to confirm this).

As mentioned above requesting 2000 samples (using a seed of 198631) gives boundary and sampling acceptance rates of 0.00. Using a seed of 198631 the posterior sample dataset contains 2000 values but they are all numbered as sample 0, σ_b^2 is 14.482 for

TABLE 6.12: **Posterior samples for σ^2_{batch} produced by IC, IS and RS sampling algorithms**

IC		IS		RS	
sample #	**value**	**sample #**	**value**	**sample #**	**value**
1	46.106	1	46.106	1	46.106
2	30.266	2	30.266	2	31.062
3	35.499	3	22.090	3	22.090
4	24.063	4	83.723	4	19.244
		5	70.380	5	64.737
5	10.380	6	10.380	6	104.534
6	26.163	7	52.803	7	34.416
7	173.558	8	21.824	8	14.994
8	27.630	9	28.011		
9	41.780	10	36.335	9	41.780
10	97.590	11	63.453	10	15.425
11	207.05	12	107.514	11	107.514
12	41.329	13	122.043	12	30.112
13	59.043	14	59.043	13	59.043
14	28.393	15	28.393	14	55.572
15	28.444	16	28.444	15	21.125
16	31.236	17	29.236	16	61.120
17	25.970	18	21.233		
18	40.015	19	40.015	17	40.015
19	45.908	20	64.053	18	45.908
20	108.222			19	40.754
				20	26.149

the first sample and -467.784 for the remainder, σ^2_k is 0.46565 for all samples and σ^2_p is 5.44499 for all samples. The output dataset also shows LOGF=0 for all observations including the first. This does not seem correct and is not consistent with the output from other sampling algorithms where LOGF is in the region of -465. Despite this an attempt was made to change the incremental (base) density used. There didn't seem to be an option to change the step size but the parameters of the incremental distribution could be changed. The covariance parameters (Parm4-Parm9) of the default base density were scaled smaller and larger between 0.00001 and 1000 times the original and the degrees of freedom (Parm10) were varied (taking values 1,4,10,100). The covariance parameters were also set to small values, Parm4=Parm6=Parm9=0.001 and Parm5=Parm7=Parm8=0. None of these changes made to the parameters of the proposal distribution for the RWC gave a sampling acceptance rate greater than 0.00.

Note for a mean model Y $\sim$ N(μ,σ^2) when IC algorithm is specifically requested or when the default sampling algorithm is used (stated as being IC) the sampling algorithm is reported as being Importance Sampling (IS) in the SAS software output. However, as for the two-way variance components model the IC sampling algorithm is reported when it is requested (or by default) this was not investigated further.

6.4.1.4 Choice of Sampling Algorithm

Liu (1996) showed that, unless the proposal distribution is the posterior distribution and c=1, rejection sampling is less efficient than Metropolized independent sampling (independent sampling is denoted as IC in this thesis and PROC MIXED). Note that rather than requiring c=1 for rejection sampling to be as efficient, using the terminology in Section 6.4 we only require $f(\theta_x) = cg(\theta_x)$ where $\frac{f(\theta_x)}{c}$ is the posterior probability density function. Liu showed that importance sampling (as defined by Liu) is asymptotically more efficient than rejection sampling in most cases, it can be very efficient depending on the quantity of interest. Liu states that the advantage of rejection sampling is that the distribution of samples is always the desired one, whereas that for importance sampling is always biased from the target and the distribution of that obtained from the Metropolized independent sampling theoretically converges to the target when sample sizes grow to infinity, but practically differs from the target distribution. Liu indicates that the advantage of importance sampling and IC sampling is that the constant c is not required.

The IS algorithm was not considered advantageous over IC given that summary statistics from the posterior are unlikely to be a simple mean and thus their estimation using posterior samples accompanied by weights may be difficult. The RWC provides a correlated sample from the posterior. Given this and the problems which were seen in Section 6.4.1.3 in trying to obtain a sample in the simple balanced two-way variance components example with variance components not close to zero the RWC algorithm seemed unlikely to offer advantages over the default IC algorithm. If the proposal distribution is the same as the posterior distribution the IC and RS algorithms should not have an advantage over each other. However if the distributions differ, the IC algorithm is likely to offer speed advantages (no samples are rejected though more will be required given the autocorrelation, and the RS algorithm may start again if the bounding constant used is found actually not to be a bound) whilst the RS algorithm offers independent samples.

For the investigations in the rest of this chapter the default IC sampling method is used, with the exception in Section 6.4.3.4 when investigating modifications to the analysis the effect of changing the sampling method was examined.

6.4.2 Initial Choice of Number of Samples from Posterior

A number of samples from the posterior consistent with at least 2000 independent samples was sought. This is aligned with the rationale in Section 6.1 for choosing to evaluate at least 2000 datasets and also with that in Section 6.3 for those priors evaluated using the Gibbs sampling algorithm. Since the proposal distribution used by the independence chain sampling algorithm are expected to be approximately the same as the posterior

distribution of the linear combination of the variance components, the sampling acceptance probability is expected to be close to 1. Thus there should be few duplicates of previous observations and the required number of samples was not expected to be much above 2000. Thus initially 2500 samples were requested from the posterior.

6.4.3 Exploring when a Full Posterior Sample is Not Obtained

If the acceptance rate is too low the algorithm in SAS software stops sampling from the posterior with the warning message in the program log "WARNING: Sampling stopped because total acceptance rate is too low.". Sections 6.4.3.1 - 6.4.3.3 examine when and how often this occurs and aspects of the analysis which may make it more or less likely. Section 6.4.3.4 explores whether changes to the analysis can result in a full sample. Note datasets where sampling from the posterior stopped are referred to as "stopped" datasets. The conclusions and decisions made are summarised in Section 6.4.3.5.

6.4.3.1 Examining Number of Samples Obtained

Two thousand five hundred samples were requested from the posterior for the analysis of each dataset. Ten thousand datasets were analysed for each design scenario described in Section 4.3 and values of σ_b^2=0.5, 6 and 24. These were analysed in PROC MIXED using the various inverse gamma, JEFF and FLAT priors as described in Sections 5.2.4-5.2.6. 2500 samples were obtained from the posterior for the informative priors I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, and IhE6 for scenario k2p16_0.5 and σ_b^2=0.5, 6 and 24. The remaining priors had some datasets for which 2500 samples were not obtained from the posterior for some scenarios and the number of datasets where 2500 posterior samples were not obtained is given in Table 6.13 for each prior, scenario and batch variance component.

It is seen that when the true batch variance component is smallest (0.5) the number of stopped datasets is greatest. In general increasing the true batch variance component decreases the number of stopped datasets. This was not seen when the keg variance component was increased for the design with 2 kegs/batch and 16 portions/keg. The smaller designs with 2 portions per keg (k2p2_0.5 and k6p2_0.5) have a higher number of stopped datasets than the designs with 16 portions per keg (k2p16_0.5 and k6p16_0.5). There are also differences between the priors. The FLAT prior applied to the datasets from the b3k2p2_0.5 scenario has most datasets with fewer than the full set of posterior samples. This may be due to the posterior being improper (see Section 5.2.6). However for the other scenarios the FLAT prior has least datasets with fewer than 2000 samples from the posterior. Aside from the FLAT prior for the b3k2p2_0.5 scenario, the IG0 prior had most datasets with fewer than 2500 samples from the posterior. The other priors had broadly similar numbers of datasets with fewer than 2500 samples.

TABLE 6.13: **Number of analysed datasets with < 2500 samples**

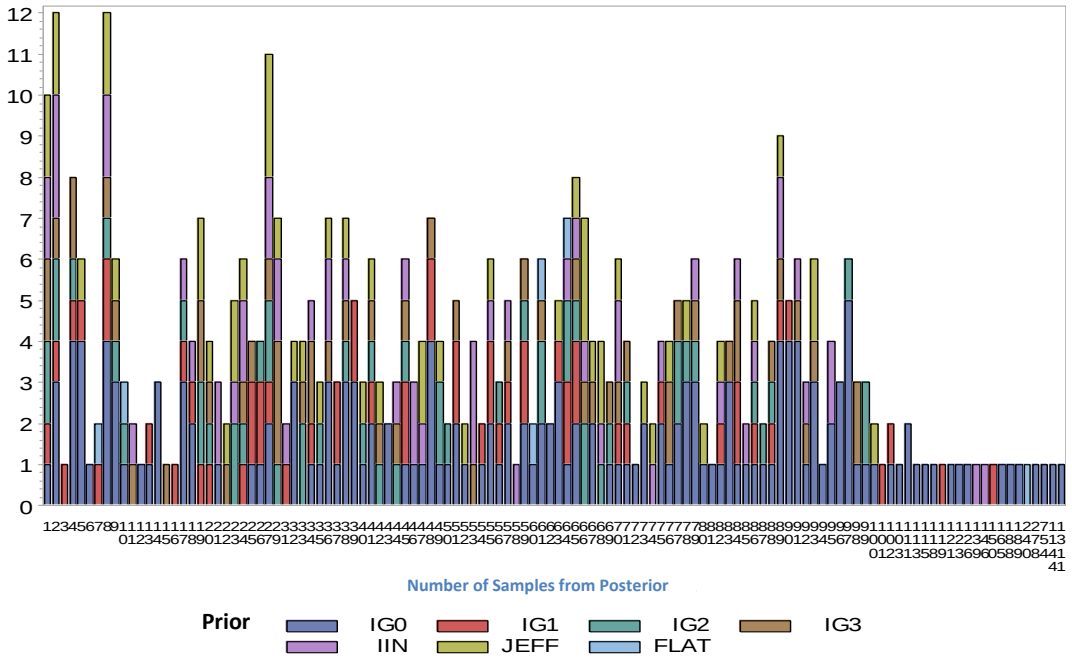
		Scenario						
Prior	σ_b^2	b3k2p2 _0.5	k2p2 _0.5	k2p16 _0.5	k2p16 _6	k2p16 _24	k6p16 _0.5	k6p2 _0.5
	0.5	1359	903	110	215	261	14	257
IG0	**6**	572	203	31	35	146	0	33
	24	285	134	26	3	17	0	36
	0.5	688	479	36	94	95	3	120
IG1	**6**	292	121	14	10	47	0	26
	24	153	82	9	1	6	0	30
	0.5	633	442	31	84	88	3	112
IG2	**6**	276	114	14	10	45	0	25
	24	146	78	8	1	5	0	29
	0.5	624	440	32	80	84	3	113
IG3	**6**	272	113	14	10	44	0	25
	24	149	80	8	1	5	0	29
	0.5	833	580	34	80	83	3	138
IIN	**6**	358	151	14	10	44	0	31
	24	199	108	8	0	5	0	36
	0.5	620	440	30	81	83	3	110
JEFF	**6**	273	113	14	10	43	0	25
	24	151	80	8	0	5	0	28
	0.5	8347	58	5	8	8	0	32
FLAT	**6**	6585	10	1	0	7	0	20
	24	4657	9	0	0	0	0	22

The number of samples actually achieved from the posterior was examined. Figure 6.10 shows the frequency of the number of samples achieved for stopped datasets from the original batch sampling design (6 batches, 2 kegs/batch, 16 portions/keg and σ_k^2=0.5) classified by the prior and pooled across datasets with σ_b^2 of 0.5, 6, 24. Note that the plot is not a histogram - only those numbers of samples which actually occur are plotted on the x axis.

In Figure 6.11 the frequency of the number of samples achieved for stopped datasets are shown for various design scenarios and in Figure 6.12 the classification is by prior. The results for the FLAT prior are excluded and plotted separately in Figures 6.13 and 6.14. The plots are in two parts for clarity with the x axis being divided into 0-149 and 150 to 2499 samples.

From Figures 6.10, 6.11 and 6.12 it is seen that most of the datasets without a full sample stopped with fewer than 100 samples from the posterior. The reduced number stopping above 100 suggests not only that those with very low acceptance rates are likely to be stopped with number of samples $\leq$ 100 but also that the algorithm for deciding when

FIGURE 6.10: **Distribution of number of samples from posterior of stopped datasets for original batch sampling design pooled over σ^2_{batch}**



to stop takes into consideration when the number of samples from the posterior reaches 100. Figure 6.13 shows results for the FLAT prior for the various design scenarios. Given that the smallest design with 3 batches (b3k2p2_0.5) had a much higher number of datasets stopping, for clarity this was plotted separately in Figure 6.14. Figure 6.14 splits the x axis into number of samples < 100 and ≥ 100 for clarity. It is seen that for the smallest design which has much higher number of datasets stopping, the pattern appears different with very high spikes at some numbers greater than 100.

FLAT prior: 2498 and 2499 samples It is seen that there is a very large spike at 2499 samples. For datasets where 2498 or 2499 posterior samples were obtained there was no SAS software warning indicating that sampling had stopped. The datasets with < 2500 samples seen in Figure 6.14 for the FLAT prior and scenario b3k2p2_0.5 are summarised in Table 6.14 split by σ^2_b, together with a count of those datasets from which 2500 samples were obtained. Analyses for all three true batch variance components had datasets where SAS software only provided 2498 or 2499 samples and whilst the number of datasets was smaller for σ^2_b=24, so were the number of datasets stopping.

Ten datasets with 2499 posterior samples were re-run with a request for 4000 samples from the posterior and 3999 samples were obtained, again with no warning. Thus it was confirmed that there wasn't an issue with these datasets, just that the algorithm hadn't quite given the requested number of samples. In the original analyses with the default

FIGURE 6.11: **Distribution of number of samples from posterior of stopped datasets for various scenarios pooled over σ^2_{batch} and priors investigated (excluding FLAT prior)**

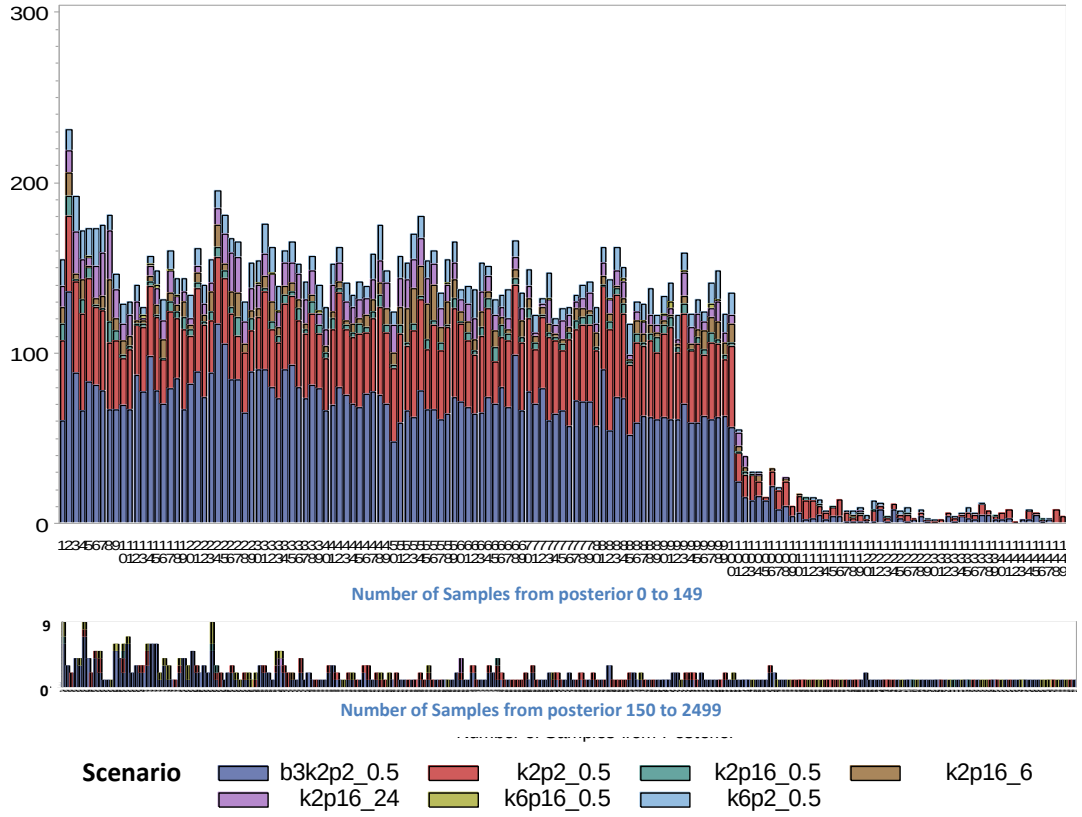
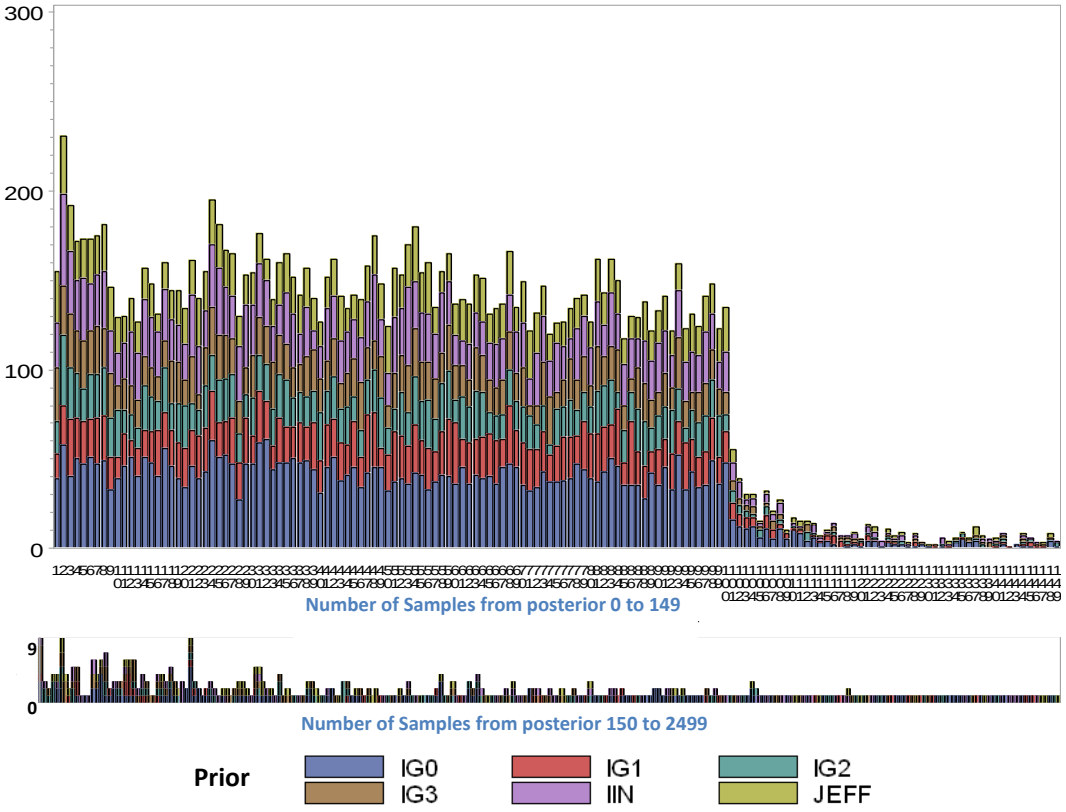


TABLE 6.14: **Number of datasets for FLAT prior for 3 batch design categorised by σ^2_{batch} and number of samples from posterior**

Number of Samples	σ^2_b 0.5	6	24
≤1100	4816	2957	2250
>1100 and <2498	0	0	0
2498	134	43	8
2499	3397	3585	2399
2500	1653	3415	5343

proposal distributions the sampling acceptance rates are almost entirely 1 except for the FLAT prior for scenario b3k2s2_0.5. Thus a plausible hypothesis was that the stopping with 1 or 2 samples less than requested was in some way related to this. However, when the proposal distributions were modified and the sampling acceptance rates were less than 1 it was seen that only the FLAT prior for scenario b3k2s2_0.5 has samples with 1 or 2 less than the number requested. Thus no explanation was found though it is noted that for the FLAT prior for scenario b3k2s2_0.5 the posterior is improper.

FIGURE 6.12: **Distribution of number of samples from posterior of stopped datasets for various priors (excluding FLAT prior) pooled over σ^2_{batch} and design scenarios investigated**



FLAT prior: Number of samples for stopped datasets For scenario b3k2s2_0.5, from 0 to 100 samples the number of datasets stopping is fairly uniform (see Figure 6.14) but then there are large spikes at 101,201,301,401,501,601,901 samples with smaller spikes at 200, 300, 400, 500, 600, 900. This suggests that the monitoring of the sampling rate is not continuous but predominantly checks after groups of samples have been obtained. The smaller spikes at 1 fewer samples than the larger spikes together with datasets producing 2499 samples rather than the requested 2500 despite not being stopped, suggests an anomaly where the algorithm counting the sample misses by one. No explanation was found.

In the forthcoming evaluation of the priors the posterior estimates from the reduced sample size could be used or the dataset excluded from the evaluation. In either case, given the number of stopped datasets is fairly high and stopping is related to the variance component size and design size, it is likely to substantially affect the evaluation. Thus further work was performed to understand why the sampling stopped and possible strategies for obtaining a full sample from the posterior. Given that a Bayesian analysis of the dataset from an actual study may also incur the problem of not gaining a full

FIGURE 6.13: **Distribution of number of samples from posterior of stopped datasets for FLAT prior and various scenarios pooled over σ^2_{batch}**

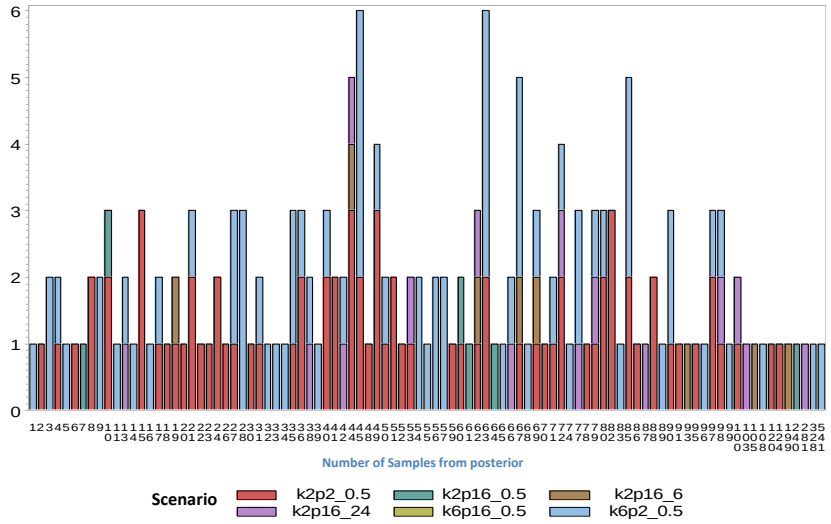
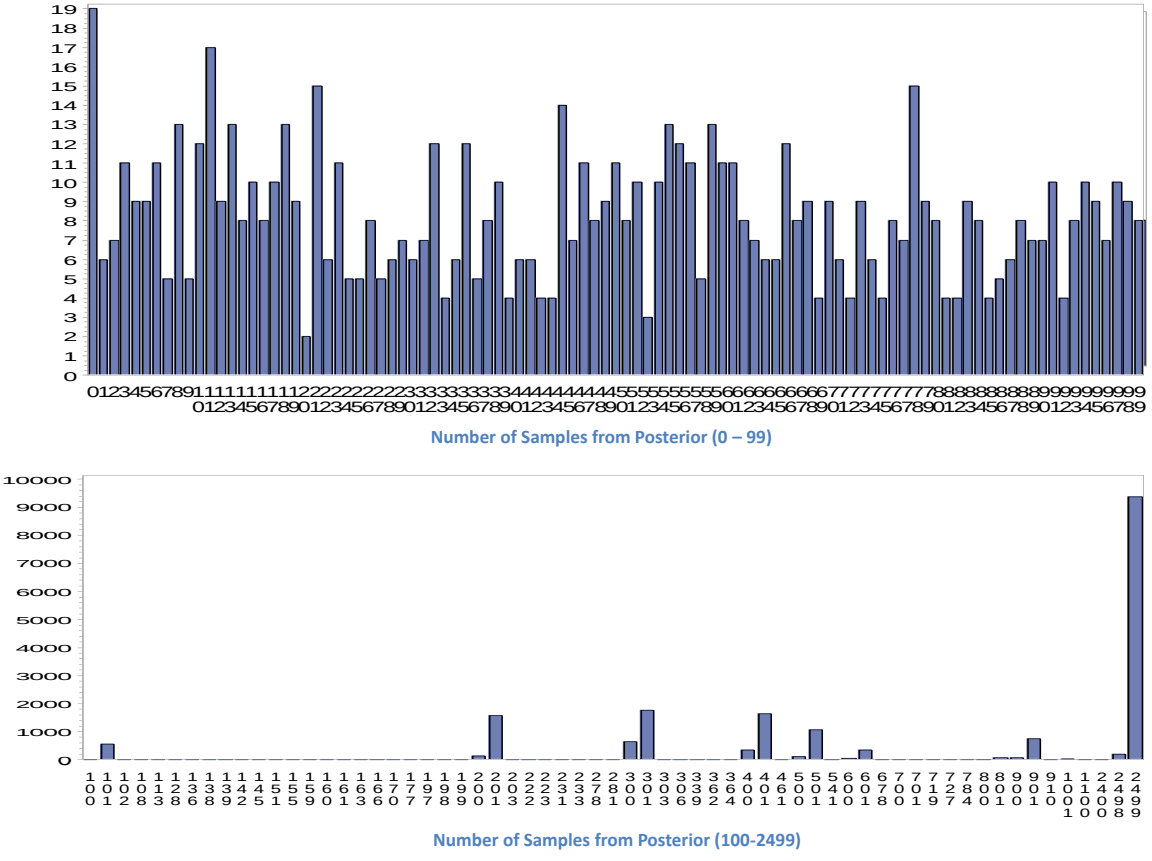


FIGURE 6.14: **Distribution of number of samples from posterior of stopped datasets for FLAT prior for 3 batch design pooled over σ^2_{batch}**



sample, any solutions may also be useful in that situation. Given that it was seen above that SAS software considered it had provided the requested sample when only 2498 and 2499 samples had been obtained from the posterior (i.e. no warning message concerning stopping was given and if more samples were requested they were provided), in the subsequent investigations a posterior sample of the requested number of samples or up to two fewer was considered a full dataset and not a stopped dataset.

6.4.3.2 Investigating Acceptance Rates

SAS software provides information on the acceptance rates for the proposals from the base distribution. Proposals may be rejected due to the base density not being the same as the true posterior distribution as described in Section 6.4.1.1 and detailed in Equation (6.1) (sampling acceptance rate) or the proposal not meeting constraints on the parameters (boundary acceptance rate).

Since SAS software stops sampling due to low acceptance rates this section investigates the acceptance rates to see how they are affected by the design scenario, batch variance and prior applied. Figure 6.15 shows box plots summarising the boundary acceptance rates for the various priors and for values of σ^2_{batch} of 0.5, 6 and 24 over the 10,000 datasets evaluated using the default proposal distribution. This was evaluated for the scenario of 6 batches, 2 kegs per batch and 16 portions per keg, σ^2_{keg}=0.5 and $\sigma^2_{portion}$=6 for the default proposal distribution. It is seen that in most cases the median acceptance rate is above 80% and many of the 25% quartile values are above 70%. The acceptance rate is lowest for σ^2_{batch} =0.5 and increases as σ^2_{batch} increases (which is perhaps to be expected as the σ^2_{batch} values become further away from zero). The mildly informative priors on the higher variance components generally have higher acceptance rates. Again this is perhaps to be expected as the implied marginal priors for σ^2_{batch} and σ^2_{keg} shown in Figures 5.10 and 5.11 have greater density for positive values compared with those seen in Figures 5.6-5.9. Even where the median is high, most prior/ σ^2_{batch} combinations have some datasets where the acceptance rate is low (the bottom of the whisker indicating the min is close to zero).

Boundary acceptance rates were also evaluated for the various scenarios which varied the numbers of kegs per batch and portions per keg and size of keg variance component. These are shown in Figures 6.16-6.17 for a subset of the priors shown in Figure 6.15 and batch variance components of 0.5, 6, and 24. In Figure 6.16 which compares designs with varying number of kegs/batch and portions/keg for σ^2_{keg}=0.5, it is seen that increasing the number of portions/keg from 2 to 16 increases the boundary acceptance sampling rates and increasing the kegs/batch from 2 to 6 increases the boundary acceptance rates. However, it is noted that for σ^2_{batch}=6 and 24 it is not until the design has 6 kegs/batch and 16 portions/keg that there are no datasets with a sampling rate < 0.05 and for σ^2_{batch}=0.5 there are still some datasets with boundary acceptance rates < 0.05. In

FIGURE 6.15: **Box plot summary of boundary acceptance rates for original batch sampling design (median, interquartile range and range with line joining medians)**

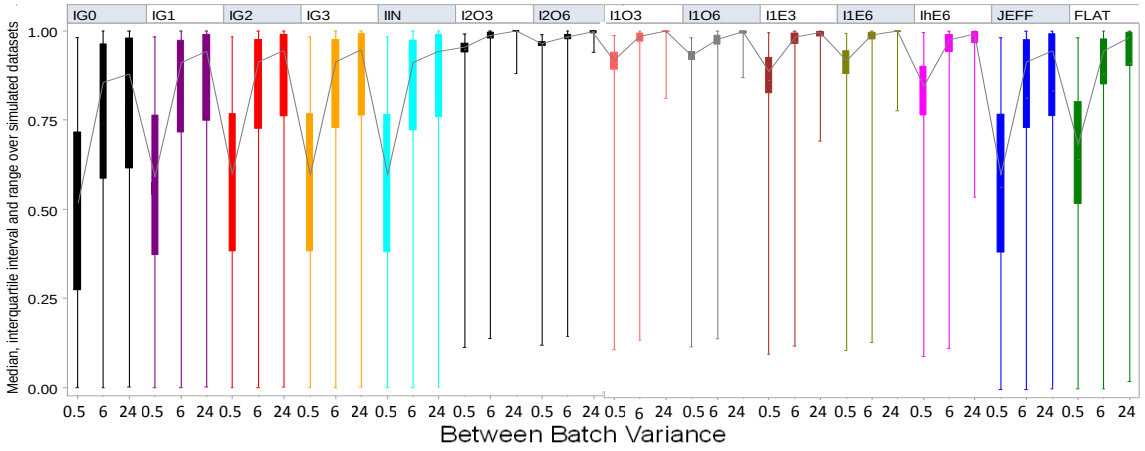
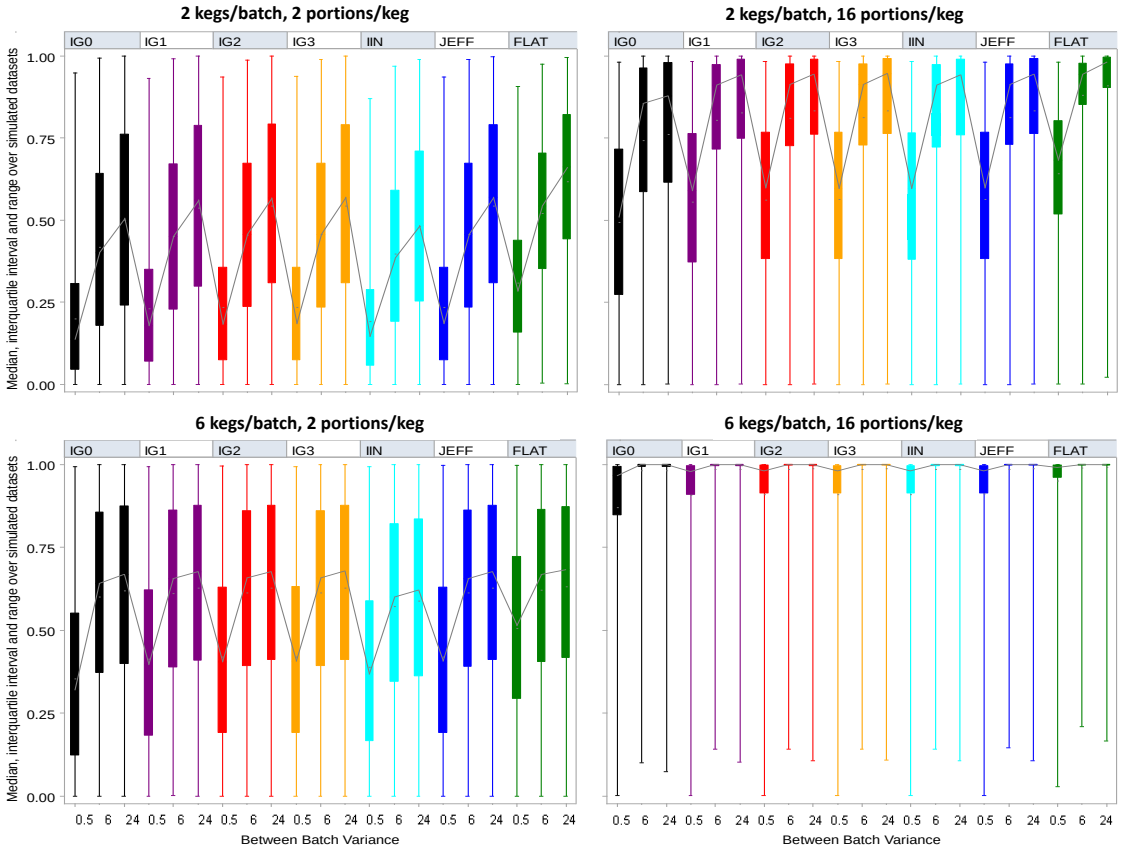


Figure 6.17 which shows results for the design with 2 kegs/batch and 16 portions/keg it is seen that there are some differences between the boundary acceptance rates when changing the keg variance component between 0.5, 6 and 24. σ^2_{keg}=6 tends to have slightly higher acceptance rates compared to σ^2_{keg}=0.5 and σ^2_{keg}=24, this being most pronounced for σ^2_{batch}=24. Figure 6.18 shows the effect of reducing the number of batches from 6 to 3 for designs with 2 kegs/batch and 16 portions/keg and σ^2_{keg}=0.5, where 3 batches tended to have slightly lower rates with the effect being more pronounced for the FLAT prior where the posterior will be improper. For all scenarios looked at in Figures 6.16-6.17, priors IG1, IG2, IG3, IIN and JEFF had similar boundary acceptance rates whilst the IG0 prior often had slightly lower rates and the FLAT prior had slightly higher, except for design b3k2s2_0.5 where the posterior is improper.

In contrast to the boundary acceptance rates, the sampling acceptance rates are typically one. They were evaluated for the various priors and for values of σ^2_{batch} of 0.5, 6 and 24 over the 10,000 datasets evaluated. This was evaluated for the scenario of 6 batches, 2 kegs per batch and 16 portions per keg, σ^2_{keg}=0.5 and $\sigma^2_{portion}$=6 and are summarised in Figure 6.19. Here it is seen that only the lower whiskers of the box plots are visible indicating that in most cases the sampling acceptance rate is 1 and thus the proposal (base) distribution is close to the posterior distribution.

Sampling acceptance rates were also evaluated for the various scenarios which varied the numbers of kegs per batch and portions per keg and size of keg variance component and are shown in Appendix E.3. Increasing the number of kegs/batch from 2 to 6 increases the sampling acceptance rates to the extent that for those evaluated, the only datasets with sampling acceptance rates less than 1 are those for a batch variance component of 0.5 and changing the keg variance component from 0.5 to 6 removes even the few

FIGURE 6.16: **Boundary acceptance rates with varying number of kegs and portions (median, interquartile range and range with line joining medians)**



datasets with low acceptance rates. Reducing the number of portions/keg from 16 to 2 or decreasing the number of batches from 6 to 3 decreases the sampling acceptance rates for a few datasets except in the case of the FLAT prior where with only 3 batches, most datasets have a low sampling acceptance rate.

FIGURE 6.17: **Boundary acceptance rates with varying σ^2_{keg} (median, interquartile range and range with line joining medians)**

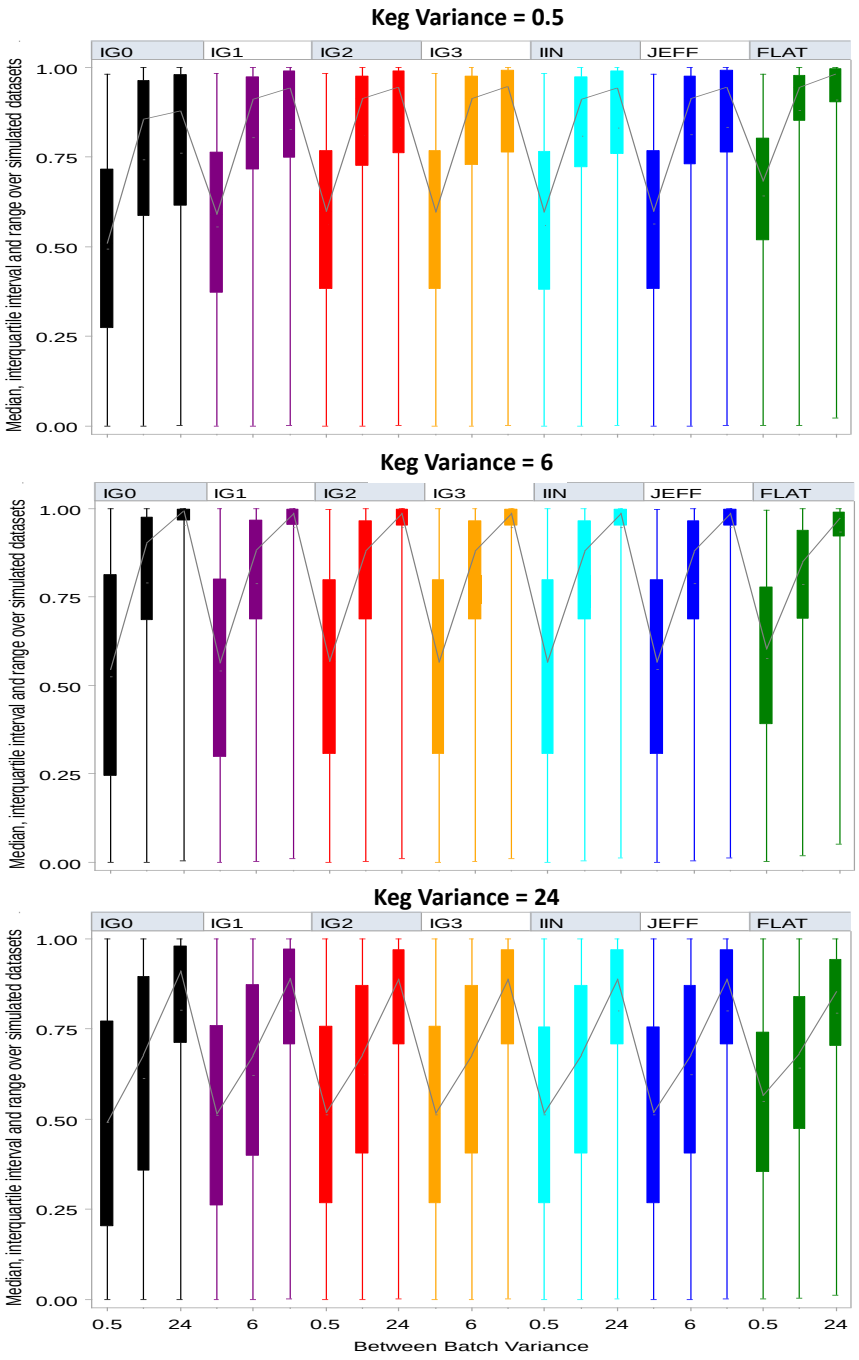


FIGURE 6.18: **Boundary acceptance rates with varying number of batches (median, interquartile range and range with line joining medians)**

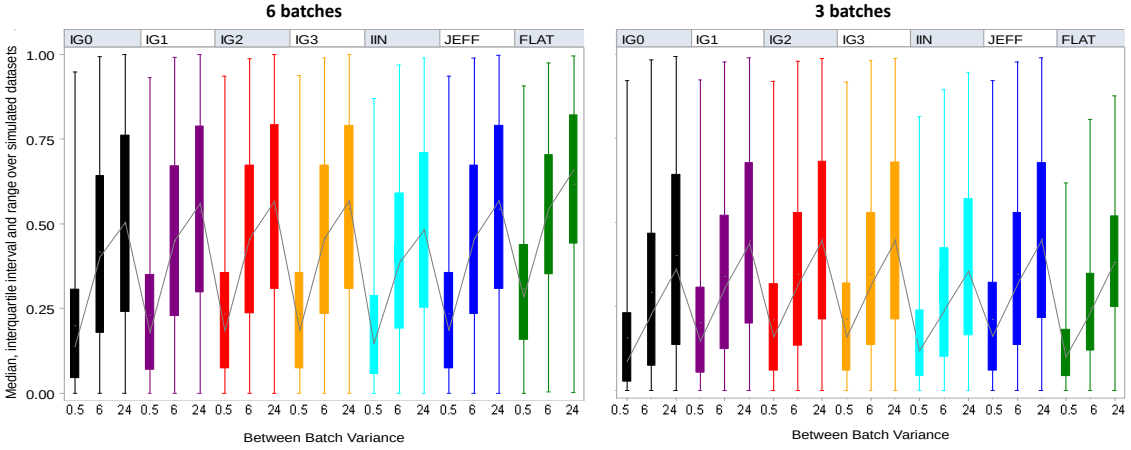
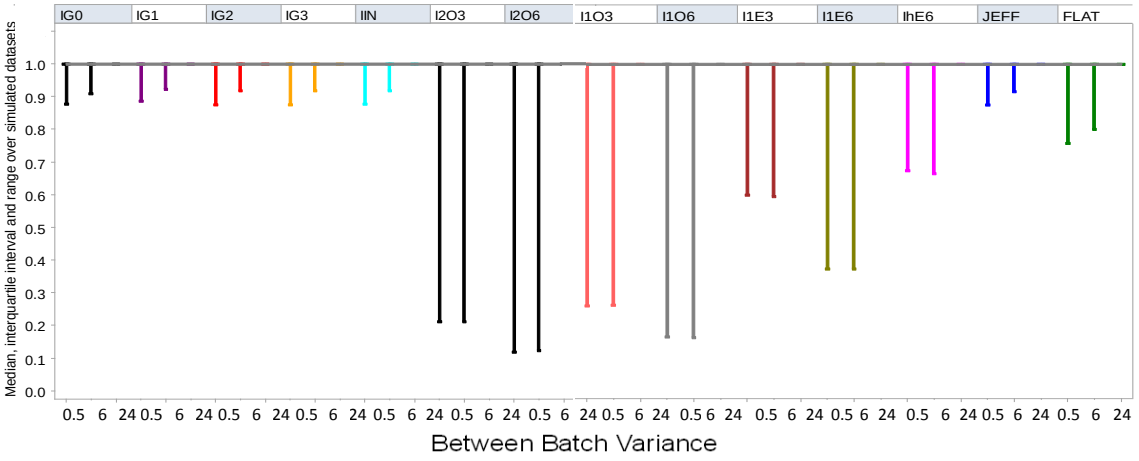


FIGURE 6.19: **Box plot summary of sampling acceptance rates for original batch sampling design (median, interquartile range and range with line joining medians)**



6.4.3.3 Association of Stopping with REML Variance Component Estimates and Posterior Median Estimates

This section examines whether sampling from the posterior is related to the values of the REML estimates or posterior median estimates. The number of stopped datasets for the scenarios were previously summarised in Table 6.13. The design with 2 portions/keg and 2 kegs/batch which had lower boundary acceptance rates (see Figure 6.16) was chosen for further investigation in conjunction with a selection of priors. Given that variance components are restricted to be ≥ 0 it seemed likely that stopping would be associated with datasets where the REML estimates of the variance components are 0 or very small.

TABLE 6.15: **Number of datasets categorised by whether the REML estimates are practically zero ($< \delta$=0.0001) or not, and whether sampling from posterior stopped**

		Sampling from Posterior							
		Full Sample				Stopped			
		Batch Var Estimate							
True Batch Variance		$< \delta$		$\geq \delta$		$< \delta$		$\geq \delta$	
		Keg Var Estimate							
Component	**Prior**	$< \delta$	$\geq \delta$	$< \delta$	$\geq \delta$	$< \delta$	$\geq \delta$	$< \delta$	$\geq \delta$
	IG0	1477	2488	2698	2434	655	109	139	0
	IG1	1776	2545	2766	2434	356	52	71	0
	IG2	1799	2551	2774	2434	333	46	63	0
0.5	**IG3**	1804	2551	2771	2434	328	46	66	0
	IIN	1707	2538	2741	2434	425	59	96	0
	JEFF	1800	2552	2774	2434	332	45	63	0
	FLAT	2084	2594	2830	2434	48	3	7	0
	IG0	192	461	3991	5153	61	7	135	0
	IG1	222	463	4041	5153	31	5	85	0
	IG2	227	463	4043	5153	26	5	83	0
6	**IG3**	227	464	4043	5153	26	4	83	0
	IIN	216	463	4017	5153	37	5	109	0
	JEFF	228	463	4043	5153	25	5	83	0
	FLAT	252	466	4119	5153	1	2	7	0
	IG0	9	45	4250	5562	2	1	131	0
	IG1	10	45	4301	5562	1	1	80	0
	IG2	10	45	4305	5562	1	1	76	0
24	**IG3**	10	45	4303	5562	1	1	78	0
	IIN	10	45	4275	5562	1	1	106	0
	JEFF	10	45	4303	5562	1	1	78	0
	FLAT	11	46	4372	5562	0	0	9	0

Table 6.15 examines whether the REML estimates were approximately zero ($< \delta$ where δ=0.0001) or not for datasets from which either the full sample was obtained or sampling

stopped for the scenario of 6 batches, 2 kegs per batch and 2 portions from each keg and a between keg variance component of 0.5. It is seen that all of the datasets which stopped had a batch and/or a keg variance component REML estimate which was less than δ. It is also seen however, that there are many datasets with a keg and/or batch variance component REML estimate of less than δ which did not stop. There were fewer datasets with stopped sampling for the FLAT prior compared to the other priors.

As expected, when the true batch variance component is increased to 6 or 24 the number of datasets with a batch variance component REML estimate of zero also decreases. Hence to enable comparison across the four groups generated by the two way classification of whether batch or keg variance component REML estimates were $< \delta$ or $\geq \delta$ where δ= 0.0001, Table 6.16 shows the percentage of datasets in each group for which sampling from the posterior stopped. It is seen that if both the batch and the keg variance component are estimated as approximately zero (noting that for a true batch variance component of 24 there are few datasets where this occurs so the estimate of the percentage will be imprecise), sampling is considerably more likely to stop.

TABLE 6.16: **Percentage of datasets where sampling from posterior stopped ($\delta = 0.0001$)**

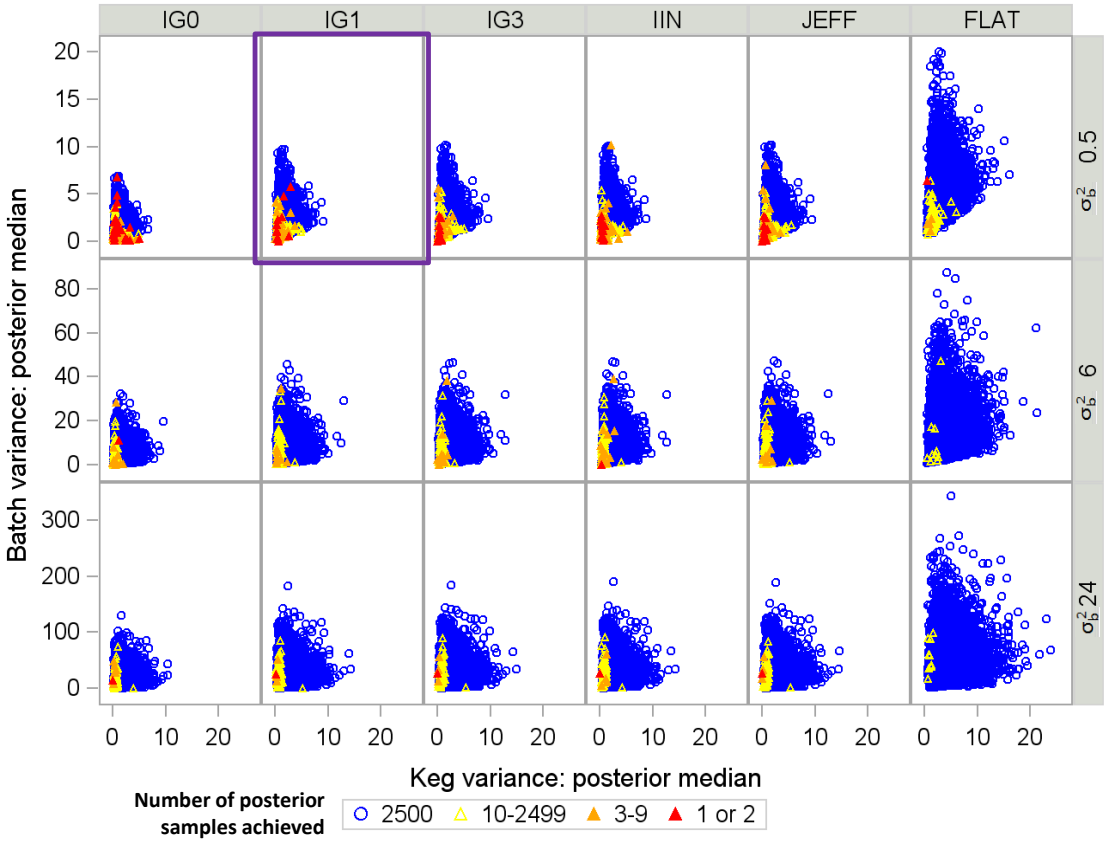
True Batch Variance Component	**Prior**	**Batch Variance Estimate** $< \delta$ **Keg Variance Estimate** $< \delta$	$\geq \delta$	$< \delta$	$\geq \delta$
	IG0	30.7	4.2	4.9	0
	IG1	16.7	2.0	2.5	0
	IG2	15.6	1.8	2.2	0
0.5	IG3	15.4	1.8	2.3	0
	IIN	19.9	2.3	3.4	0
	JEFF	15.6	1.7	2.2	0
	FLAT	2.3	0.1	0.2	0
	IG0	24.1	1.5	3.3	0
	IG1	12.3	1.1	2.1	0
	IG2	10.3	1.1	2.0	0
6	IG3	10.3	0.9	2.0	0
	IIN	14.6	1.1	2.6	0
	JEFF	9.9	1.1	2.0	0
	FLAT	0.4	0.4	0.2	0
	IG0	18.2	2.2	3.0	0
	IG1	9.1	2.2	1.8	0
	IG2	9.1	2.2	1.7	0
24	IG3	9.1	2.2	1.8	0
	IIN	9.1	2.2	2.4	0
	JEFF	9.1	2.2	1.8	0
	FLAT	0	0	0.2	0

Given that, though most datasets for which sampling from the posterior stopped had an approximately zero REML variance component estimate, many more with approximately zero REML variance components did not stop, the medians of the posterior samples for the variance components were also examined.

Figure 6.20 plots the posterior median for the batch variance component against the posterior median for the keg variance component for various priors and batch variance components of 0.5, 6, and 24. The datasets are identified using symbols to show whether a full sample of 2500 was achieved or smaller number of samples.

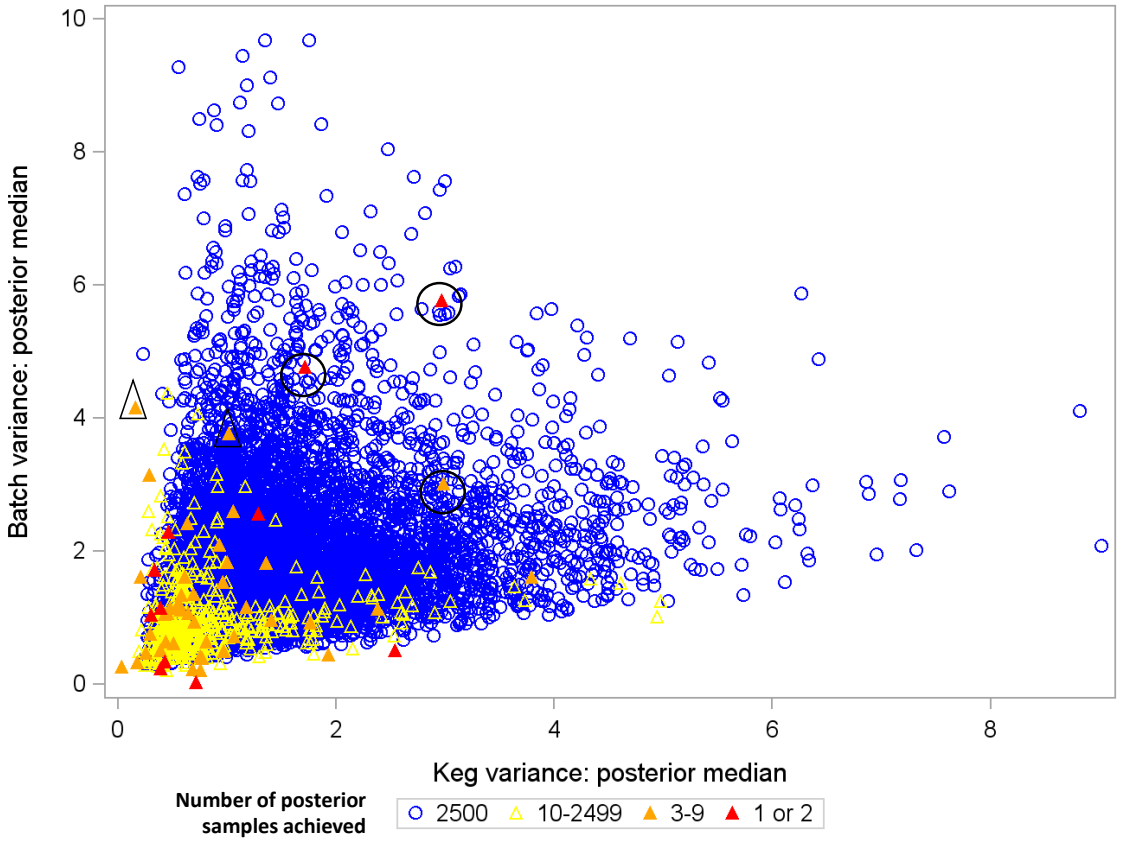
Figure 6.20 shows that those stopping due to boundary conditions do generally have smaller estimates on one or other of the batch or keg variance components. When the batch variance component investigated is higher than 0.5 most of the analyses which stop have smaller estimates for the keg variance component, rather than the batch variance component. If the larger design with 2 kegs/batch and 16 portions/keg is examined (Appendix E.4, Figure E.4) it has a smaller number of datasets stopping and their posterior median encompassing a smaller range of values.

FIGURE 6.20: **Scatterplot of posterior median estimates of σ^2_{batch} against σ^2_{keg} coloured by the number of posterior samples by prior and σ^2_{batch}**



The graph outlined in Figure 6.20 for the IG1 prior and batch variance component of 0.5 is shown in Figure 6.21. Firstly it is noted that, due to the overplotting of symbols, Figure 6.20 may visually overemphasise those datasets with low numbers of samples from the posterior, though the plots are still very useful for comparisons across priors and across batch variance components. It is seen that whilst almost all datasets which stop tend to have low posterior median batch and/or keg variance there are three datasets circled which appear outlying from the stopping datasets. These datasets have stopped very early. Note it cannot be concluded from the scatterplot in Figure 6.21 that the distribution of the points does differ according to whether sampling has stopped or not due to the high density of the plotted points. Confirmation of this is provided in a bivariate density plot of the data in Figure E.5 in Appendix E.4.

FIGURE 6.21: **Scatterplot of posterior median estimates of σ^2_{batch} against σ^2_{keg} for IG1 prior coloured by the number of posterior samples**

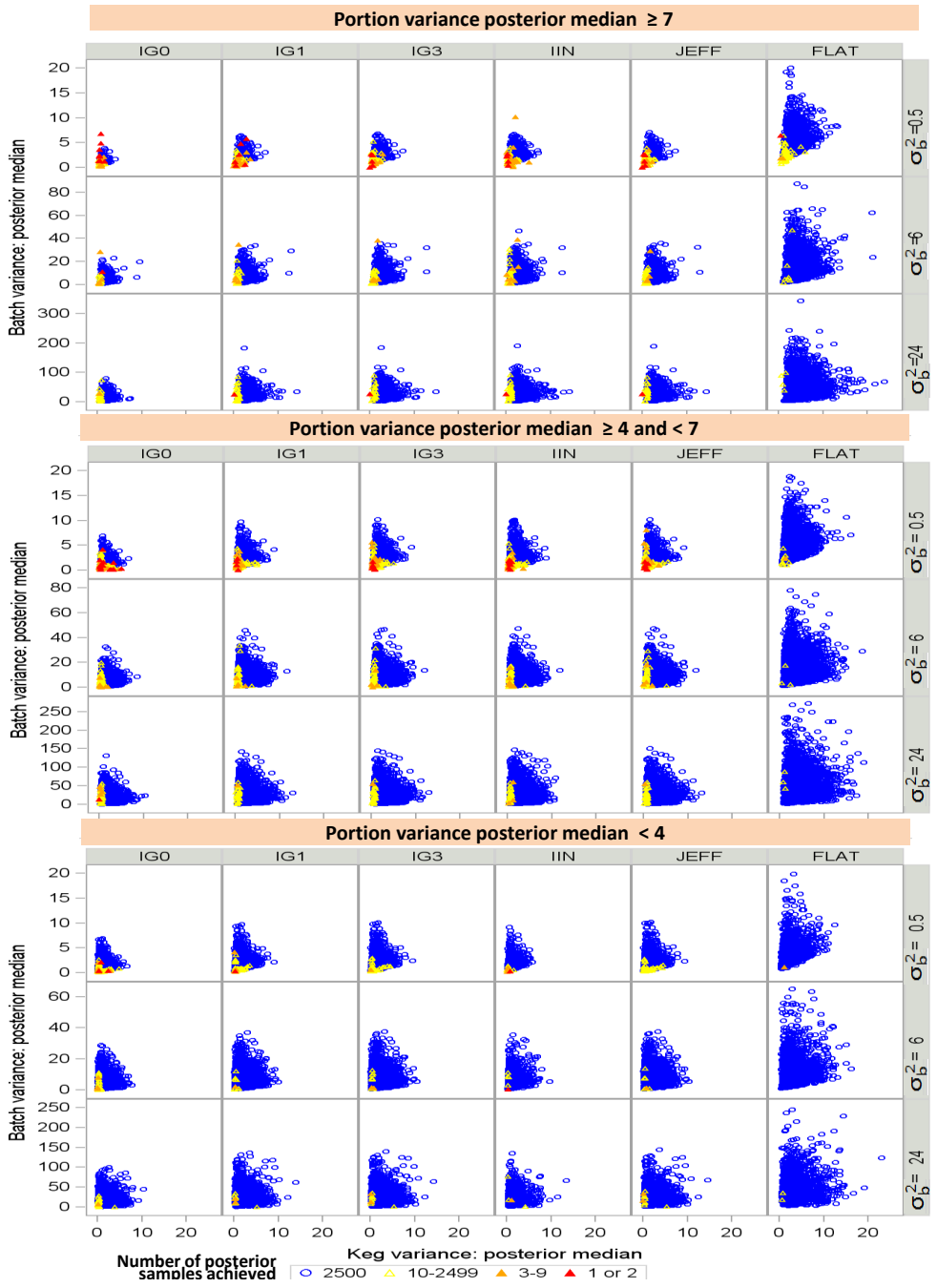


Some datasets (denoted 2128, 2292, 3155, 4782, 7737 and identified in Figure 6.21 by being enclosed with an open circle or triangle) were selected for further investigation as detailed in Section 6.4.3.4. These all stopped with fewer than 10 samples from the posterior, had a true batch variance component of 0.5 and a posterior median estimate for batch variance component of greater than 3. From the SAS software PROC MIXED analysis datasets 2128 and 2292 (enclosed by an open triangle) had a keg variance

component estimate of zero, dataset 3155 had a batch variance component estimate of zero and datasets 4782 and 7737 have both batch and keg variance estimates of zero.

Figure 6.22 splits the plots shown in Figure 6.20 by the magnitude of the portion variance. It is seen that datasets with higher median estimates of portion variance were more likely to stop.

FIGURE 6.22: **Scatterplot of posterior median estimates of σ^2_{batch} against σ^2_{keg} coloured by the number of posterior samples, classified by magnitude of $\sigma^2_{portion}$**



6.4.3.4 Investigating Modifications to Analysis

In this section various modifications to the analysis are used to see whether they affected whether sampling stopped. Firstly the seed used to produce the random samples is changed and secondly the sampling algorithm. Neither of these is found to enable full samples from the posterior to be obtained. Thirdly the relationship between the parameters of the IG proposal distribution and the sampling stopping was investigated. The parameters for the default proposal distribution were examined, particularly for those where sampling had stopped. The effect of changing the parameters for the IG proposal distribution for a set of example datasets from which sampling had stopped was examined. It is found that modifying the IG parameters did enable some datasets to provide a complete posterior samples. However, the required distribution is a careful balance of increasing the boundary acceptance rate whilst not reducing the sampling acceptance to a point where that results in posterior sampling stopping. Thus an approach is developed which attempts to find IG distributions for the strata which provide 10 or 20% of the variance component samples above zero. Further investigations of the modified proposal distributions and the posterior for the various priors and bounded and unbounded analyses are performed for one dataset to further understand why acceptance rates are low and sampling may stop.

Selection of Datasets to Investigate Some datasets were selected for further investigation to see whether a full posterior sample could be obtained by modifying the analysis options. Those datasets chosen were from the design with 6 batches, 2 portions/keg and 2 kegs/batch, a batch variance component of 0.5 and the IG1 prior was used. The following were chosen:

1. Five datasets (simulated datasets 10, 4090, 4409, 5378 and 7042) were chosen to be typical of those stopping; these had small posterior median estimates for both batch and keg variance components ≤ 0.6.
2. Five datasets (simulated datasets 2128, 2292, 3155, 4782 and 7737 as described in Section 6.4.3.3) were chosen which were not typical of those stopping; these all had a posterior median estimate for batch variance component of ≥ 3 and were highlighted in Figure 6.21.
3. Five datasets (simulated datasets 170, 869, 1379, 6517 and 6984) were chosen which were less typical of those stopping; these had small (≤ 3) posterior median estimates for the portion variance. None of these datasets had stopped as early as those in groups 1 and 2.

A summary of aspects of the analysis for these datasets is given in Table 6.17.

Changing the Seed Firstly the seed used to produce the random samples from the proposal distribution was changed. This typically resulted in a slightly different number of samples from the posterior before the analysis stops (see Table 6.18) with the dataset itself having more effect on the number of samples obtained than the seed. Given there were only small differences in the number of samples, changing the seed is unlikely to enable the full 2000 samples to be obtained. It also indicates that the stopping of the sampling is not due to an unfortunate bad stream of random numbers but is related to the location of the chain.

Changing the Sampling algorithm Secondly the sampling algorithm was changed (see Sections 6.4.1.1 and 6.4.1.2 for details of those available in SAS software). The number of samples obtained are shown in Table 6.19. For algorithms IC, IS and RS this has resulted in only a slightly different number of samples from the posterior before the analysis stops and thus they are unlikely to provide a full sample. For the random walk algorithm (RWC) the output dataset contained 2000 values. However these had sample numbers as 0 and were not true samples see Section 6.4.1.3 for further details. Thus no true samples were obtained and a number 0 is recorded in Table 6.19.

Investigating the default proposal distribution parameters The default proposal distributions used by SAS software were then investigated. For this design and IC, IS and RS sampling algorithms, the procedure places inverse gamma distributions on portion variance ($\sigma^2_{portion}$), keg stratum variance ($2\sigma^2_{keg} + \sigma^2_{portion}$), and batch stratum variance ($4\sigma^2_{batch} + 2\sigma^2_{keg} + \sigma^2_{portion}$). Figure 6.23 shows the default proposal distribution IG(3.1,β_k) placed on keg stratum variance and IG(2.6,β_b) on the batch stratum variance for the datasets listed in Table 6.17. The key in the figure indicates the dataset group, the dataset number and the scale parameters associated with the two strata (Note an additional proposal distribution with β_b=50 and β_k=50 which will be discussed later is also plotted). Figure 6.24 shows the default proposal distribution IG(6.1,β_p) placed on portion variance. Figure 6.25 shows the marginal distributions placed on the batch and keg variance components given the distributions placed on the batch and keg strata and portion variance. They are obtained by simulating from the proposal distributions for the batch, keg and portion strata. The key in the figure indicates the dataset number and the associated IG proposal distributions on the batch and keg strata. The proposal IG distribution for the portion variance had a shape value of 6.1 and scale parameters within 16-107 for the datasets being examined (see Table 6.17 for the actual scale value associated with each dataset).

TABLE 6.17: **Summary of datasets for further investigation**

	Acceptance Rates			**Batch**				**Keg**				**Portion**			
				Posterior	REML	IG(α,β)		Posterior	REML	IG(α,β)		Posterior	REML	IG(α,β)	
Dataset	Boundary	Sampling	n	Median	Est.	α	β	Median	Est.	α	β	Median	Est.	α	β
10	0.00	1.00	3	0.51	0.00	2.6	5.96	0.39	0.00	3.1	1.59	5.05	4.36	6.1	42.85
4090	0.00	1.00	9	0.61	0.00	2.6	4.94	0.50	0.00	3.1	3.17	3.35	4.23	6.1	40.79
4409	0.00	1.00	7	0.47	0.00	2.6	2.30	0.25	0.00	3.1	5.88	4.81	4.30	6.1	41.60
5378	0.00	1.00	3	0.26	0.00	2.6	4.31	0.04	0.00	3.1	3.96	8.26	6.59	6.1	67.77
7042	0.00	1.00	3	0.22	0.00	2.6	1.92	0.69	0.00	3.1	15.78	5.36	7.14	6.1	64.73
2128	0.00	1.00	7	3.76	1.99	2.6	33.90	1.01	0.00	3.1	1.53	6.29	5.55	6.1	48.59
2292	0.00	1.00	4	4.15	2.44	2.6	33.20	0.16	0.00	3.1	0.71	3.59	3.49	6.1	30.92
3155	0.00	1.00	5	3.01	0.00	2.6	1.11	2.98	2.21	3.1	56.71	7.77	6.06	6.1	36.47
4782	0.00	1.00	1	4.76	0.00	2.6	4.95	1.71	0.00	3.1	3.14	8.77	5.43	6.1	54.68
7737	0.00	1.00	1	5.77	0.00	2.6	7.31	2.97	0.00	3.1	5.92	17.07	10.38	6.1	106.50
170	0.00	1.00	31	0.60	0.31	2.6	10.30	0.44	0.00	3.1	1.30	2.94	2.84	6.1	24.45
869	0.00	1.00	23	0.33	0.00	2.6	1.14	0.25	0.00	3.1	6.42	2.71	2.48	6.1	21.29
1379	0.00	1.00	40	0.27	0.00	2.6	3.27	0.37	0.00	3.1	3.22	2.88	2.85	6.1	26.58
6517	0.00	1.00	23	0.46	0.00	2.6	1.28	1.80	1.58	3.1	31.31	2.56	2.73	6.1	16.46
6984	0.00	1.00	23	0.72	0.00	2.6	8.05	0.32	0.00	3.1	1.68	2.76	3.36	6.1	29.21

TABLE 6.18: **Effect of changing the seed on number of posterior samples obtained**

	Datasets														
Seed	**10**	**170**	**869**	**1379**	**2128**	**2292**	**3155**	**4090**	**4409**	**4782**	**5378**	**6517**	**6984**	**7042**	**7737**
198631	3	31	23	40	7	4	5	9	7	1	3	23	23	3	1
248651	1	23	28	35	7	4	2	14	19	3	3	16	21	4	4
289341	3	16	24	34	6	2	3	14	9	1	1	21	22	4	1
397257	1	29	24	41	11	6	2	12	17	1	0	19	24	9	0
482281	0	25	21	29	9	3	2	13	9	4	3	15	21	4	2
571423	4	26	28	32	8	5	2	11	18	4	2	12	20	1	2
617823	2	27	27	30	11	5	0	12	8	3	1	17	30	5	6
672513	2	25	27	38	13	4	2	11	10	4	2	23	28	7	1
783467	3	21	22	32	12	4	6	11	8	3	2	18	16	3	2
934249	3	38	33	26	5	3	3	15	9	4	2	21	18	2	2

TABLE 6.19: **Effect of changing the sampling algorithm on number of posterior samples obtained**

	Datasets														
Algorithm	**10**	**170**	**869**	**1379**	**2128**	**2292**	**3155**	**4090**	**4409**	**4782**	**5378**	**6517**	**6984**	**7042**	**7737**
IC	3	31	23	40	7	4	5	9	7	1	3	23	23	3	1
IS	3	28	26	38	6	4	6	10	6	1	3	24	19	2	1
RS	2	28	26	35	8	4	5	12	6	2	1	23	20	2	1
RWC	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*

*** The output dataset contained 2500 values but these were not true samples**

Figure 6.25 shows that the datasets 10, 4090, 4409, 5378, 2128, 2292, 4782, 7737, 170, 1379 and 6984 have their marginal proposal distribution for the keg variance component almost entirely below zero, explaining why the boundary sampling acceptance rate is close to zero. The scale parameter for the keg stratum for these datasets is less than 6. Figure 6.25 also shows that datasets 7042, 3155, 869 and 6517 have their marginal proposal distribution for the batch variance component almost entirely below zero, again explaining why the boundary sampling acceptance rate is close to zero. The scale parameter for the batch stratum for these datasets is less than 2. Even though datasets 10, 4090, 4409, 5378, 4782, 7737 1379 and 6984 have REML estimates of batch variance component of zero, the proposal distribution for the batch variance has reasonable density for values greater than zero suggesting that for those datasets it is the proposal density for the keg variance component which is causing the boundary sampling rate to be low.

Since the distributions placed on the batch and keg variance components are a consequence of the IG distributions placed on the strata the IG parameters were explored in more detail. It was seen in Table 6.17 that the shape parameters (α) were the same for all 15 datasets being investigated. For each prior the shape parameter α was evaluated over the 10,000 datasets for each of σ_b^2 values of 0.5, 6 and 24 (30,000 instances per prior) for scenario k2s2_0.5. Table 6.20 shows the values of α found for the parameters related to the batch, keg or portion strata categorised by prior. The count for each value is also shown e.g. for the batch and portion stratum and a particular prior, all 30,000 occurrences had the same value, whilst for the keg stratum 143 occurrences had a value of 1 for each prior and 29857 occurrences had the same value within a prior but this differed across the priors.

TABLE 6.20: **Values of inverse gamma shape parameters by prior**

	Batch	Keg		Portion
	IG(α)	IG(α)	IG(α)	IG(α)
Count	30000	143	29857	30000
Prior				
IG0	3.50	1.00	4.00	7.00
IG1	2.60	1.00	3.10	6.10
IG2	2.51	1.00	3.01	6.01
IG3	2.50	1.00	3.00	6.00
IIN	2.50	1.00	3.00	7.00
JEFF	2.50	1.00	3.00	6.00
FLAT	1.50	1.00	2.00	5.00

Table 6.20 has been ordered with decreasing shape parameter for the batch and keg strata. It is seen that the larger shape parameters for these strata are associated with the more informative priors IG1 and IG0. The FLAT prior has the smallest shape parameters. There were 143 occurrences of the IG shape parameter for the keg stratum

FIGURE 6.23: **Inverse gamma proposal distributions IG$(2.6, \beta_b)$ on batch stratum variance and IG$(3.1, \beta_k)$ on keg stratum variance, plotted in three groups**

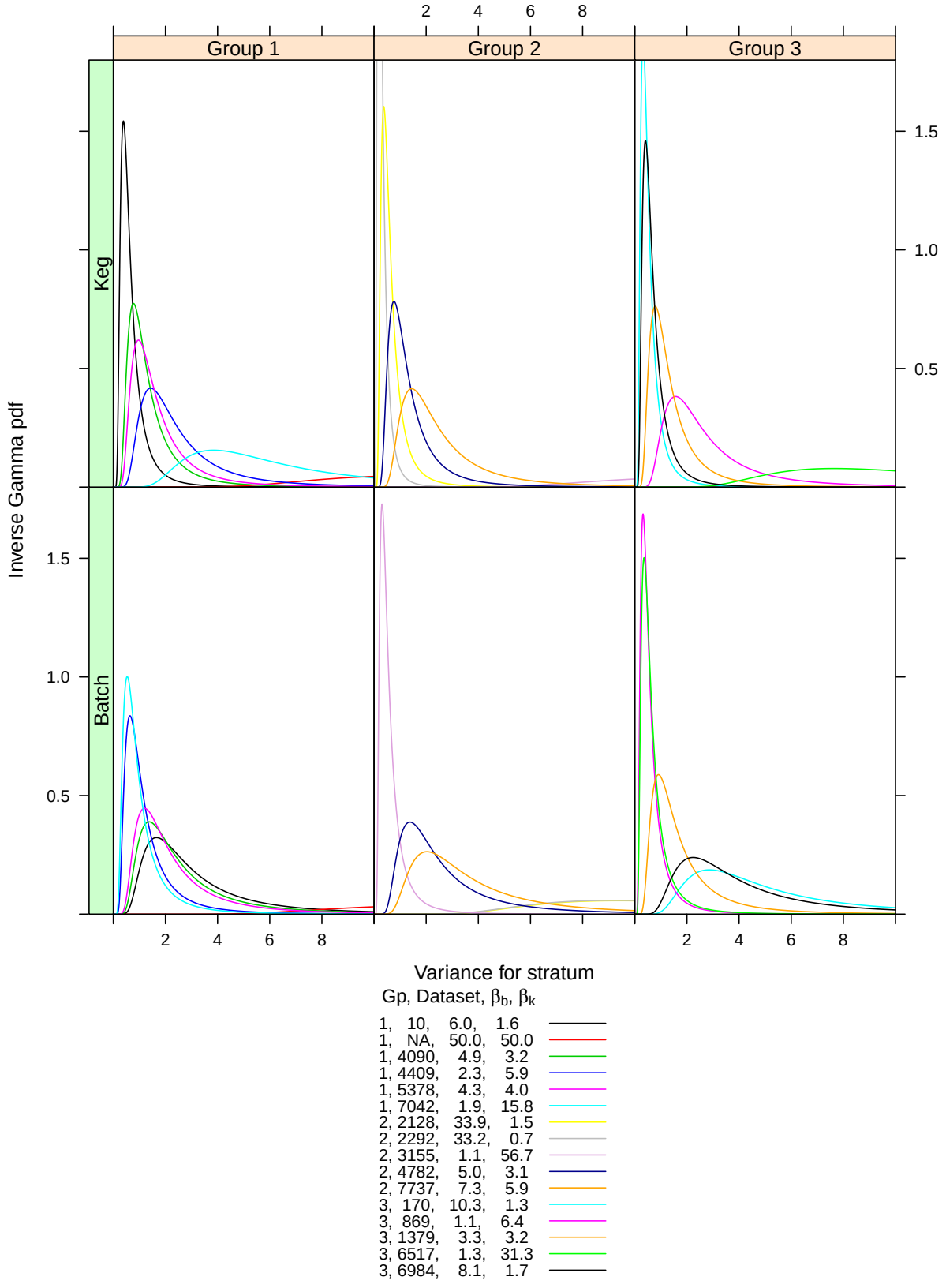
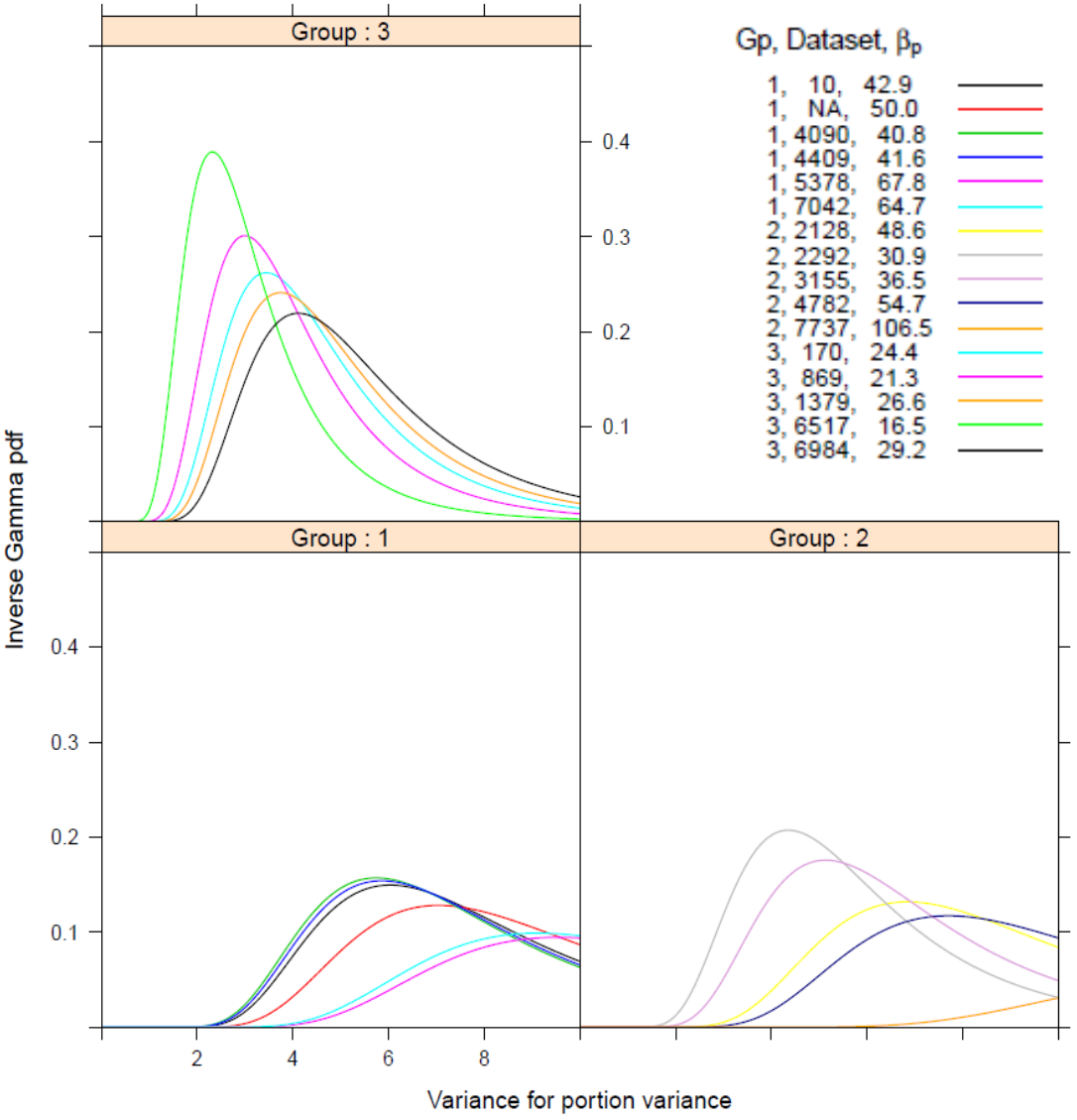


FIGURE 6.24: **Inverse gamma proposal distributions IG(6.1, β_p) on $\sigma^2_{portion}$**



having a value of 1 for each prior. Of these 111 occurred for a σ^2_b value of 0.5, 29 for σ^2_b=6 and 3 for σ^2_b=24. The same 143 datasets had an IG shape parameter for the keg stratum having a value of 1 for all priors and this corresponded with the IG scale parameter also having a value of 1. In contrast the datasets with values other than 1 for the IG shape parameter for the keg stratum had a range of values (approximate summary statistics are: median=19, interquartile range=12-28 and range=0.5-115) for the scale parameter. Figure 6.26 shows the shape of the IG(1,1) distribution compared to IG distributions typical for other datasets. It is seen that it corresponds to higher density for smaller keg stratum variances. All datasets achieved the requested number of samples from the posterior and thus it did not indicate a cause of the stopped datasets.

Figure 6.27 plots the IG scale parameter (β_b) for the batch stratum against the IG scale

FIGURE 6.25: **Marginal proposal distributions on σ^2_{batch} and σ^2_{keg} given IG(2.6,β_b) on batch stratum variance, IG(3.1,β_k) on keg stratum variance**

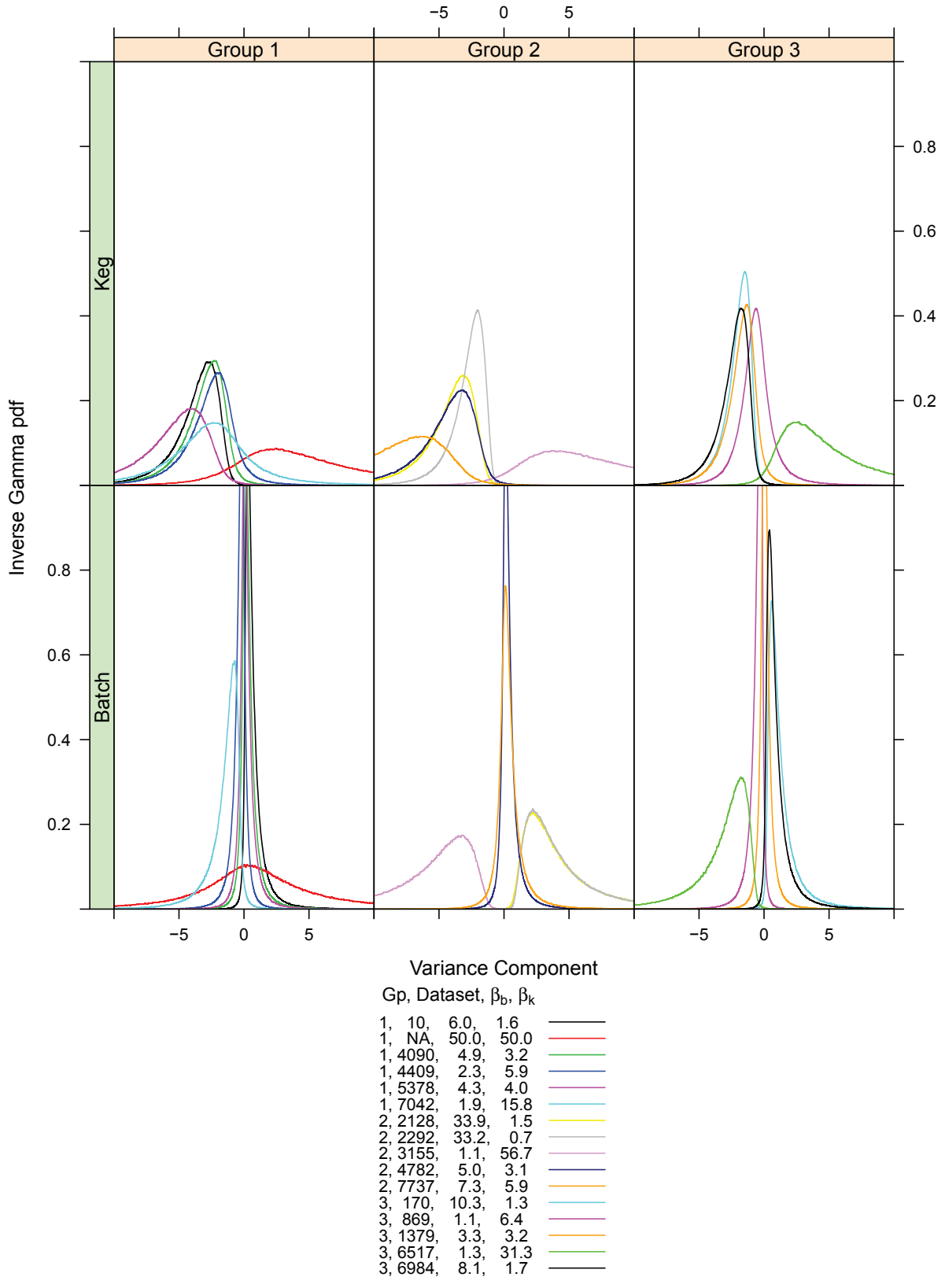
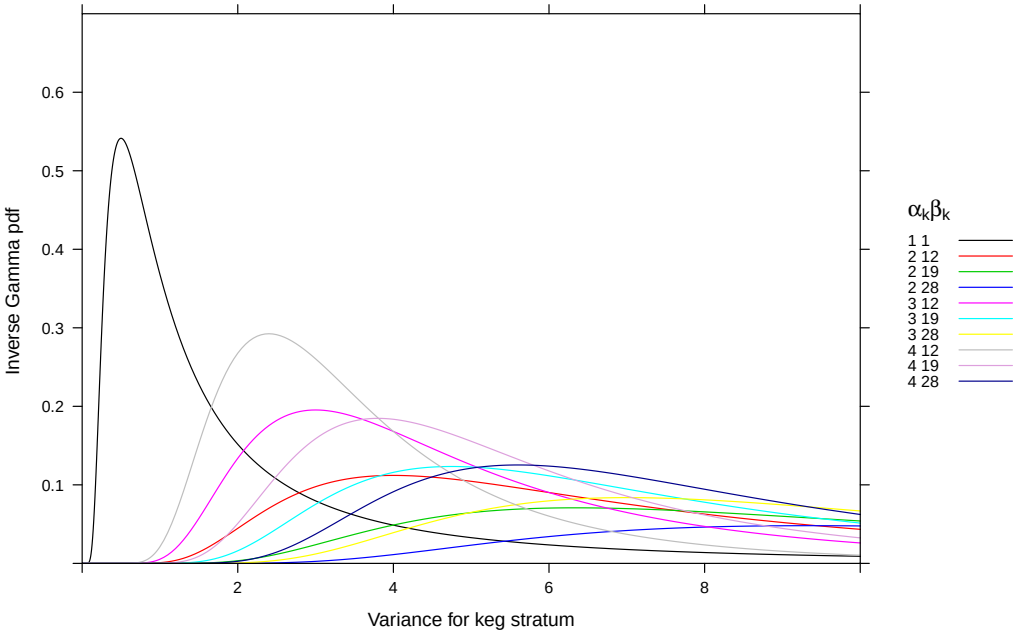


FIGURE 6.26: **Inverse gamma proposal distributions IG(α_k, β_k) on keg stratum variance**

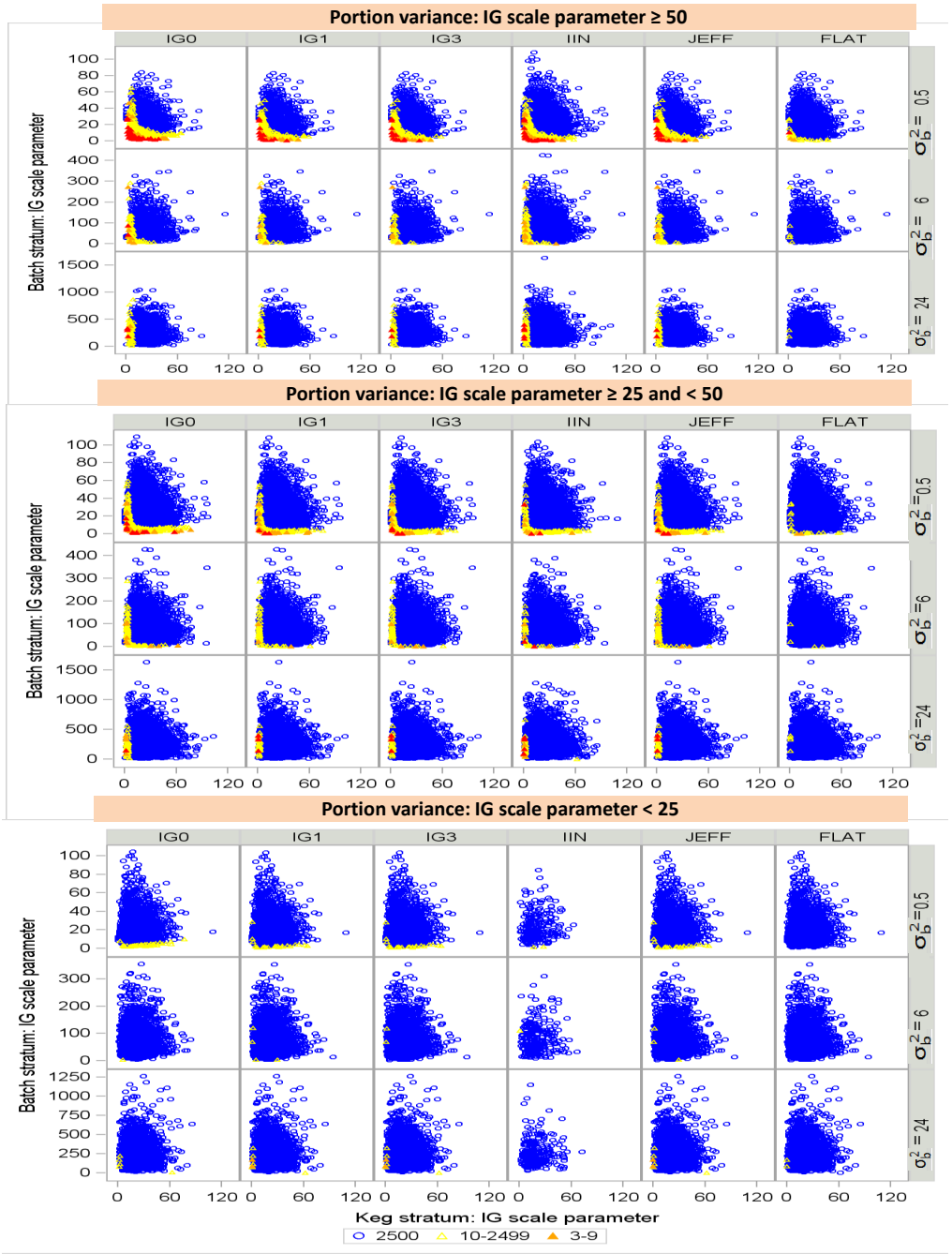


parameter (β_k) for the keg stratum for various priors and batch variance components of 0.5, 6, and 24. The datasets are identified using symbols to show whether a full sample of 2500 was achieved or a smaller number of samples. The plot is split by the magnitude of the IG scale parameter (β_p) for the portion variance.

It is seen that datasets with larger scale parameters for the portion variance IG proposal distribution tend to have more stopped datasets. Those datasets which stopped tend to have small scale parameters for the batch and/or keg strata. If the batch and keg stratum scale parameters are both above 50 there don't appear to be any stopped datasets.

Changing the proposal distribution parameters Having examined the parameters for the default IG proposal distributions, the effect of changing the proposal distribution parameters was investigated. IG(3.1, 50) was chosen for the keg stratum and IG(2.6, 50) was chosen for the batch stratum given that datasets with these parameters had not stopped. IG(6.1, 50) was chosen for the portion variance, though it was noted from Figure 6.27 a smaller value may be more likely to produce a full sample. The corresponding proposal distributions for the keg and batch variance components were plotted in Figure 6.25 (group 1, dataset=N/A, red curve). It is seen that the proposal distributions for both the batch and keg variance components have reasonable area above zero.

FIGURE 6.27: **Scatterplot of β_{batch} for IG of batch stratum variance against β_{keg} for IG on keg stratum variance coloured by the number of posterior samples, classified by magnitude of $\beta_{portion}$ for IG on $\sigma^2_{portion}$**



Initially the 15 datasets listed in Table 6.17 which had a true batch variance component of 0.5 were run with the changed proposal distribution. Having seen that changing the proposal distribution affected the number of samples obtained from the posterior, a further 8 datasets were evaluated. These consisted of four datasets (1225, 6309, 8402, 9281) with true batch variance component of 6 and four datasets (2066, 4155, 6039, 6431) with true batch variance component of 24. In all a total of 23 datasets were run with the changed proposal distribution. With the default proposal distribution they all stopped with less than 10 samples from the posterior.

The number of samples taken from the posterior when 2500 were requested are shown in Table 6.21 for the 23 datasets. It is seen that whilst for some seeds less than 2500 samples were obtained for some datasets, the majority succeeded in obtaining 2500 samples (denoted F in the table) when the changed proposal distribution was used. A number of seeds (A=198631, B=248651, C=289341, D=397257, E=482281, F=571423, G=617823, H=672513, I=783467, J=934249) were used in order to assess whether any increase seen in the number of samples obtained was robust to the choice of seed.

TABLE 6.21: **Number of posterior samples with changed proposal distribution with scale parameters=50, for various seeds for the 23 datasets**

		Seed									
σ_b^2	**Dataset**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
	10	F	F	F	F	F	F	F	F	F	F
	170	F	F	F	F	F	F	F	F	F	F
	869	F	F	F	F	F	F	F	F	F	F
	1379	F	F	F	F	F	F	F	F	F	F
	2128	F	F	F	F	F	F	F	F	F	F
	2292	F	F	F	F	F	F	F	F	F	F
	3155	F	F	F	F	F	F	F	F	F	F
0.5	**4090**	F	F	F	F	F	F	F	F	F	F
	4409	F	F	F	F	F	F	F	F	F	F
	4782	F	F	F	F	F	501	F	F	F	F
	5378	F	F	F	F	F	401	F	F	F	F
	6517	F	F	501	F	F	F	F	F	F	901
	6984	F	F	F	F	F	F	F	F	F	F
	7042	F	F	F	F	F	F	F	F	F	F
	7737	F	F	F	F	F	F	F	F	F	F
	1225	F	F	F	F	F	F	F	F	F	F
6	**6309**	F	F	F	F	F	F	F	F	F	F
	8402	F	F	F	F	F	F	F	F	F	F
	9281	F	F	F	F	F	F	F	F	F	F
	2066	301	F	F	F	F	F	201	F	F	F
24	**4155**	F	F	F	F	F	F	F	F	F	F
	6039	F	F	F	F	F	F	F	F	F	F
	6431	301	F	F	F	F	F	201	F	F	F

F = Full sample = 2500

Two other alternatives for the set of IG(α,β) proposal distributions were also investigated. Firstly IG(2.6, 1000) and IG(3.1, 50) were chosen for the batch and keg stratum respectively to see whether increasing the scale parameter for the batch strata above 50 improved the sampling further e.g. by obtaining a full sample over all seeds investigated, though it is noted that for σ^2_{batch}=0.5 or 6 none of the datasets had batch stratum scale parameters this high. Secondly IG(2.6, 20) and IG(3.1, 20) were chosen for the batch and keg stratum respectively to see the effect of decreasing these parameters from 50. The proposal distribution for $\sigma^2_{portion}$ was IG(6.1,50) in both cases. The results shown in Tables 6.22 and 6.23. Both alternative proposal distributions also succeeded in the majority of cases in providing the full posterior sample requested. The set with IG(2.6, 20) on the batch stratum was similar to that with IG(2.6, 50) on the batch stratum stopping on 10 occasions compared to 8. However, the set with IG(2.6, 1000) on the batch stratum stopped on 29 occasions.

TABLE 6.22: **Number of posterior samples with changed proposal distribution with scale parameters=1000 and 50, for various seeds for the 23 datasets**

		Seed									
σ^2_b	**Dataset**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
	10	F	F	F	F	F	F	F	F	F	F
	170	F	901	1301	F	F	F	2101	F	F	F
	869	F	601	1001	F	F	F	1201	F	F	F
	1379	F	901	1801	F	F	F	F	F	F	F
	2128	F	F	F	F	F	F	F	F	F	F
	2292	F	1001	1601	F	F	F	F	F	F	F
	3155	F	1101	F	F	F	F	F	F	F	F
0.5	**4090**	F	F	F	F	F	F	F	F	F	F
	4409	F	F	F	F	F	F	F	F	F	F
	4782	F	F	F	F	F	F	F	F	F	F
	5378	F	F	F	F	F	F	F	F	F	F
	6517	F	1101	2401	2301	F	1401	F	F	F	F
	6984	F	1001	F	F	F	F	F	F	F	F
	7042	F	F	F	F	F	F	F	F	F	F
	7737	F	F	1501	F	501	F	F	1101	F	F
	1225	F	F	F	F	F	F	F	F	F	F
6	**6309**	F	F	F	F	F	F	F	F	F	F
	8402	F	1001	1601	F	F	F	F	F	F	F
	9281	F	F	1801	F	F	F	F	F	F	F
	2066	1801	601	F	F	F	101	2101	F	F	F
24	**4155**	F	701	1001	F	F	F	1701	F	F	F
	6039	F	F	F	F	F	F	F	F	F	F
	6431	F	701	F	F	F	F	F	F	F	F

F = Full sample = 2500

TABLE 6.23: **Number of posterior samples with changed proposal distribution with scale parameters=20, for various seeds for the 23 datasets**

		Seed									
σ_b^2	**Dataset**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
	10	F	F	F	F	F	F	F	F	F	F
	170	F	F	F	F	F	F	F	F	F	F
	869	F	2301	F	F	F	F	501	F	F	F
	1379	F	F	F	F	F	F	F	F	F	F
	2128	F	F	F	F	F	F	F	F	F	F
	2292	F	F	F	F	F	F	F	F	F	F
	3155	F	F	F	F	F	F	F	F	F	F
0.5	**4090**	F	F	F	F	F	F	F	F	F	F
	4409	F	F	F	F	F	F	F	F	F	F
	4782	F	F	F	F	F	F	F	F	F	F
	5378	F	F	F	F	F	F	F	F	F	F
	6517	F	F	F	F	F	F	F	F	F	F
	6984	F	F	F	F	F	F	F	F	F	F
	7042	F	F	F	F	F	F	F	F	F	F
	7737	F	F	F	F	F	F	F	F	F	F
	1225	F	F	F	F	F	F	F	F	F	1601
6	**6309**	F	F	F	F	F	F	F	F	F	F
	8402	F	F	F	F	F	F	F	F	F	F
	9281	F	F	F	F	F	F	F	F	F	F
	2066	F	F	1801	901	F	1001	F	F	F	F
24	**4155**	F	F	1601	F	F	F	F	F	F	F
	6039	F	F	F	F	F	F	F	F	F	F
	6431	F	F	1501	F	F	F	F	F	2001	201

F = Full sample = 2500

In Table 6.24 the boundary and sampling acceptance rates are shown for all three alternative sets of IG(α,β) proposal distributions: i) β=20 for the batch and keg strata; ii) β=50 for the batch and keg strata and iii) β=1000 for the batch stratum and β=50 for the keg stratum. It is seen that the boundary acceptance rates are reasonably consistent across different datasets, even when the datasets have a different true batch variance. As discussed in Section 6.4.1.1, with the assumption that the posteriors differ between the datasets this might suggest that proposal samples are tested whether they meet the boundary constraints before the comparison of the densities. The boundary acceptance rate increased from 0.14 for β_{batch}=20 to 0.85 for β_{batch}=1000. The sampling acceptance rates are also plotted in Figure 6.28. The sampling acceptance rate varies according to the dataset and with the true batch variance, with higher batch variance having, in general, lower sampling acceptance rates. When the true batch variance component is 24 the sampling rate is very low with some of the datasets having a sampling rate as low as 0.01. Thus whilst the proposal distributions may be fairly easily changed to increase

the boundary acceptance rate, this may cause a problem with the sampling acceptance rate for specific datasets.

TABLE 6.24: **Boundary and sampling rates for various datasets and changed proposal distributions**

				Acceptance Rates		
β_{batch}	β_{keg}	σ_b^2	**Dataset**	**Boundary**	**Sampling**	σ^2**=0**
20	20	0.5	10	0.14	0.29	Keg, Batch
20	20	0.5	170	0.14	0.13	Keg
20	20	0.5	869	0.14	0.08	Keg, Batch
20	20	0.5	1379	0.14	0.12	Keg, Batch
20	20	0.5	2128	0.14	0.57	Keg
20	20	0.5	2292	0.14	0.21	Keg
20	20	0.5	3155	0.14	0.45	Batch
20	20	0.5	4090	0.14	0.28	Keg, Batch
20	20	0.5	4409	0.14	0.30	Keg, Batch
20	20	0.5	4782	0.14	0.46	Keg, Batch
20	20	0.5	5378	0.14	0.56	Keg, Batch
20	20	0.5	6517	0.14	0.13	Batch
20	20	0.5	6984	0.14	0.16	Keg, Batch
20	20	0.5	7042	0.14	0.62	Keg, Batch
20	20	0.5	7737	0.14	0.28	Keg, Batch
20	20	6	1225	0.14	0.03	Keg
20	20	6	6309	0.14	0.24	Keg
20	20	6	8402	0.14	0.10	Keg
20	20	6	9281	0.14	0.38	Keg
20	20	24	2066	0.14	0.03	Keg
20	20	24	4155	0.14	0.06	Keg
20	20	24	6039	0.14	0.07	Keg
20	20	24	6431	0.14	0.01	Keg

Table 6.24 – *Continued from previous page*

				Acceptance Rates		
β_{batch}	β_{keg}	σ_b^2	**Dataset**	**Boundary**	**Sampling**	σ^2**=0**
50	50	0.5	10	0.47	0.07	Keg, Batch
50	50	0.5	170	0.48	0.04	Keg
50	50	0.5	869	0.48	0.03	Keg, Batch
50	50	0.5	1379	0.48	0.04	Keg, Batch
50	50	0.5	2128	0.48	0.13	Keg
50	50	0.5	2292	0.48	0.10	Keg
50	50	0.5	3155	0.47	0.36	Batch
50	50	0.5	4090	0.48	0.07	Keg, Batch
50	50	0.5	4409	0.48	0.07	Keg, Batch
50	50	0.5	4782	0.48	0.14	Keg, Batch
50	50	0.5	5378	0.48	0.16	Keg, Batch
50	50	0.5	6517	0.48	0.06	Batch
50	50	0.5	6984	0.48	0.05	Keg, Batch
50	50	0.5	7042	0.47	0.21	Keg, Batch
50	50	0.5	7737	0.47	0.26	Keg, Batch
50	50	6	1225	0.47	0.07	Keg
50	50	6	6309	0.47	0.35	Keg
50	50	6	8402	0.47	0.08	Keg
50	50	6	9281	0.46	0.36	Keg
50	50	24	2066	0.47	0.01	Keg
50	50	24	4155	0.47	0.07	Keg
50	50	24	6039	0.47	0.10	Keg
50	50	24	6431	0.47	0.01	Keg

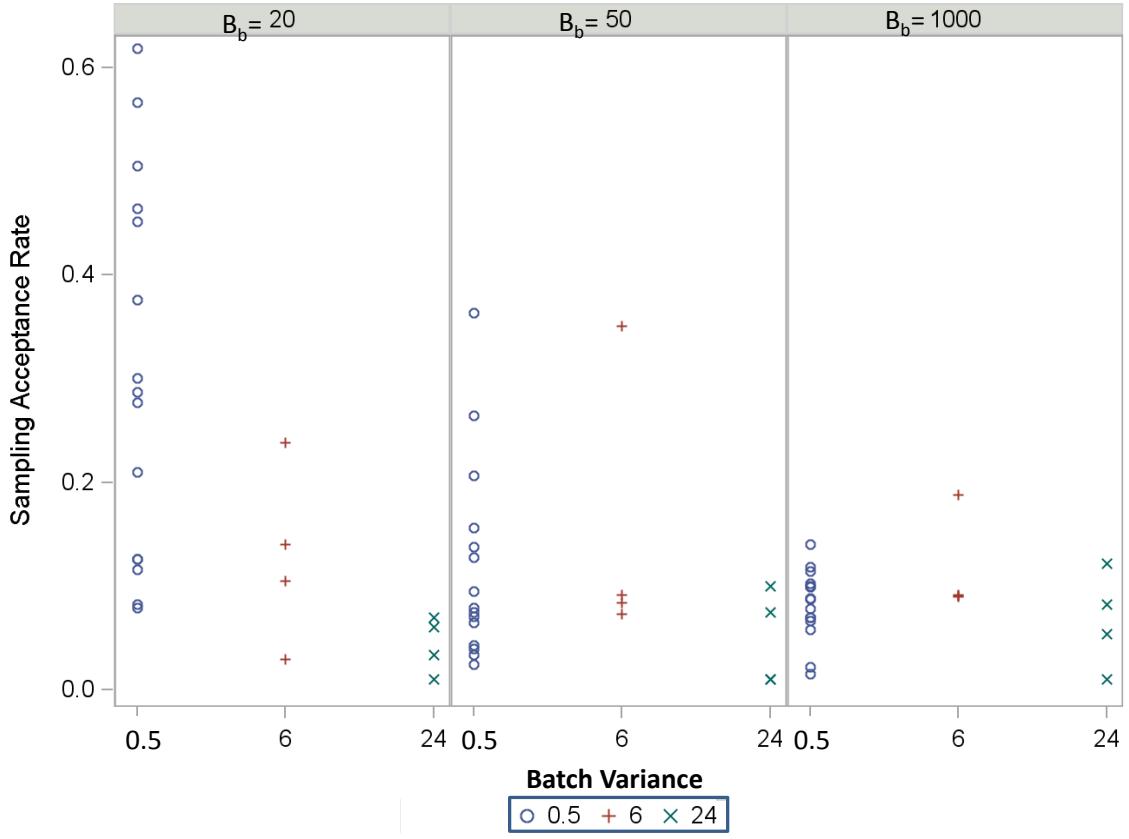
Table 6.24 – *Continued from previous page*

				Acceptance Rates		
β_{batch}	β_{keg}	σ_b^2	**Dataset**	**Boundary**	**Sampling**	σ^2**=0**
1000	50	0.5	10	0.85	0.10	Keg, Batch
1000	50	0.5	170	0.85	0.06	Keg
1000	50	0.5	869	0.85	0.08	Keg, Batch
1000	50	0.5	1379	0.85	0.07	Keg, Batch
1000	50	0.5	2128	0.85	0.12	Keg
1000	50	0.5	2292	0.85	0.09	Keg
1000	50	0.5	3155	0.85	0.07	Batch
1000	50	0.5	4090	0.85	0.11	Keg, Batch
1000	50	0.5	4409	0.85	0.11	Keg, Batch
1000	50	0.5	4782	0.84	0.10	Keg, Batch
1000	50	0.5	5378	0.85	0.09	Keg, Batch
1000	50	0.5	6517	0.85	0.02	Batch
1000	50	0.5	6984	0.85	0.08	Keg, Batch
1000	50	0.5	7042	0.85	0.09	Keg, Batch
1000	50	0.5	7737	0.85	0.07	Keg, Batch
1000	50	6	1225	0.85	0.19	Keg
1000	50	6	6309	0.84	0.09	Keg
1000	50	6	8402	0.85	0.09	Keg
1000	50	6	9281	0.84	0.10	Keg
1000	50	24	2066	0.86	0.01	Keg
1000	50	24	4155	0.85	0.08	Keg
1000	50	24	6039	0.85	0.12	Keg
1000	50	24	6431	0.85	0.05	Keg

Modifying proposal distributions on strata to obtain adequate density for positive variance components Given that for the proposal distributions evaluated increasing the acceptance rate with regard to the boundary decreases the sampling acceptance rate, an approach was sought to change the inverse gamma parameters of the proposal distribution so that the acceptance rate with respect to the boundaries was increased sufficiently to obtain a full posterior sample whilst also maintaining a sufficient sampling acceptance rate. Thus an approach for the stopped datasets of changing the IG parameters to provide just some of the distribution with batch and keg variance above zero was sought.

PROC MIXED places inverse gamma proposal distributions on the portion variance ($\sigma^2_{portion}$), keg stratum variance ($\phi_k = \mathrm{P}\sigma^2_{keg} + \sigma^2_{portion}$), and batch stratum variance ($\phi_b = \mathrm{KP}\sigma^2_{batch} + \mathrm{P}\sigma^2_{keg} + \sigma^2_{portion}$) where P=number of portions/keg and K=number of

FIGURE 6.28: **Sampling rates for various datasets and changed proposal distributions, classified by magnitude of β_{batch} for IG on batch stratum variance and by σ^2_{batch}**



kegs/batch. Thus the effective distributions on σ^2_{batch} and σ^2_{keg} are required to evaluate when the boundary conditions will not be met. Initially simulation was used to find these distributions but the computing time required to find suitable changes to parameters was too long to be used routinely. Witkovsky (2001) provides results on computing the distribution of a linear combination of inverted gamma variables and these results are used to provide an alternative method. Witkovsky gives the following results: Let $Y \sim IG(\alpha, \beta))$ be an inverted gamma random variable with a probability density function $f_Y(y)$ defined for $y \geq 0$ by:

$$f_Y(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{1}{y}\right)^{\alpha+1} \exp\left\{-\frac{1}{\beta y}\right\}. \tag{6.3}$$

Then the characteristic function of Y is

$$\varphi_Y(t) = E\left(e^{itY}\right) = \frac{2\left(-it\beta\right)^{\frac{1}{2}\alpha} K_\alpha\left\{\frac{2}{\beta}\left(-it\beta\right)^{\frac{1}{2}}\right\}}{\beta^\alpha \Gamma(\alpha)}, \tag{6.4}$$

where $K_\alpha(z)$ denotes the modified Bessel function of second kind. For a sample of

independent variables $Y_{(\alpha_1,\beta_1)},...,Y_{(\alpha_n,\beta_n)}$ where $Y_{(\alpha_k,\beta_k)} \sim IG(\alpha_k,\beta_k)$, with $\alpha_k > 0$ and $\beta_k > 0$, $k = 1,...,n$, and define $X = \Sigma_{k=1}^{n}\lambda_k Y_{(\alpha_k,\beta_k)}$ as a linear combination of n inverted gamma variables, with real coefficients λ_k. Let $\varphi_k(t) = E\left(\exp\left\{itY_{(\alpha_k,\beta_k)}\right\}\right)$ denote a characteristic function of the distribution of $Y_{(\alpha_k,\beta_k)}$. Then the characteristic function $\varphi_X(t)$ of X is given by

$$\varphi_X(t) = \varphi_{X_1}(\lambda_1 t)...\varphi_{X_n}(\lambda_n t), \tag{6.5}$$

and the distribution function $F_X(x) = Pr\{X \leq x\}$ is given by

$$F(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \left(\frac{e^{-itx}\varphi(t) - e^{itx}\varphi(-t)}{2it}\right)dt = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty Im\left(\frac{e^{-itx}\varphi(t)}{t}\right)dt\,. \tag{6.6}$$

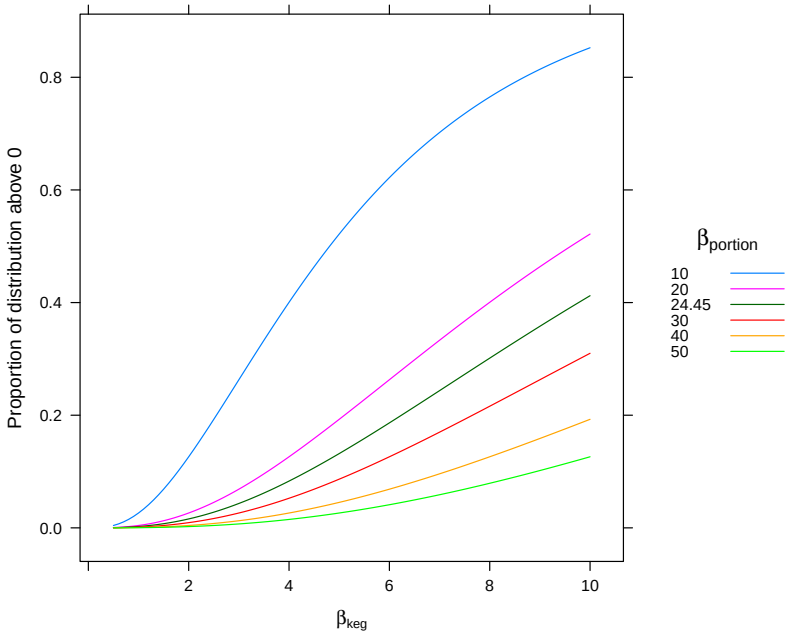
Witkovsky also warns that numerical integration should be carried out carefully if $\alpha_k \in (0,1)$ for some k. Table 6.20 shows that for the datasets and priors investigated the shape parameter was not less than 1. However 143 datasets did have a shape parameter of 1 for the keg stratum and thus in those cases there could be an issue. Rearranging the keg stratum variance (ϕ_k), $\sigma^2_{keg} = \frac{\phi_k}{P} - \frac{\sigma^2_{portion}}{P}$ where IG distributions are placed on ϕ_k and $\sigma^2_{portion}$. Thus letting $Y_1 = \phi_k$, $Y_2 = \sigma^2_{portion}$ $\lambda_1 = \frac{1}{P}$ and $\lambda_2 = \frac{-1}{P}$ and using Equations (6.4), (6.5), and (6.6) the distribution of σ^2_{keg} can be derived via numerical integration.

I implemented this methodology in a program in R for a linear combination of two inverse gamma distributions (provided in Appendix E, Section E.5). I used the Bessel package (Maechler, 2013) in R to supply the function BesselK which is needed to apply the methodology. The answers obtained agreed with those Witkovsky gives for examples $X_1 - X_4, X_6 - X_8$ in Table 1 of Witkovsky (2001) (examples X_5, X_9 and X_{10} involved more than two inverse gamma variables and thus were not implemented), though it should be noted that the parameterisation is different where β in Witkovsky (2001) is the inverse of the IG scale parameter used in this thesis. I developed an optimisation procedure in R which changed the scale parameter to seek values for which 10 and 20% of the distribution of σ^2_{keg} was above 0. A similar procedure was applied to the batch stratum variance (ϕ_b) where $\sigma^2_{batch} = \frac{\phi_b}{KP} - \frac{\phi_k}{KP}$. Large shape parameters caused problems with applying the method as the gamma function becomes very large e.g. when IG shape parameter=100, $\Gamma(100) = 10 \times 10^{155}$. So if the IG shape parameter was > 50, the IG scale and shape parameters were modified so that newshape=50 and newscale=$\frac{\text{scale}\times 50}{\text{shape}}$ thus making the gamma function calculation feasible but also adjusting the scale parameter to keep the same mean. The optimisation algorithm still gave errors occasionally which have not been explored further. One possibility is that this could be due to the shape parameter being 1 (or less), where as explained above this causes problems to the numerical integration. It will be seen that this did not pose a serious problem to obtaining a full sample from the posterior so has not been investigated further. Whilst the aim for each individual variance component was to achieve 10 or 20% above zero, given that both need to be above zero to be accepted, in practice the

boundary acceptance rate will be lower than 10 or 20%. Note subsequent to the original explorations 50% above zero was also looked at.

An example is given for dataset 170, with the design 6 batches, 2 kegs/batch and 2 portions/keg and an IG1 prior. The default proposal distribution gave IG(3.1,1.30) on the keg stratum variance ($2\sigma^2_{keg} + \sigma^2_{portion}$) and IG(6.1,24.45) on the portion variance. Figure 6.29 shows the proportion of the distribution of σ^2_{keg} above 0 for the IG(3.1,β_{keg}) on the keg stratum variance as β_{keg} is increased and for the default proposal distribution on the portion variance and for various alternative values of $\beta_{portion}$.

FIGURE 6.29: **Proportion of distribution on σ^2_{keg} above 0 given proposal distribution of IG(3.1,β_{keg}) on keg stratum variance and IG(6.1,$\beta_{portion}$) on $\sigma^2_{portion}$**



It is seen that for the range of IG scale parameters explored for the portion variance, a value of $\beta_{keg} < 2$ typically gave a very small proportion of the distribution of σ^2_{keg} above 0 and thus if the default proposal distribution had a value of $\beta_{keg} < 2$ the search algorithm was started with a value of 2.

Further exploration of the Posterior and Proposal Distributions A strategy for modifying the proposal distributions which incorporates the above methods and some additional elements is implemented in Section 6.4.4. However, some further exploration of the effect of the modifications for one dataset are now described to gain further understanding of why acceptance rates are low and sampling may stop, together with

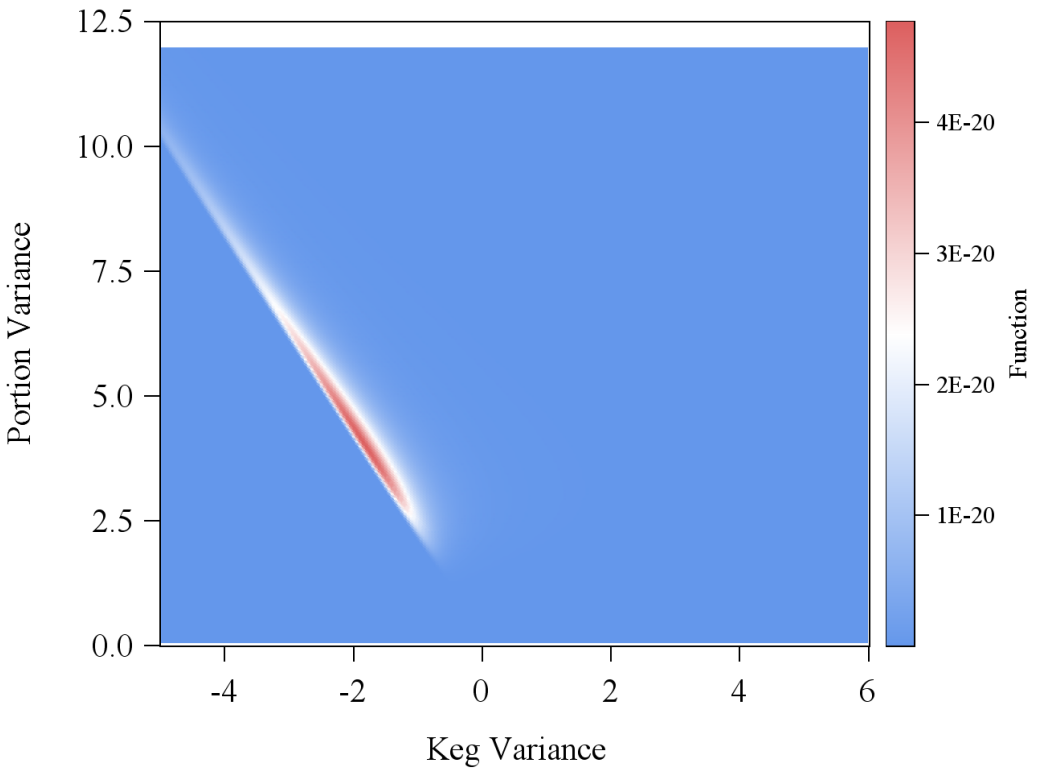
ideas for possible alternatives to changing the inverse gamma distribution. For dataset 170 and the design 6 batches, 2 kegs per batch and 2 portions per keg the likelihood and posterior distributions are examined for a subset of the priors. Firstly the distributions over keg and portion variances are examined for specified batch values. Secondly the distributions over batch and keg variances are examined for specified portion values. Finally the default proposal distribution and some modified proposal distributions are explored and compared to posterior distributions.

Firstly the distributions over keg and portion variances are examined for specified batch values. Though it was intended to run PROC MIXED using the option which bounds the variance component estimates to be ≥ 0 the analysis which allows negative estimates is initially explored. This helps to provide understanding why the boundary acceptance rate becomes low. For dataset 170 (for which sampling from the posterior stopped) the residual likelihood is plotted in Figure 6.30 over a range of σ^2_{keg} and $\sigma^2_{portion}$ values, for $\sigma^2_{batch} = 3$. Figure 6.31 shows the residual likelihood for $\sigma^2_{batch} = 1$. It is seen that:

- the area where the likelihood is highest is within the region where $\sigma^2_{keg} \leq 0$.
- The likelihood has a cliff at the line $2\sigma^2_{keg} + \sigma^2_{portion} = 0$ which represents the keg stratum variance which must be ≥ 0 for these models. To the lower left, the likelihood is effectively zero (the natural log value is reported as -8.9884656743116E307 i.e. SAS software notation for -8.9884656743116 $\times 10^{307}$).
- The highest residual likelihood value does not lie on the cliff.
- For $\sigma^2_{batch} = 3$ (Figure 6.30), σ^2_{keg} values within the range -2.5 to -15 have some density within the reddish region (representing a value at least half that of the maximum) whilst for $\sigma^2_{batch} = 1$ (Figure 6.31), the reddish region corresponds to σ^2_{keg} in the range -1 to -5 i.e. the highest region shifts towards less negative σ^2_{keg} values.

The posterior is then plotted for five priors: IG1, IG3, IIN, FLAT and JEFF (Figure 6.32), the colour axis being scaled by the maximum value for each individual plot. It is seen that the unbounded posterior has the same cliff and diagonal region as the likelihood but that the posterior constrains the reddish area to have keg variance greater than or approximately equal to -3, in comparison to the likelihood where the reddish area had keg variances as low as -5. The posteriors have different maximums ranging from 2.5 $\times 10^{-30}$ to 4 $\times 10^{-20}$. However, the patterns are fairly similar across the priors though subtle differences are seen. The FLAT prior has a broader peak (comparing the reddish pink region where the value is above approximately half the maximum) compared to the other priors. The IIN prior has the peak nearer to a portion value of 6 compared to the other priors where the peak is at lower portion values, which can be expected due the more informative prior on the portion variance for the IIN prior. For the FLAT prior it is seen that the posterior is the same as the likelihood plotted in Figure 6.30, as expected.

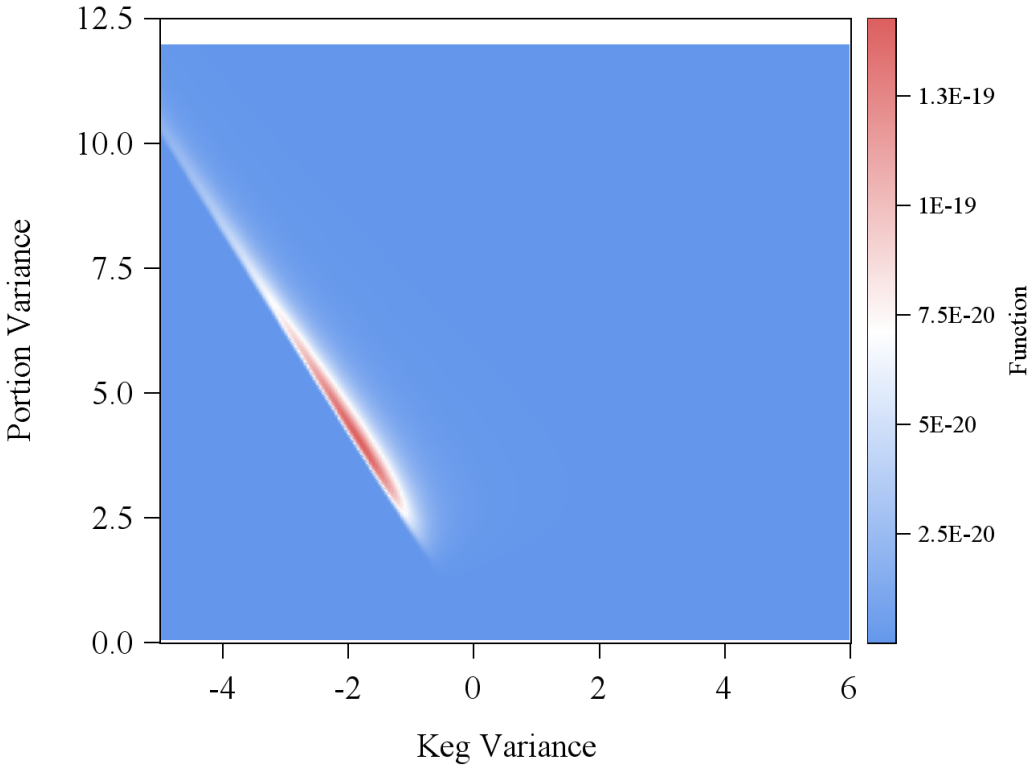
FIGURE 6.30: **Residual unbounded likelihood plotted for σ^2_{keg} and $\sigma^2_{portion}$, for $\sigma^2_{batch} = 3$**



In Figure 6.33 the posterior for the bounded analysis is plotted for priors IG1, IG3, IIN, FLAT and JEFF the colour axis being scaled by the maximum value in the bounded region for each individual plot. Note the peak shown in Figure 6.33 is not visible in Figure 6.32 as the region has very low density values in comparison and on the colour scale of Figure 6.32 it is too low to be distinguished. Again it is seen that the IIN prior has a peak at slightly higher portion values and that the FLAT prior has a broader peak (comparing the reddish pink region where the value is above approximately half the maximum) compared to the other priors. For the bounded analysis the posterior values in the region with variance components above zero is the same as that for the unbounded analysis. Thus the priors have effectively been truncated, rather than censored. The inverse gamma priors form proper priors on the strata but the truncation implies a truncated prior which is not a proper density function, rather than a censored prior with a spike at zero.

Secondly the distributions over keg and batch variances are examined for specified portion values. The residual likelihood for the unbounded analysis is shown in Figure 6.34 for the batch and keg variance components for portion variances of 1, 3 or 6. The cliff represented by the keg stratum constraint $2\sigma^2_k + \sigma^2_p \geq 0$ is seen at σ^2_k = -0.5 for σ^2_p = 1, -1.5 for σ^2_p = 3 and -3 for σ^2_p = 6 can be seen. There is also a restriction that

FIGURE 6.31: **Residual unbounded likelihood plotted for σ^2_{keg} and $\sigma^2_{portion}$, for $\sigma^2_{batch} = 1$**



$\sigma^2_b \geq -\frac{\sigma^2_k}{2} - \frac{\sigma^2_p}{4}$. However, over the range of keg variances which meet the keg stratum constraint this represents a maximum constraint that $\sigma^2_b \geq 0$ at the minimum σ^2_k which with this dataset is not very apparent but just seen in the plot where portion=1. The posterior for the unbounded analysis is then plotted for the batch and keg variance components. The posterior for the JEFF prior is plotted for a portion value of 3 in Figure 6.35. It is seen that the density is highest for negative values of keg variance (as was also seen in Figure 6.32) and for positive values of batch variance. The shape is similar to that of the likelihood for a portion value of 3 but not as broad. To compare across priors and portion variances the posterior is shown in Figure 6.36 for the batch and keg variance components for portion variances of 1, 3 or 6. The effect of portion on where the maximum posterior density occurs on the keg axis is seen and is as expected from Figure 6.32 – the higher the portion variance the lower the keg variance. The maximum with respect to batch does not seem particularly dependent on keg, portion or prior. The FLAT prior has a broader peak (comparing the reddish pink region where the value is above approximately half the maximum) compared to the other priors. Figure 6.37 shows the maximum posterior over portion variances of 1, 3 and 6, unbounded, plotted for batch and keg variance for various priors to facilitate comparison across portion values and priors. Note the peak for a portion value of 1 is too small to be seen. For the

FIGURE 6.32: **Posterior, unbounded, plotted for σ^2_{keg} and $\sigma^2_{portion}$ for σ^2_{batch}=3 for various priors**

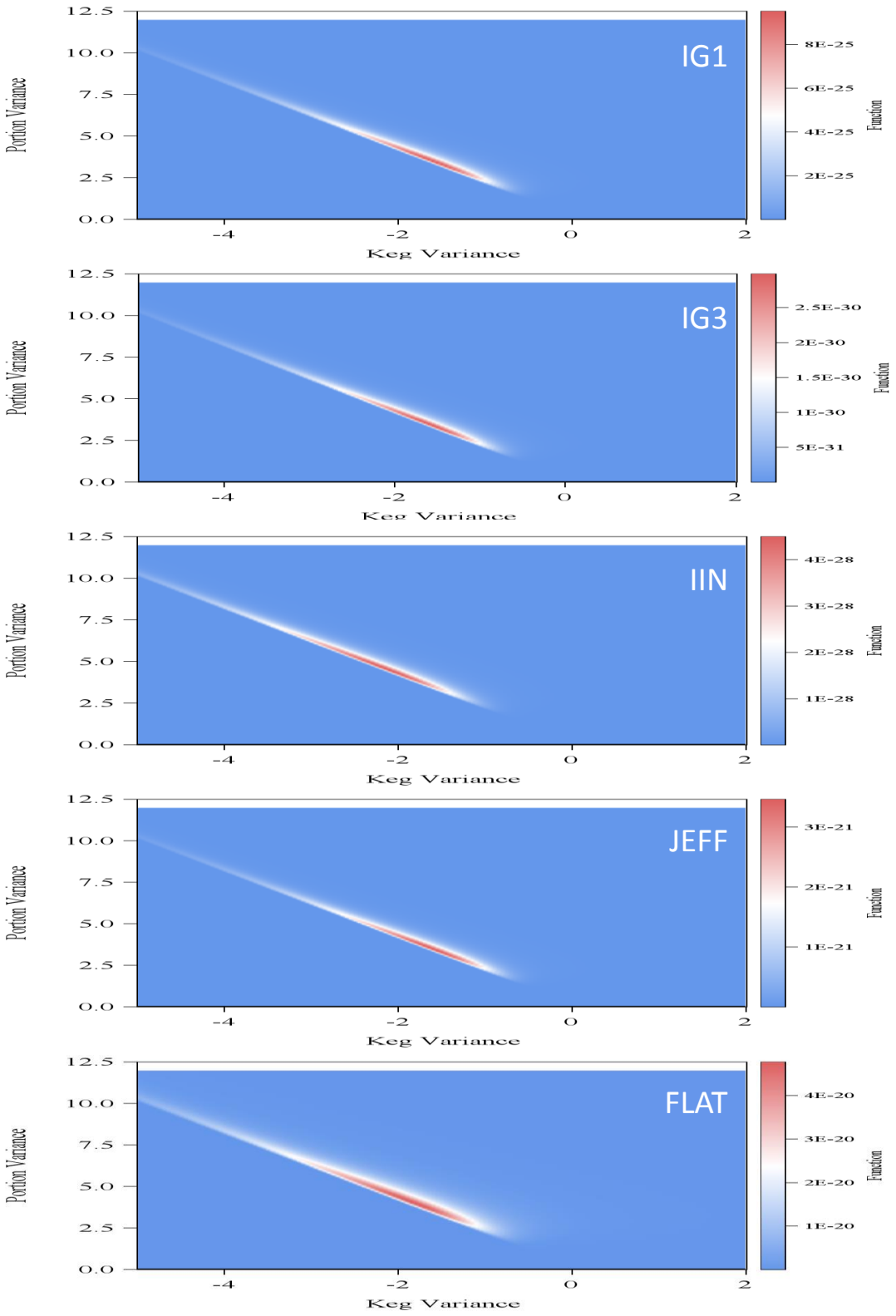
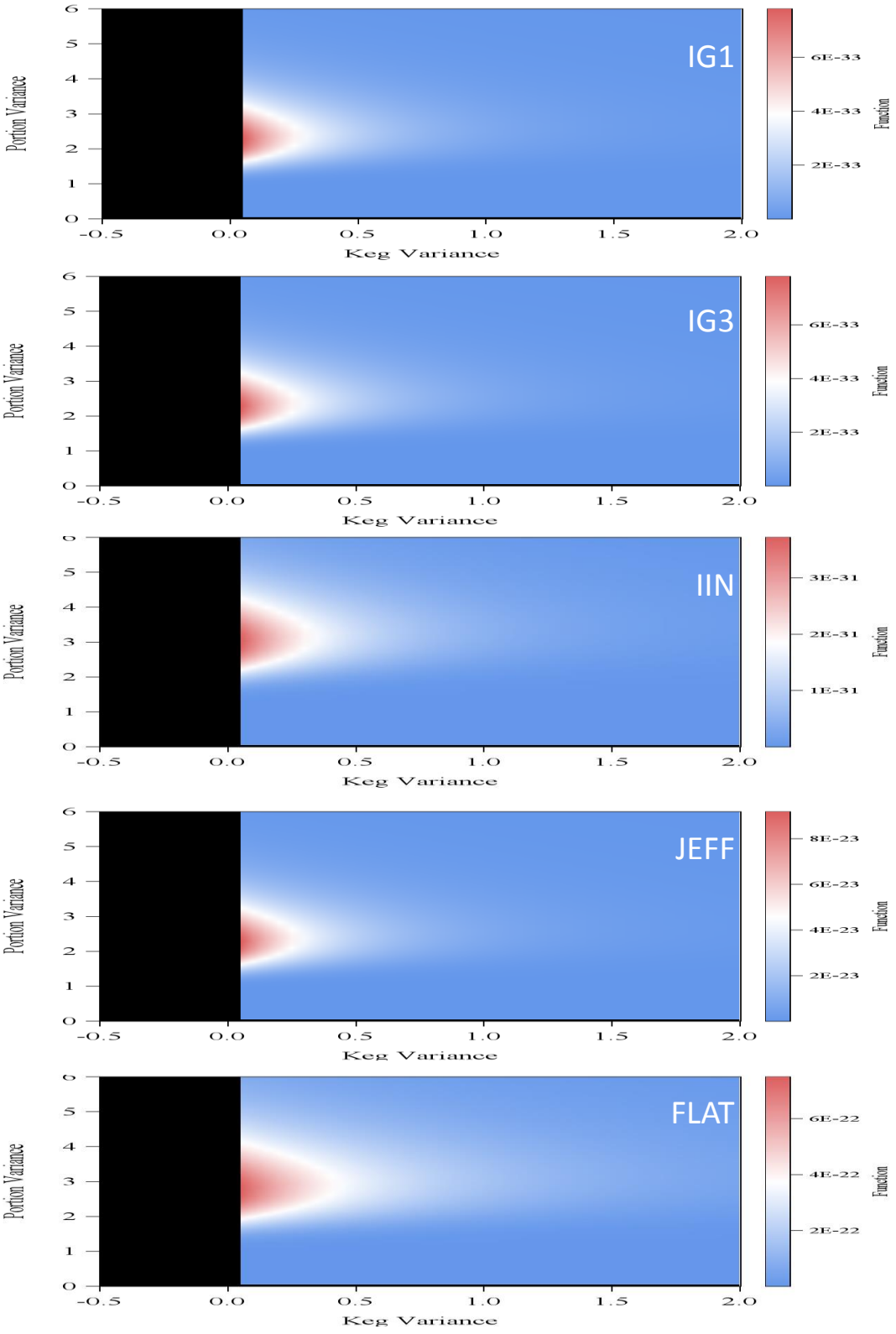


FIGURE 6.33: **Posterior, bounded, plotted for σ^2_{keg} and $\sigma^2_{portion}$ for σ^2_{batch}=3 for various priors**



IG1 and JEFF priors the highest peak is at a keg variance value of approximately -1.5 (corresponding to a portion variance of 3), for the FLAT prior the two peaks are more similar (corresponding to portion variances of 3 and 6), whereas for the IIN prior the largest peak is at a keg variance value of approximately -3. The latter corresponds to a portion variance of 6 which given the informative nature of the IIN prior for the portion variance can be expected.

Figure 6.38 shows the bounded posterior plotted for the batch and keg variance components for portion variances of 1, 3 or 6 and various priors. It is seen that for a portion variance of 3 or 6 the highest point is at $\sigma_k^2 = 0$ and $\sigma_b^2 = 0$. For a portion variance of 1 the highest point is at $\sigma_k^2 = 0$ but at a positive value for σ_b^2.

Finally the inverse gamma proposal distributions were examined for dataset 170 for priors IG1, IG3, IIN, JEFF and FLAT. The same inverse gamma proposal distribution is provided whether or not a bounded or unbounded analysis is requested. Aside from a multiplying constant, the density values provided by PROC MIXED for the proposal distributions were the same as the unbounded posterior. Thus for this balanced example for all the priors investigated, the unbounded posterior distributions on the strata are inverse gamma. The parameters for the default inverse gamma proposal distributions are shown in Table 6.25 for various priors.

TABLE 6.25: **Parameters for the default inverse gamma proposal distributions for various priors**

	Batch Stratum		**Keg Stratum**		**Portion Stratum**	
Prior	α	β	α	β	α	β
IG1	2.600	10.2977	3.100	1.29636	6.100	24.4481
IG3	2.501	10.1987	3.001	1.19736	6.001	24.3491
IIN	2.501	10.1987	3.001	1.19736	7.000	24.3481
JEFF	2.500	10.1977	3.000	1.19636	6.000	24.3481
FLAT	1.500	10.1977	2.000	1.19636	5.000	24.3481

It is seen that the FLAT prior has the smallest parameter values for the proposal distributions with the shape parameters being smaller than all the other priors. JEFF is almost the same as IG3. IIN is same as IG3 except for the proposal distribution on portion variance (as expected since the prior IIN is the same as IG3 except for placing a more informative prior on the portion variance) and IG1 is similar to IG3 but with slightly larger parameter values. Some modified proposal distributions for dataset 170 and an IG1 prior were then explored. One modification is where the scale parameters for the batch and keg strata are changed to 50 (S50). Another is where the scale parameters are changed to attempt to get 10, 20 or 50% of the distribution above zero (P10, P20 and P50 respectively. The IG proposal distributions for the batch stratum and portion variance were unchanged but the scale parameter for the keg stratum was changed to 4.364063 for the IG intended to provide 10% above zero and changed to

FIGURE 6.34: **Residual unbounded likelihood plotted for σ^2_{batch} and σ^2_{keg} for $\sigma^2_{portion}$=1,3 or 6**

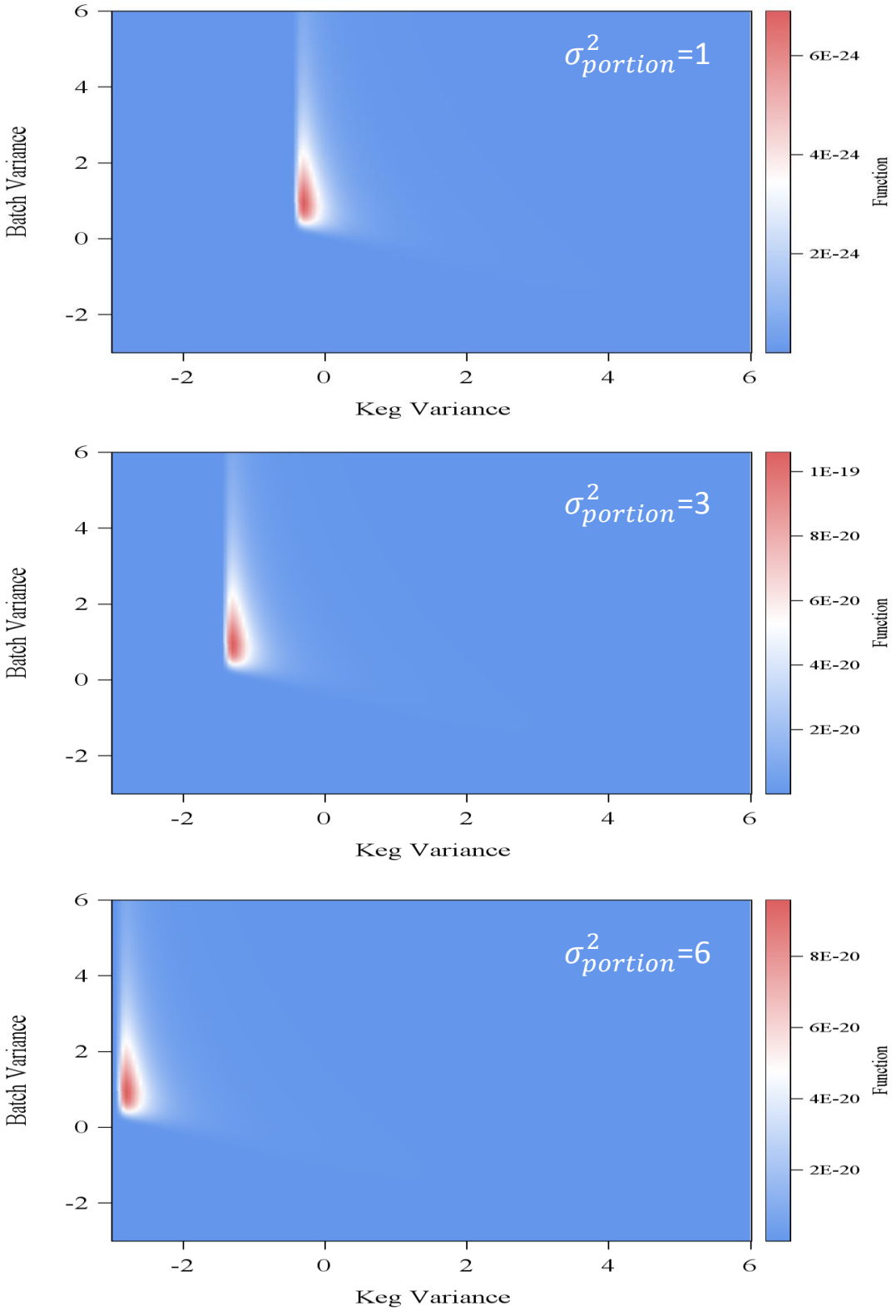
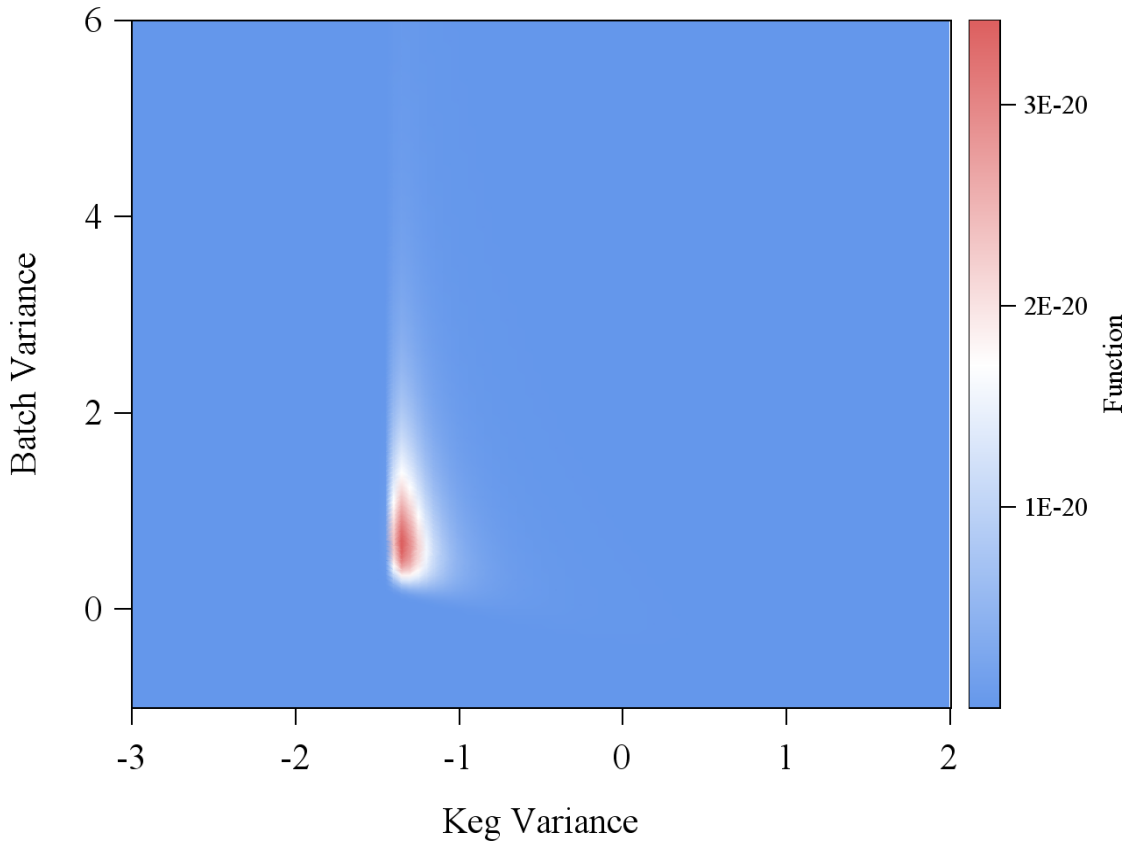


FIGURE 6.35: **Posterior, unbounded, plotted for σ^2_{batch} and σ^2_{keg} for JEFF prior and $\sigma^2_{portion}$=3**



6.240098 for the IG intended to provide 20% above zero and changed to 11.75503 for the IG intended to provide 50% above zero as per the algorithm in the previous section (Modifying Proposal Distributions on Strata to obtain part of distribution on variance component above zero). Table 6.26 shows the acceptance rates for the bounded analysis for the various proposal distributions (the sampling rates for the unbounded analysis are also shown for information). It is seen that as the boundary acceptance rate increases, the sampling acceptance rate decreases.

TABLE 6.26: **Acceptance rates for the various proposal distributions**

Proposal	Scale Param. Batch	Keg	Bounded: Accept.Rate Boundary	Sampling	Unbounded: Accept.Rate Boundary	Sampling
Default	10.30	1.30	0.00	1.00	1.00	1.00
S50	50	50	0.58	0.08	1.00	0.05
P10	10.30	4.36	0.05	0.82	1.00	0.25
P20	10.30	6.24	0.09	0.71	1.00	0.14
P50	10.30	11.76	0.17	0.51	1.00	0.10

FIGURE 6.36: **Posterior, unbounded, plotted for σ^2_{batch} and σ^2_{keg} for $\sigma^2_{portion}$ =1, 3 or 6**

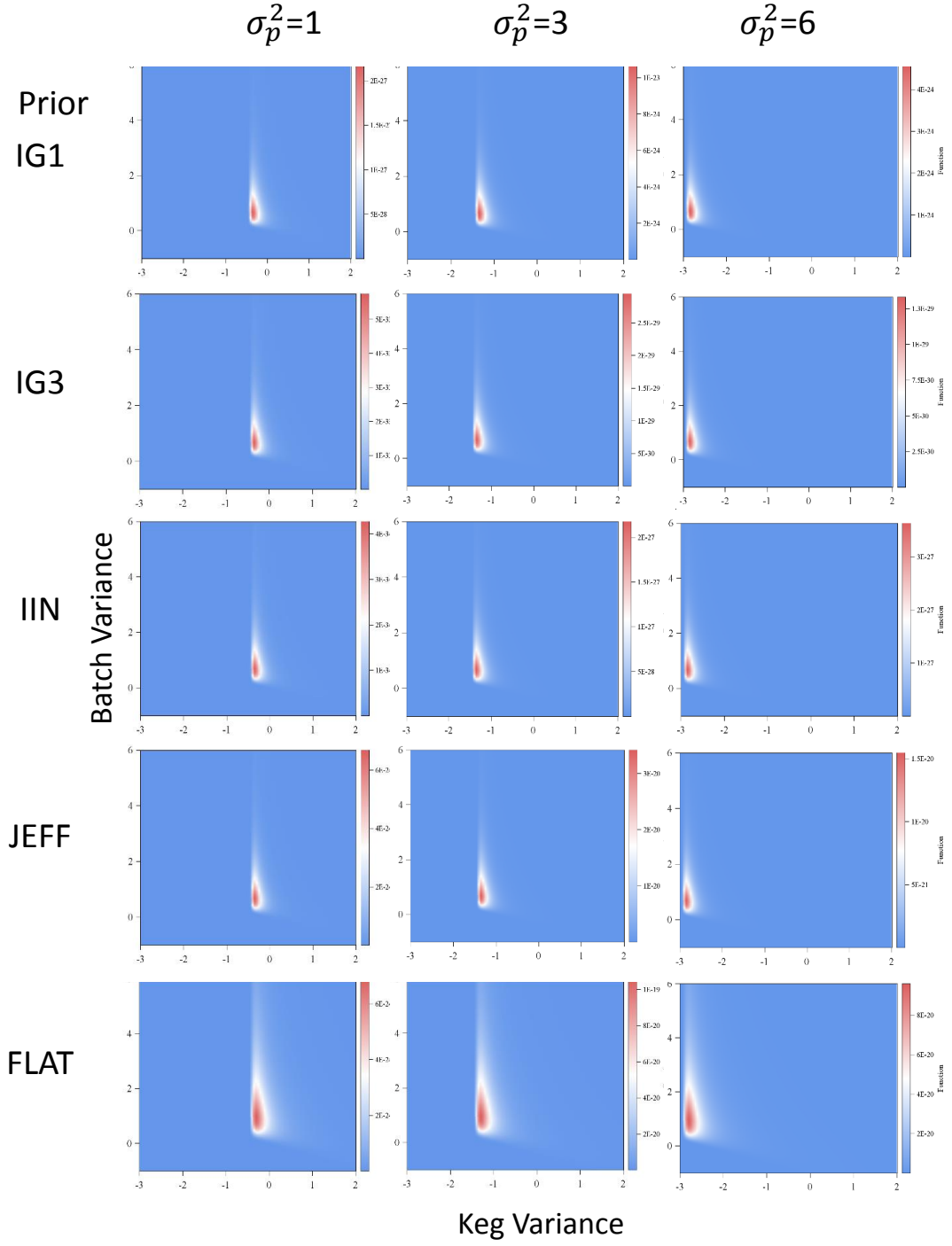


FIGURE 6.37: **Maximum posterior over $\sigma^2_{portion}$ of 1,3 and 6, unbounded, plotted for σ^2_{batch} and σ^2_{keg}**

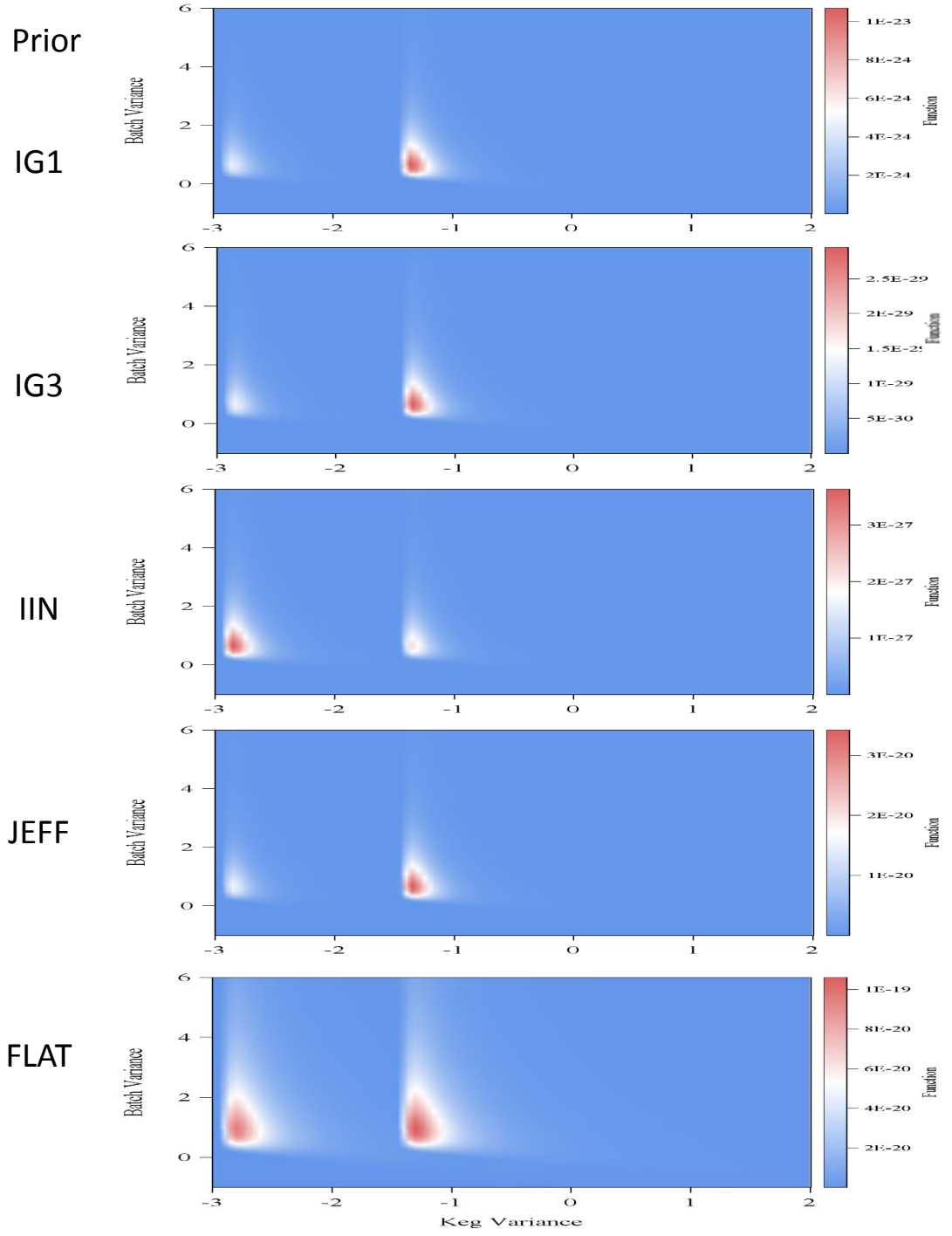
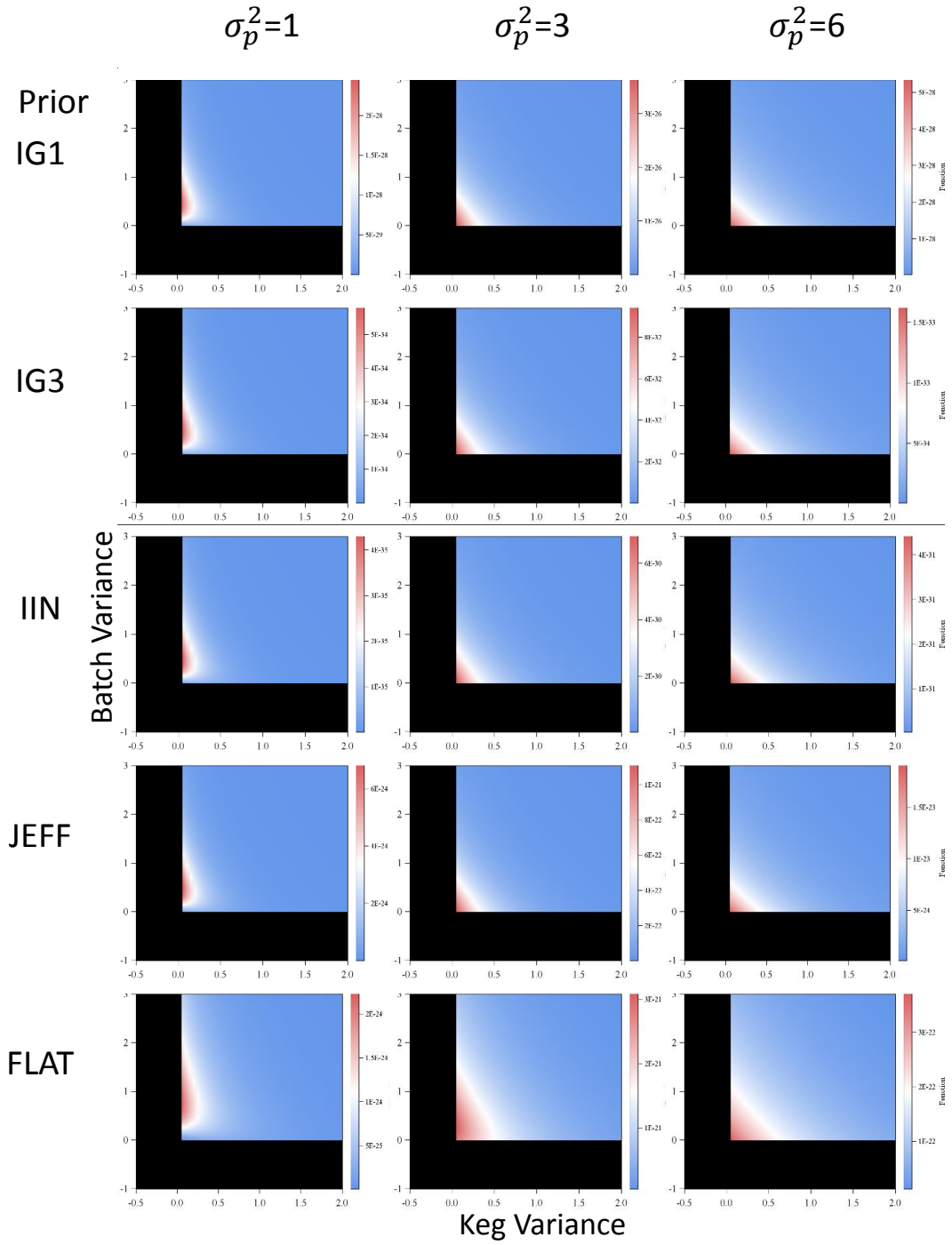


FIGURE 6.38: **Posterior, bounded, plotted for σ^2_{batch} ($\leq$3) and σ^2_{keg} ($\leq$2) for $\sigma^2_{portion}$ =1, 3 or 6(bounded)**



To understand further the relationship between boundary and sample acceptance rates Figure 6.39 shows the posterior for the bounded analysis with an IG1 prior, the default proposal distribution and the modified proposal distributions. It is seen that the peak of default proposal distribution is in the region where $\sigma_k^2 < 0$ and the density is low where $\sigma_k^2 > 0$. In contrast the peak of the modified distribution S50 is above zero and the density is low where $\sigma_k^2 < 0$. However, S50 now has part of the peak region where $\sigma_b^2 < 0$ and whilst the posterior has its peak at $\sigma_k^2 = 0$, the modified distribution has its peak at a $\sigma_k^2 > 4$ and no visible density close to 0. For modified proposal distributions P10 and P20 the peaks are still in the region where $\sigma_k^2 < 0$ as designed but closer to $\sigma_k^2 = 0$ and there is some visible density where $\sigma_k^2 > 0$. It is seen that some of this density occurs for $\sigma_b^2 < 0$. For P50 the peak lies on the boundary $\sigma_k^2 = 0$ as designed. However it is seen that where the $\sigma_k^2 > 0$ much of the peak is where $\sigma_b^2 < 0$. Figure 6.40 shows the posterior for the bounded analysis with an IG1 prior and some modified proposal distributions restricted to non-negative batch and keg variances. Note the default proposal distribution is not plotted since for this balanced design the proposal is the same as the posterior. It is seen that P10 which has a sampling acceptance rate of 0.82 seems most aligned with the distribution of the posterior, whereas P50 which has a sampling acceptance rate of 0.51 is slightly less so. S50 appears quite different and only has a sampling acceptance rate of 0.08.

The final strategy adopted for modifying the parameters of the IG proposal distributions is given in Section 6.4.4. In the analysis of a single dataset the inefficiency of low acceptance rates and consequently a high number of samples requested is not expected to be an issue. However, when performing simulations this does become an issue in the time taken.

6.4.3.5 Summary

For some analyses using the default IG proposal distribution in PROC MIXED in SAS software, sampling stopped and the requested number of posterior samples was not obtained. This occurred because the boundary acceptance rate was too low. Changing the seed did not resolve the issue, indicating that it was not due to an unfortunate bad stream of random numbers but related to the location of the chain. Nor did changing the sampling algorithm resolve the issue. The stopping was found to be associated with datasets where the REML estimate for the batch or keg variance was zero or very close to zero, though many datasets where the REML estimate was zero or very close to zero did not have sampling stopped. It happened more frequently for datasets with higher posterior median portion variance estimates; scenarios with smaller datasets and when the true keg or batch variances were 0.5 as opposed to 6 or 24. It was found that the IG proposal distribution is the same for the bounded and unbounded analysis and the prior is essentially truncated rather than censored. The stopping occurred when the

FIGURE 6.39: **Posterior for the bounded analysis with an IG1 prior, the default proposal distribution and various modified proposal distributions, plotted at $\sigma^2_{portion}$=3**

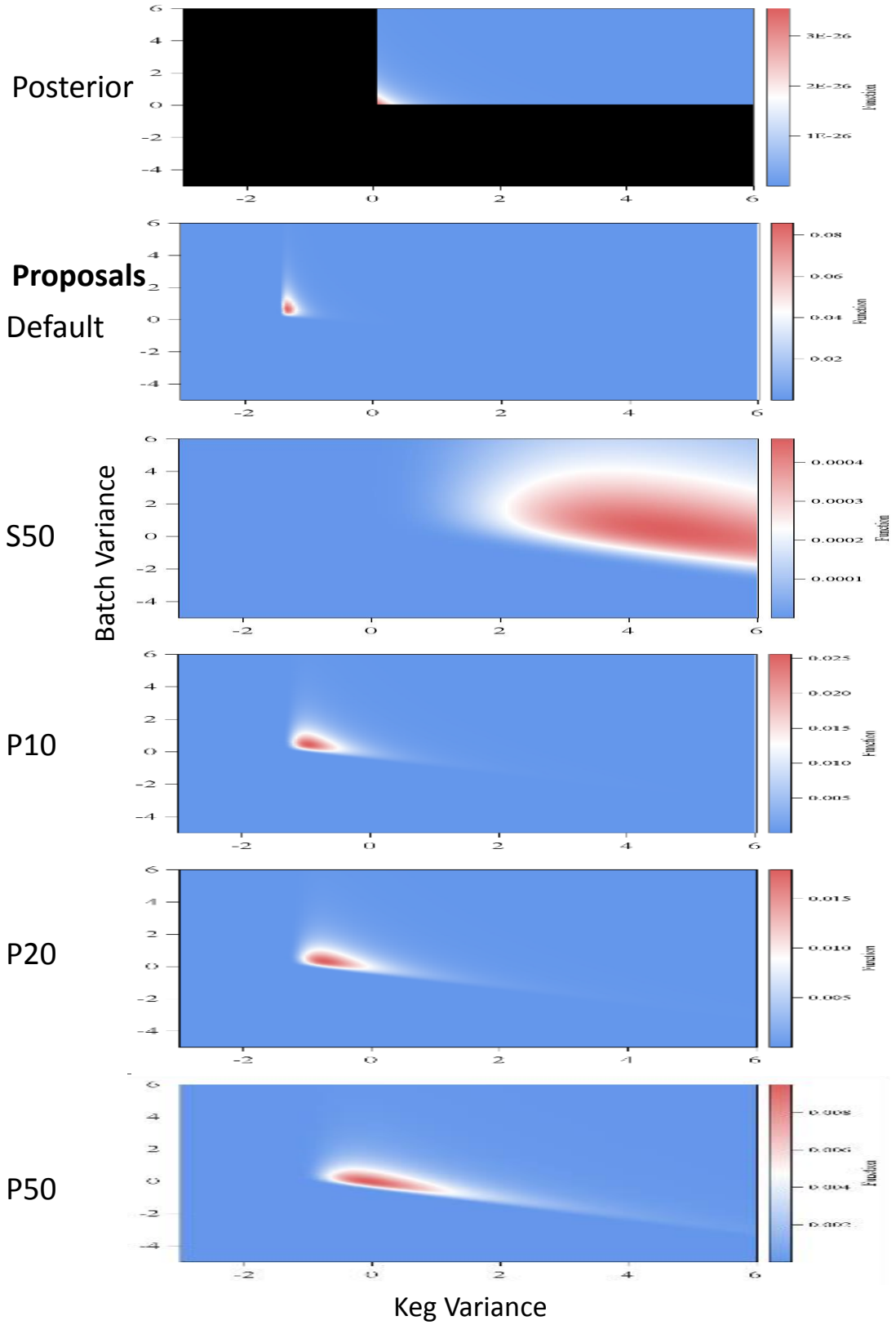
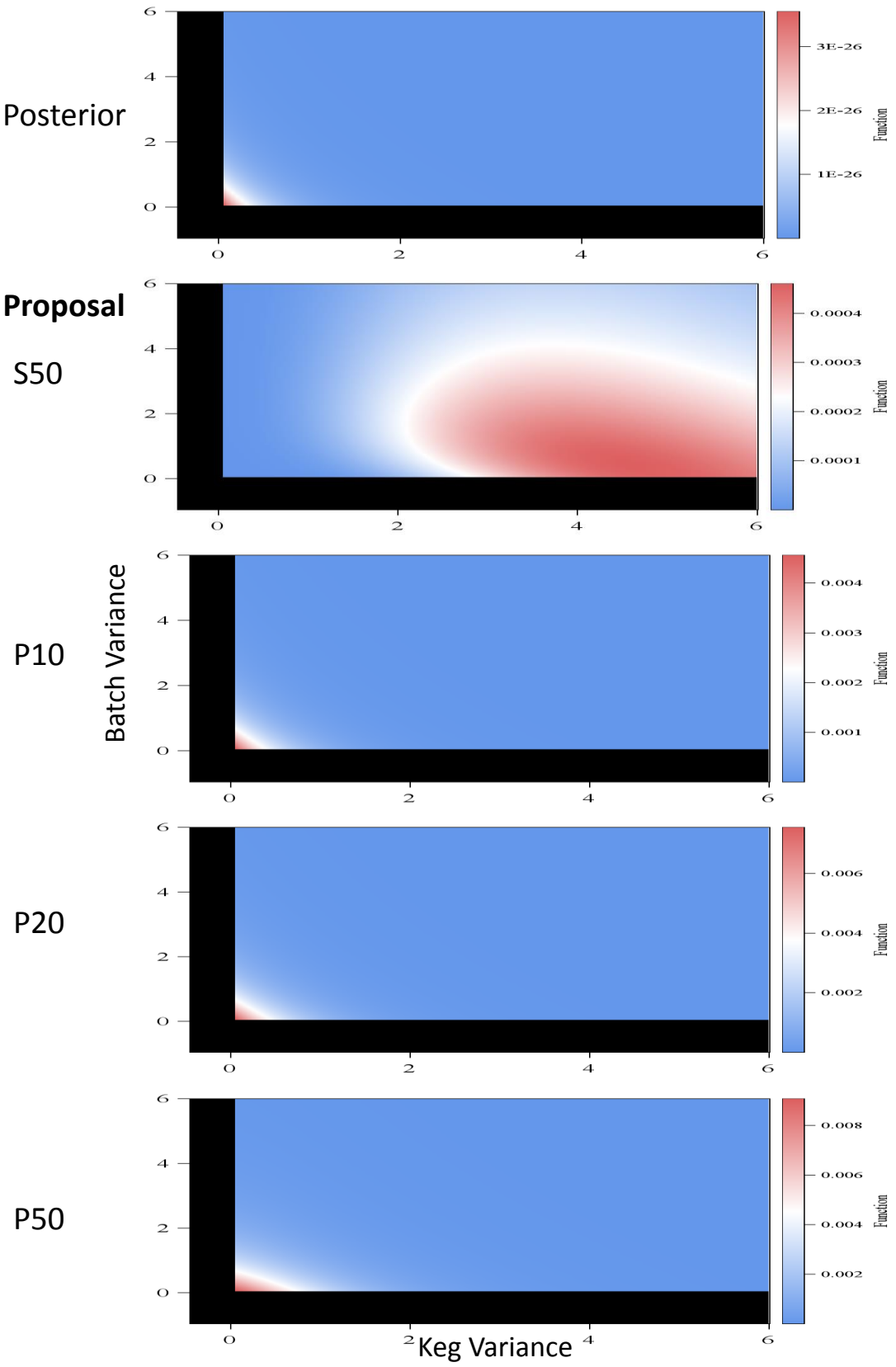


FIGURE 6.40: **Posterior for the bounded analysis with an IG1 prior and modified proposal distributions restricted to non-negative values of σ^2_{batch} and σ^2_{keg}, plotted at $\sigma^2_{portion}$=3**



proposal distributions for the stratum variances resulted in marginal densities for batch and keg variance components with low density for non-negative values. There was also different behaviour with analyses using the FLAT prior for scenario b3k2p2_0.5 where the sampling acceptance rate was typically much less than 1 for the default proposal distribution (in contrast to other scenarios or priors where the sampling acceptance rate was generally close to 1). Also often the analyses gave 1 or 2 samples less than the number requested. It seems likely this is a consequence of the behaviour of the SAS software procedure connected with an improper prior. However, some datasets for scenario b3k2p2_0.5 gave the full sample from the posterior requested for the FLAT prior and thus an improper posterior may not be detectable from the behaviour of the analysis.

Neither the SAS software documentation nor the literature provided a solution to address the situation when sampling stopped. Changing the parameters of the IG proposal distribution changed the boundary and sampling acceptance rates. Since typically changing the parameters to increase the boundary acceptance rate resulted in a distribution less like that of the posterior the sampling acceptance rate reduced. Thus an alternative IG distribution needed to be found with high enough boundary and sampling acceptance rates. It was seen that a default fixed proposal distribution could not be used across all datasets which originally failed to give a complete sample from the posterior as though it might give a satisfactory boundary acceptance rate, for some datasets the sampling acceptance rate dropped too low and the sampling from the posterior stopped for that reason. Thus an R program was developed that attempted to identify IG distributions which gave 10 or 20% of the marginal distributions for the batch or keg variance above zero. This made use of an approximation by Witkovsky (2001) for the linear combination of inverse gamma distributions. Occasionally the optimisation algorithm failed in which case other changes were made to the IG parameters to seek a full sample as described in Section 6.4.4. Further investigation of the proposal, likelihood and posterior illustrated that the difficulty in analyses stopping is due to the IG proposal distribution being derived from the unbounded posterior, even when a bounded analysis has been selected. This is a particular issue with small datasets and small variance components. Thus with (effectively) only IG proposal distributions available in the SAS software PROC MIXED, in order to obtain a full sample it is necessary to try to find IG proposal distributions for which the resultant marginal distributions for the variance components have not minimal density for non-negative values and also have some density where the posterior function is high.

6.4.4 Strategy Applied when Posterior Sampling Stopped

Strategy for Modifying Proposal Distribution For analyses where the independence chain algorithm stopped sampling from the posterior using the SAS software

default proposal distribution, a sequence of modifications were made to the parameters of the inverse gamma proposal distribution to attempt to get a full sample:

(a) Modification of inverse gamma proposal distributions using a program I developed in R. The resultant distributions for the batch and keg variance components were derived from the default inverse gamma proposal distributions on the batch, keg and portion strata by implementing Witkovsky (2001) results for computing the distribution of a linear combination of inverted gamma variables. The parameters of the inverse gamma distributions were then modified in order to try to obtain 10 or 20% of the distributions for the batch and keg variance components above 0.

(b) Increasing the scale parameter of the inverse gamma distributions for batch or keg strata by multiplying by 10

(c) Using the proposal distributions from ten other datasets from which the full sample from the posterior was obtained

(d) Setting the scale parameter for the inverse gamma distributions for the batch and keg strata to 50.

(e) Systematic screen of setting values for the scale parameters of inverse gamma distributions for the keg and batch strata e.g. all 25 combinations of keg and batch strata scale parameters set to 1,21,41,61,81.

In implementing this procedure, the datasets which stopped initially were re-run using the default proposal distribution to retain the default proposal distribution before proceeding to modify the distribution. Occasionally on this re-run, the full posterior sample was obtained and when summarised the dataset is included as reaching a full sample under the default proposal distribution.

Results of applying strategy for obtaining a full sample For the original batch sampling study design a full sample was obtained for each dataset with either the default proposal distribution or the modified proposal distribution for all analyses. For each set of 10,000 datasets, Table 6.27 shows the number of datasets where the default distribution gave a full sample, and the number which required a modified distribution. The results are shown for each prior and true batch variance used to generate the datasets.

The number of datasets with full samples obtained by the default or requiring a modified proposal distribution are summarised in Table 6.28 for the various design scenarios. There were just two analyses where a full sample was not achieved (scenario b3k2s2_0.5, batch variance = 0.5, FLAT prior).

Thus the strategy for dealing with stopped datasets worked for all but two stopped datasets. The two remaining stopped datasets were for the FLAT prior, scenario b3k2s2_0.5 where the posterior is improper.

TABLE 6.27: **Number of datasets with full posterior sample achieved from default proposal distribution or requiring modified proposal distribution for original batch sampling design**

	Batch Variance					
	0.5		6		24	
	Proposal Dist.		**Proposal Dist.**		**Proposal Dist.**	
Prior	Default	Mod.	Default	Mod.	Default	Mod.
IG0	9890	110	9969	31	9974	26
IG1	9964	36	9986	14	9991	9
IG2	9969	31	9986	14	9992	8
IG3	9968	32	9986	14	9992	8
IIN	9966	34	9986	14	9992	8
I2O3	10000	0	10000	0	10000	0
I2O6	10000	0	10000	0	10000	0
I1O3	10000	0	10000	0	10000	0
I1O6	10000	0	10000	0	10000	0
I1E3	10000	0	10000	0	10000	0
I1E6	10000	0	10000	0	10000	0
IhE6	10000	0	10000	0	10000	0
JEFF	9970	30	9986	14	9992	8
FLAT	9995	5	9999	1	10000	0

6.4.5 Autocorrelation and Number of Samples from Posterior

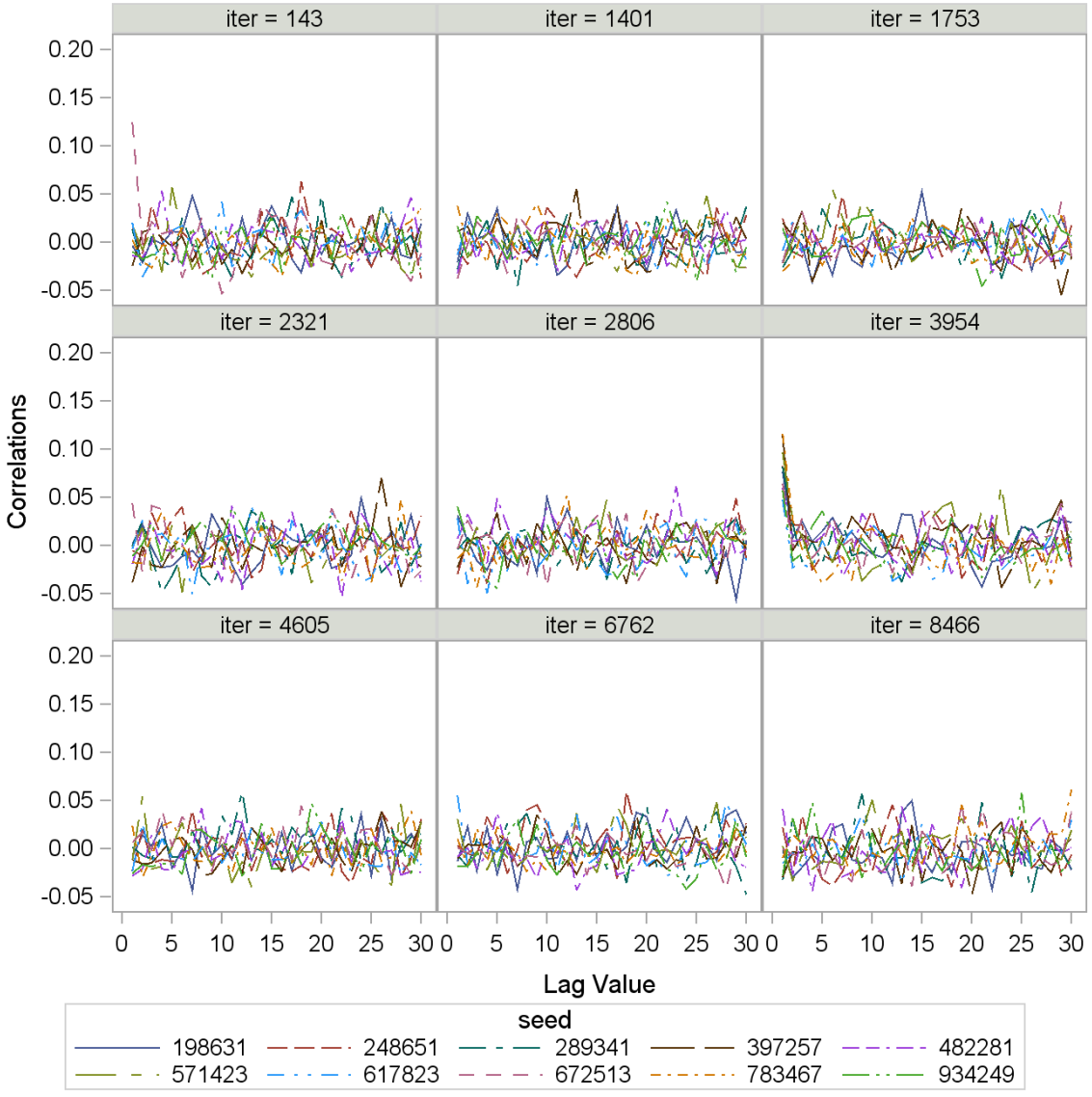
It was considered desirable to achieve the equivalent of at least 2000 independent samples from the posterior. The rationale for this was given in Section 6.3.3. For the initial analyses in SAS software 2500 samples were requested from the posterior.

The independence chain sampling algorithm was described in Section 6.4.1.1. Independent proposals are drawn from the proposal distribution. If the proposal distribution is proportional to the posterior then the proposal will be retained with probability 1 and thus the posterior samples will be independent. It is seen in Figure 6.19, Section 6.4.3.2 and Appendix E.3 that, except for the FLAT prior and scenario b3k2p2_0.5 (where the posterior is improper), the sampling acceptance rates were close to 1 for almost all analyses. However, when analyses using the default proposal distribution stopped and the analysis was repeated with a modified proposal distribution then the sampling acceptance rate is lower and some duplicate samples are added to the chain. This induces autocorrelation into the posterior samples.

Autocorrelation was investigated using analyses for the IG1 prior, 6 batches, 2 kegs/-batch and 2 portions/keg and both batch and keg variance components of 0.5. Firstly as a check that posterior samples are typically independent for the default proposal distribution and for comparison when the proposal distribution is modified the autocorrelation was plotted for some example datasets where the full 2500 samples were obtained from

the analysis using the default proposal distribution. Datasets were selected with boundary acceptance rates in the range 0.05 - 0.90. Note the sampling acceptance rates were 1. Figure 6.41 shows that for these datasets the estimated autocorrelation was fairly low, almost entirely below 0.1.

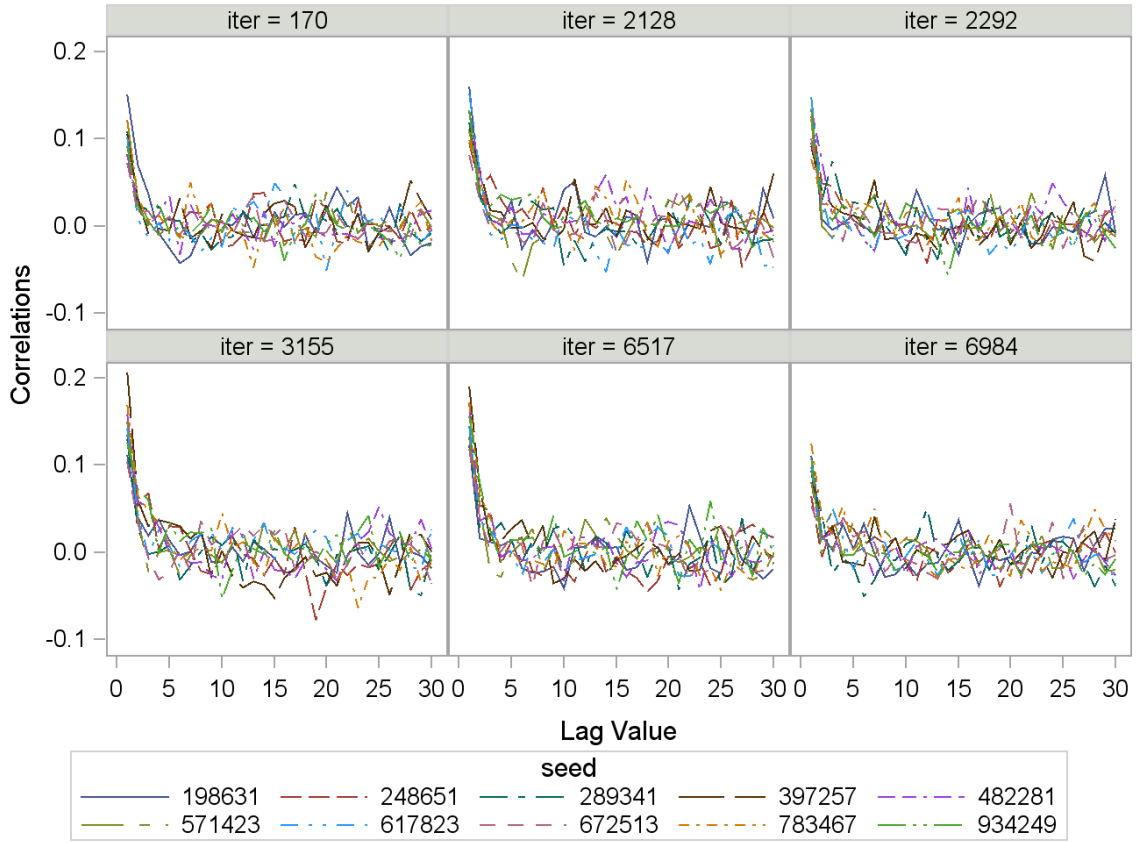
FIGURE 6.41: **Autocorrelations for keg variance for example datasets with default proposal distributions**



Secondly the autocorrelation of the samples where the dataset stopped sampling for the default proposal distribution and the proposal distribution was modified to aim for 10, 20 or 50% of the keg or batch variance proposal distributions above zero (denoted P10, P20 and P50 respectively) was examined. In addition analyses where the scale parameter values for the IG proposal distributions for the batch and keg variance strata are changed to 50 (denoted S50) are examined. The autocorrelation of the datasets

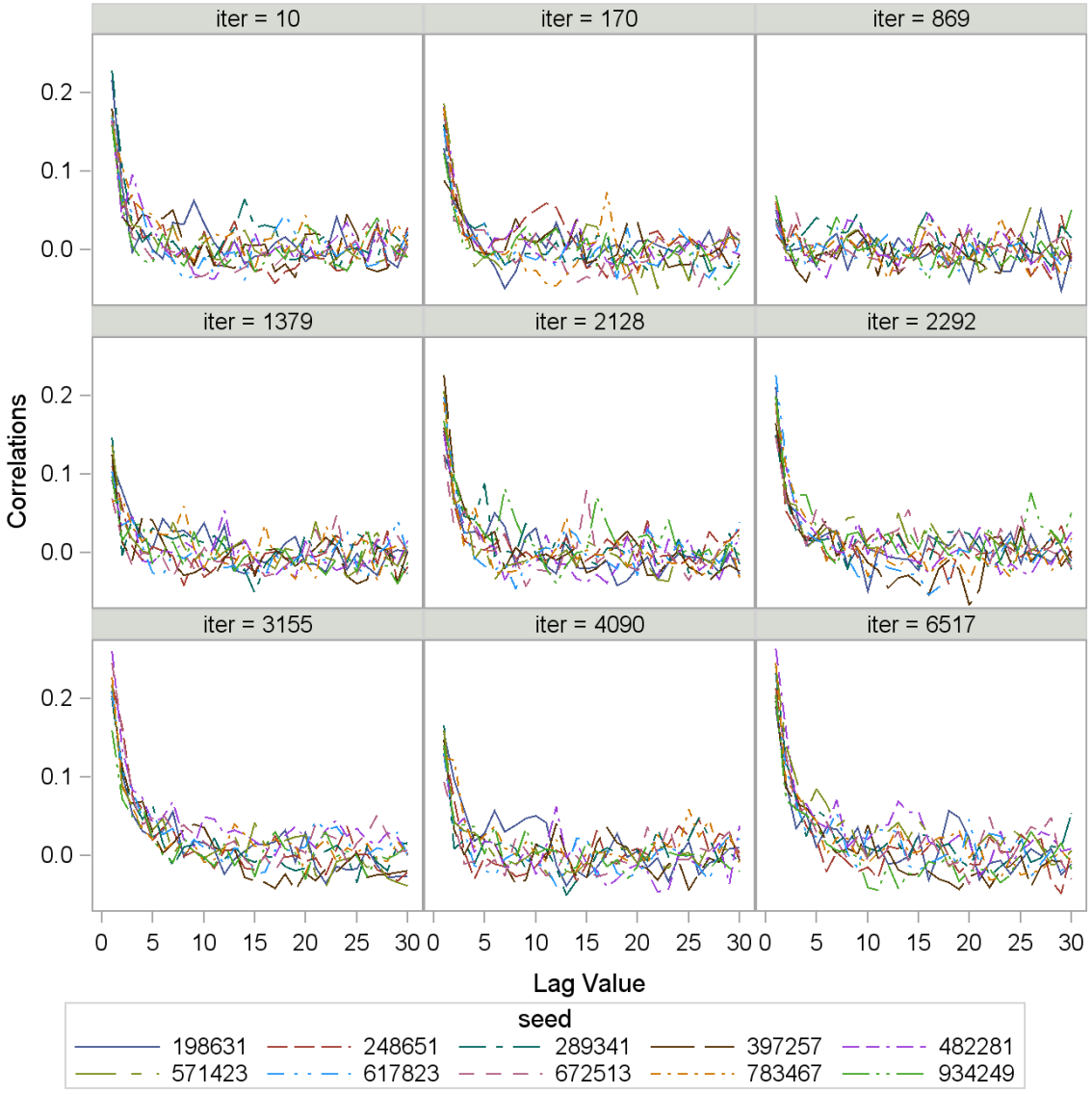
identified for further investigation at the beginning of Section 6.4.3.4 were evaluated and the autocorrelation for the first nine datasets to have 2500 samples from the posterior for the modified analyses are plotted in Figures 6.42 - 6.45 for σ_k^2. Note only 6 datasets provided full samples from the posterior with the P10 proposal distribution. Similar autocorrelation patterns were seen for σ_b^2 and these are given in Appendix E.6. It is seen that unlike in Figure 6.41 for the default proposal distribution, there is some autocorrelation for small lags. For P10 and P20 the autocorrelation doesn't persist for larger lags but for P50 it does for some of the datasets. The autocorrelation suggests that more samples should be taken from the posterior when the proposal distribution is modified and it is preferable to only increase the proportion of the distribution above zero sufficiently to overcome the boundary acceptance rate being too low for sampling to continue.

FIGURE 6.42: **Autocorrelations for keg variance for example datasets with P10 proposal distributions**



Thus a larger number of samples from the posterior were requested for the analyses with modified proposal distributions. The number chosen was 100,000 which meant that even if the sampling acceptance rate was 2%, 2,000 independent samples would be obtained from the proposal distribution along with the repeats, though the retention of previous

FIGURE 6.43: **Autocorrelations for keg variance for example datasets with P20 proposal distributions**

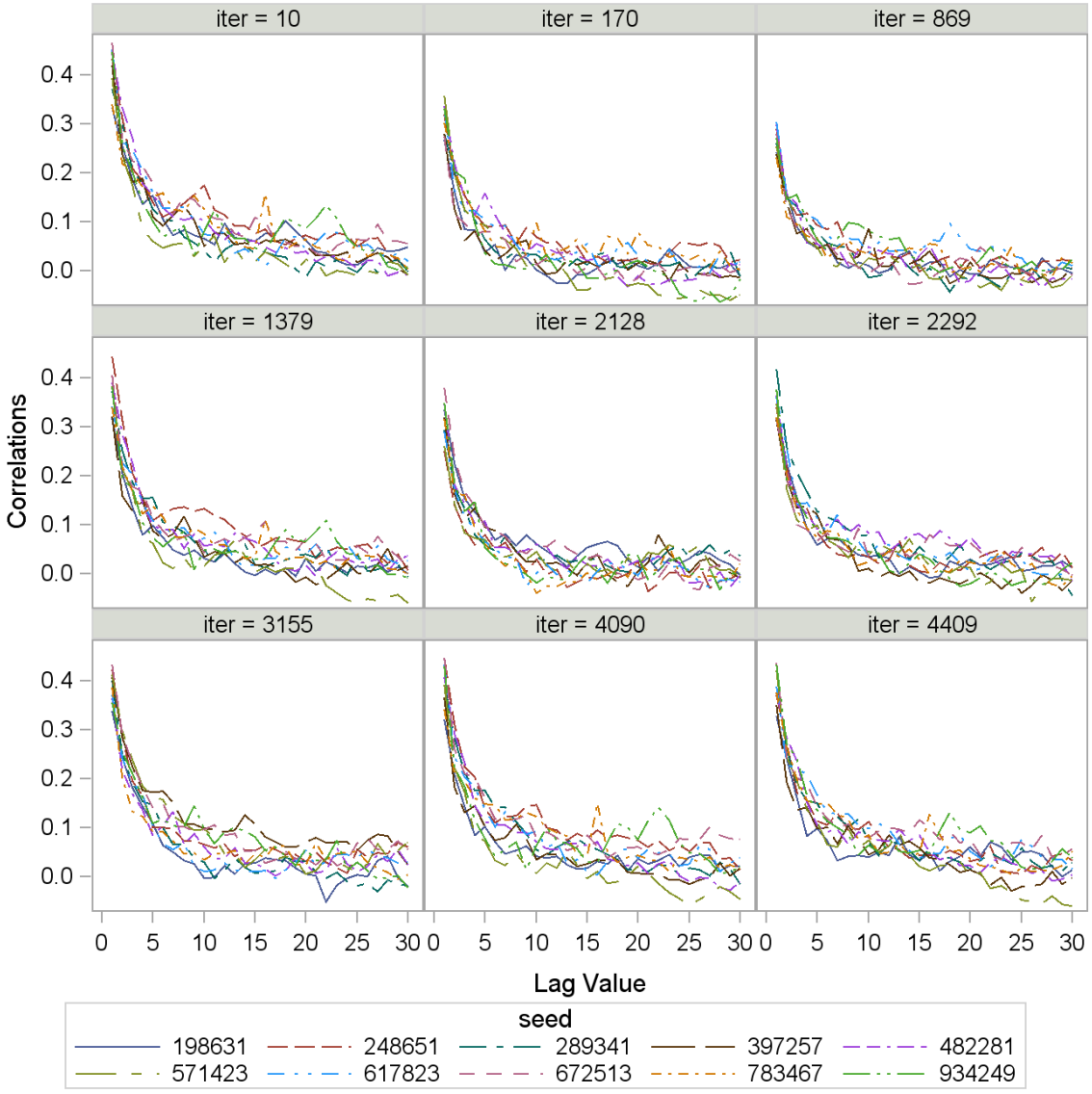


proposals when proposals are rejected means that there will be autocorrelation amongst the 100,000 samples.

The effective sample size (ESS) was calculated for all analyses with modified proposal distributions except the FLAT prior and scenario b3k2p2_0.5. The latter was not performed as it was known that the posterior is improper and given the large number of analyses to be repeated with a modified proposal distribution the additional calculations were not deemed to add value. The formula for ESS is given by

$$\frac{n}{1 + 2\Sigma_{k=1}^{\infty}\rho(k)},$$

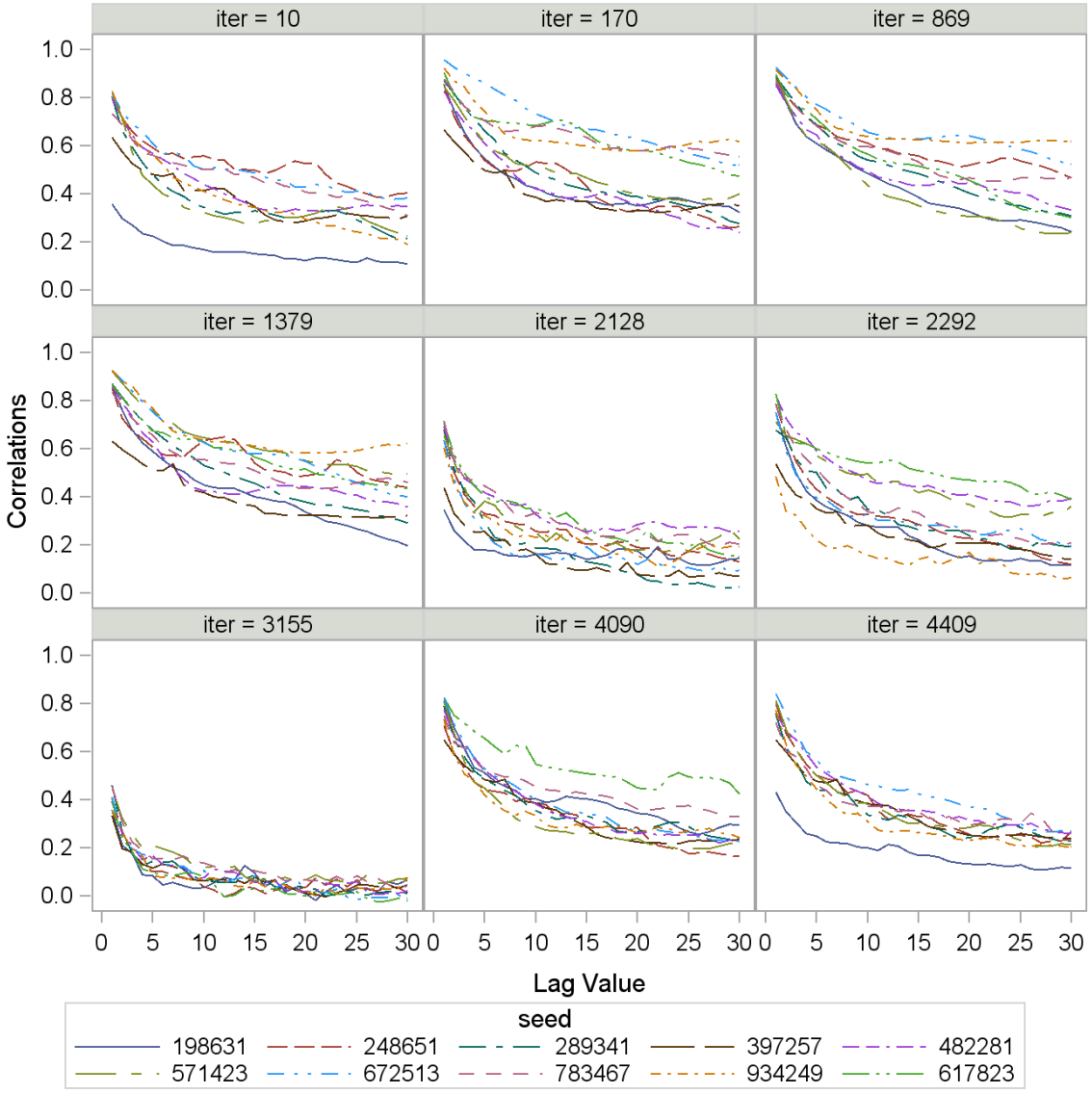
FIGURE 6.44: **Autocorrelations for keg variance for example datasets with P50 proposal distributions**



where n is the number of samples from the posterior and $\rho(k)$ is the correlation at lag k. In SAS software ∞ is replaced by $k_m = \max(500, \frac{\text{number of posterior samples}}{4})$ for the upper limit of the sum in the denominator of the equation.

Across all scenarios, priors, true batch and keg variances,excepting the FLAT prior and scenario b3k2p2_0.5 there were 1,680,000 analyses. For the default proposal distribution 1,661,891 analyses obtained the full 2500 samples requested from the posterior. Table 6.29 shows the number of analyses with modified proposal distributions classified by various levels of ESS. The number of analyses is also expressed as a percentage of all 168,000 analyses. The variance and metric with the largest percentage of analyses with ESS < 2000 is the Total SD. The metric made little difference, though the SD metric

FIGURE 6.45: **Autocorrelations for keg variance for example datasets with S50 proposal distributions**



gives slightly lower ESS than the variance metric. Table 6.30 classifies the number of analyses with ESS < 2000 by prior and scenario.

It is seen that, in general, scenario b3k2p2_0.5 had the highest percentage of datasets with ESS < 2000, though not for the more informative IG0 prior. The IG0 prior had a greater percentage for the scenario k2p2_0.5 and for all scenarios except b3k2p2_0.5 had a higher percentage than the other priors. In contrast to the effective sample sizes for the Gibbs sampling algorithm some analyses for the portion variance had ESS less than 2000. This is likely due to the nature of the algorithm in SAS software where if the boundary condition for the keg or batch variance is violated the previous proposal samples containing all parameters of interest are retained in contrast to Gibbs sampling

which samples single or subgroups of the parameters. Given the small percentage of analyses with ESS < 2000 (0.4% for Total SD and less for other variance components and metrics) and for the analyses in SAS software a total of 10,000 datasets were analysed (compared to 2000 for the Gibbs sampling) 100,000 posterior samples was considered sufficient when the proposal distribution was modified.

TABLE 6.28: **Number of datasets with full posterior sample achieved from default proposal distribution or requiring modified proposal distribution for the various scenarios**

		Batch Variance					
		0.5		6		24	
		Proposal Dist.		Proposal Dist.		Proposal Dist.	
Scenario	**Prior**	Default	Mod.	Default	Mod.	Default	Mod.
b3k2p2_0.5	**IG0**	8641	1359	9428	572	9715	285
	IG1	9312	688	9708	292	9847	153
	IG2	9367	633	9724	276	9854	146
	IG3	9376	624	9728	272	9851	149
	IIN	9167	833	9642	358	9801	199
	JEFF	9380	620	9727	273	9850	150
	FLAT*	5184	4814	7043	2957	7750	2250
k2p2_0.5	**IG0**	9097	903	9797	203	9866	134
	IG1	9521	479	9879	121	9918	82
	IG2	9558	442	9886	114	9922	78
	IG3	9560	440	9887	113	9920	80
	IIN	9420	580	9849	151	9892	108
	JEFF	9560	440	9887	113	9920	80
	FLAT	9942	58	9990	10	9991	9
k2p16_6	**IG0**	9785	215	9965	35	9997	3
	IG1	9906	94	9990	10	9999	1
	IG2	9916	84	9990	10	9999	1
	IG3	9920	80	9990	10	9999	1
	IIN	9920	80	9990	10	10000	0
	JEFF	9919	81	9990	10	10000	0
	FLAT	9992	8	10000	0	10000	0
k2p16_24	**IG0**	9739	261	9854	146	9983	17
	IG1	9905	95	9953	47	9994	6
	IG2	9912	88	9955	45	9995	5
	IG3	9916	84	9956	44	9995	5
	IIN	9917	83	9956	44	9995	5
	JEFF	9917	83	9957	43	9995	5
	FLAT	9992	8	9993	7	10000	0
k6p16_0.5	**IG0**	9986	14	10000	0	10000	0
	IG1	9997	3	10000	0	10000	0
	IG2	9997	3	10000	0	10000	0
	IG3	9997	3	10000	0	10000	0
	IIN	9997	3	10000	0	10000	0
	JEFF	9997	3	10000	0	10000	0
	FLAT	10000	0	10000	0	10000	0
k6p2_0.5	**IG0**	9743	257	9967	33	9964	36
	IG1	9880	120	9974	26	9970	30
	IG2	9888	112	9975	25	9971	29
	IG3	9887	113	9975	25	9971	29
	IIN	9862	138	9969	31	9964	36
	JEFF	9890	110	9975	25	9972	28
	FLAT	9968	32	9980	20	9978	22

* For batch variance = 0.5, FLAT prior, 2 stopped datasets

TABLE 6.29: **Effective sample size achieved from analyses with modified proposal distribution and as a percentage of all analyses, for all scenarios and priors except FLAT prior and scenario b3k2p2_0.5**

		Total		Batch		Keg		Portion	
	ESS	**SD**	**Var.**	**SD**	**Var.**	**SD**	**Var.**	**SD**	**Var.**
No.	**0-1000**	4959	4653	4216	3877	2208	2216	2360	2295
	1000-1500	996	871	920	684	1358	1151	1126	1064
	1500-2000	655	508	602	414	961	909	940	1163
	2000-4000	1194	1143	1346	1031	2247	1714	2788	2732
	>4000	10305	10934	11025	12103	11335	12119	10895	10855
%	**0-1000**	0.3	0.28	0.25	0.23	0.13	0.13	0.14	0.14
	1000-1500	0.06	0.05	0.05	0.04	0.08	0.07	0.07	0.06
	1500-2000	0.04	0.03	0.04	0.02	0.06	0.05	0.06	0.07
	2000-4000	0.07	0.07	0.08	0.06	0.13	0.1	0.17	0.16
	>4000	0.61	0.65	0.66	0.72	0.67	0.72	0.65	0.65
	2500*	98.92	98.92	98.92	98.92	98.92	98.92	98.92	98.92

* Number of posterior samples from analyses with default proposal distribution

TABLE 6.30: **Percentage of datasets with ESS < 2000 for total SD classified by scenario and prior**

	Scenario						
Prior	**b3k2p2 _0.5**	**k2p2 _0.5**	**k2p16 _0.5**	**k2p16 _6**	**k2p16 _24**	**k6p16 _0.5**	**k6p2 _0.5**
IG0	0.54	3.17	0.14	0.54	0.98	0.00	0.50
IG1	2.25	0.81	0.00	0.00	0.01	0.00	0.24
IG2	2.32	0.50	0.00	0.00	0.01	0.00	0.25
IG3	1.83	0.49	0.00	0.00	0.01	0.00	0.24
IIN	2.78	0.69	0.00	0.00	0.01	0.00	0.31
I2O3	-	-	0.00	-	-	-	-
I2O6	-	-	0.00	-	-	-	-
I1O3	-	-	0.00	-	-	-	-
I1O6	-	-	0.00	-	-	-	-
I1E3	-	-	0.00	-	-	-	-
I1E6	-	-	0.00	-	-	-	-
IaE6	-	-	0.00	-	-	-	-
JEFF	2.22	0.50	0.00	0.00	0.01	0.00	0.24
FLAT	nc*	0.25	0.02	0.03	0.04	0.00	0.12

* Not calculated

6.4.6 Summary

PROC MIXED in SAS software provides a means of sampling from the posterior primarily using inverse gamma related proposal distributions. These were used in the evaluation of the IG, JEFF and FLAT priors. The procedure provides four sampling algorithms: independence chain, importance sampling, rejection sampling and random walk chain. These were investigated in Section 6.4.1 and the independence chain algorithm chosen in Section 6.4.1.4 for the evaluation of the priors as it was expected to provide speed advantages over the others.

For some datasets/analyses the algorithm stopped sampling from the posterior and a full sample from the posterior was not obtained. Investigations were performed to understand the reason for this. It was found to be due to the default IG proposal distribution for the stratum variances resulting in marginal distributions for the batch and keg variance components with low density for non-negative values. This was found to be associated with datasets where the REML estimate for the batch or keg variance was zero or very close to zero, though many datasets where the REML estimate was zero or very close to zero did not have sampling stopped. It happened more frequently for datasets with higher posterior median portion variance estimates; scenarios with smaller datasets and when the true keg or batch variances were 0.5 as opposed to 6 or 24. It is important in these situations to be able to perform a Bayesian analysis, particularly as a main aim of the work is to be able to estimate the uncertainty of variance component estimates for small studies. Thus an approach was developed to overcome this.

Neither the SAS software documentation nor the literature provided a solution to address the situation when sampling stopped. A procedure was developed that changed the parameter values of the IG proposal distribution to find one that achieved high enough boundary and sampling acceptance rates simultaneously in order that sampling didn't stop. A default fixed proposal distribution for every dataset could not meet this requirement and thus a procedure based on the data was developed. This procedure was programmed in R and called from the SAS program. The procedure attempted to identify IG distributions which gave 10 or 20% of the marginal distributions for the batch or keg variance above zero using an approximation by Witkovsky (2001) for the linear combination of inverse gamma distributions. Occasionally the optimisation algorithm failed in which case other changes were made to the IG parameter values to seek a full sample. The full strategy was described in Section 6.4.4 and this strategy achieved a full sample from the posterior in all but two analyses.

It is also noted that there was different behaviour with analyses using the FLAT prior for scenario b3k2p2_0.5 where the sampling acceptance rate was typically much less than 1 and often the analyses gave 1 or 2 samples less than the number requested. It seems likely this is a consequence of the behaviour of the SAS software procedure connected with an improper prior. However, some datasets gave the full sample from the posterior

requested and thus it will not always be detectable from the behaviour of the analysis when a posterior is improper.

Modifying the proposal distribution results in a sampling acceptance rate less than 1. When a proposal is rejected the previous proposal is retained inducing an autocorrelation into the posterior samples. The aim was at least 2000 independent samples and in the analyses with the default proposal distributions 2500 were requested to accommodate the occasional sampling rate less than 1. For analyses with modified proposal distributions 100,000 samples were requested, meaning that even if the sampling rate was as low as 0.02%, 2000 independent samples would be obtained along with the repeats. The effective sample size was evaluated and given that $< 0.4\%$ of analyses had an effective sample size less than 2000, and for analyses in SAS software a total of 10,000 datasets were analysed, 100,000 posterior samples was considered sufficient when the proposal distribution was modified.

Chapter 7

Coverage of Credible Intervals from Various Priors

Contents

The overall aim of the thesis is to find a Bayesian analysis approach which can be used routinely in credible interval estimation in small variance component studies. This chapter assesses the suitability of various priors (previously selected for assessment in Chapter 5) for routine analysis by evaluating the coverage of credible intervals (CrIs) produced for the individual variance components and the total variance from the Bayesian analyses. The coverage for four types of CrIs (described in Section 6.2) is evaluated: highest posterior density CrIs using the posterior samples for parameter estimates on variance, SD and log variance metrics; and percentile CrIs. Six families of priors (summarised in Section 5.1.1) are evaluated. The Bayesian analyses using three of the families of priors (UNI, HCY/HT3, ICCU) were performed in WinBUGS software using Gibbs

sampling from the posterior as described in Section 6.3. The analyses for the other three families of priors (IG, JEFF, FLAT) were performed in SAS software (PROC MIXED) using an independence chain sampling algorithm, and a proposal distribution based on the inverse gamma distribution, as described in Section 6.4. The evaluation of coverage is done for the original study motivating this work and for various scenarios based on that study as previously described in Chapter 4. If during the evaluation in this chapter it is found that the coverage of the CrIs is not close to, or above the specified probability of the intervals, it will be considered that the prior and/or CrI type is not suitable for routine use.

The approach taken to evaluating the performance of priors in estimating the uncertainty of variance component estimates is described in detail in Section 7.1. The approach is based on calculating the proportion of the 95% credible intervals containing the true value for a number of scenarios (the coverage). Ideally this proportion would be 0.95, though to allow for simulation error, the estimated coverage is considered "good" if it is greater than 0.94. Firstly various priors based on the inverse gamma distribution on stratum variances were evaluated for the original batch sampling study design with 6 batches, 2 kegs per batch and 16 portions per keg and a keg variance of 0.5. A summary of the coverage achieved for the total variance and that of the individual variance components (the minimum coverage across the three σ^2_{batch} values of 0.5, 6 and 24) is given in Section 7.2.1 for several types of credible intervals. Based on the coverage of the credible intervals for the original design some of the inverse gamma priors are dropped from further evaluation. The remainder along with the priors from the other families (see Table 5.1) are then studied for a further six scenarios - four further design sizes and σ^2_{keg} values of 6 and 24 for the original batch sampling design (see Table 4.1). The results are given in Section 7.2.2. The priors and scenarios are then examined in more detail in Sections 7.3.1 and 7.3.2. In these sections the coverage is shown for all three σ^2_b values and, where the coverage is less than 0.95, the proportion of datasets with the true value lying above or below the credible interval is shown. From this it can be ascertained whether the intervals tend to be too low, too high or are just too short and which values of σ^2_b this occurs at. The results of the evaluation of coverage are summarised in Section 7.4. Firstly the choice of priors suitable for routine use is discussed in Section 7.4.1. The coverage of the hpdv and hpds credible interval types and the better priors are then compared further in Section 7.4.2. Section 7.4.3 provides other observations and conclusions drawn from the evaluation of the coverage. Tables F.1 - F.5 in Appendix F contain the coverage estimates and the proportion of datasets with the true value lying above or below the credible interval which are plotted in this chapter.

7.1 Evaluating the Performance of Priors

Given that the priors are intended to be vague, or only mildly informative, an evaluation of the priors is required to assess this. To evaluate the priors applied, the proportion of 95% credible intervals (CrIs) containing the true value was examined for each prior, CrI type, scenario and for each true σ_b^2 value investigated.

Percentile (pct) and highest posterior density (hpd) 95% credible intervals for values sampled from the posterior were obtained. Highest posterior density 95% credible intervals were obtained on the σ^2 and σ metrics, labelled hpdv and hpds respectively (see Section 6.2). In addition highest posterior density 95% credible intervals were also obtained from the logged values (labelled hpdl) for priors analysed using Gibbs sampling (note these will be the same on either the σ^2 or σ metrics). It will be seen that the log metric did not offer major advantage for priors evaluated using Gibbs sampling. Also the metric is not useful directly for interpretation unlike the variance or SD metric. Thus (except for the JEFF prior for one situation described in Section 7.2.2) this metric was not used for the other priors. The percentile intervals will be the same for any of the three metrics.

The proportion of 95% CrIs containing the true value would be 0.95 if the exact desired coverage was achieved and there was no simulation error. Given that the exact desired coverage is not likely to be possible in all cases it is considered that a desirable aim would be that the true coverage be at least 0.95. However, due to simulation error, it is not possible from the simulation results to ensure the true coverage is ≥ 0.95 without ruling out situations where the actual true coverage $= 0.95$. To have low chance of ruling out such situations and given the uncertainty in the simulations (as discussed in Section 6.1) a prior was identified as providing "good" coverage if the observed proportion was ≥ 0.94, as if the true coverage was ≥ 0.95 the estimated coverage was very unlikely to be < 0.94. It is noted that this does not ensure the true coverage is ≥ 0.95.

The chosen aim was for the true coverage to be at least 0.95 given that the nominal coverage was unlikely to be achieved in all cases (minimum coverage approach). An alternative approach would be to choose a prior on the basis of its coverage being "close" to nominal (by some measure). The former approach was chosen since typically when assessing the variability of a process or analytical method it is important that the total variance is less than some criterion and thus to achieve the nominal/stated coverage for the credible interval for the total variance is important. This is a particularly important aspect when needing to demonstrate to regulatory authorities that the process or analytical method is fit for purpose. If the aim had been for coverage close to nominal then some assurance that the coverage did not fall too far below the nominal level for individual scenarios is still likely to be required. To adopt the "closeness" approach an additional aspect is how to evaluate/predict across designs not used in the simulation

assessment. Whereas for the minimum coverage approach, provided the designs chosen provide a range of representative scenarios, it seems reasonable to assume that the coverage for other designs will not fall too far short.

The coverages of the CrIs for total variance and batch, keg and portion variance components were evaluated. The coverage associated with the total variance is considered of prime importance since in many applications the estimation of the total variance or SD is most important to the outcome of the study (e.g. for evaluating analytical or process capability) and the estimation of individual variance components is of secondary importance. As discussed in Section 5.3.2 for total variance both σ^2 and σ metrics seem relevant whilst for individual components of variation the σ^2 metric seems most relevant.

If satisfactory coverage is achieved then not only how close the coverage is to the nominal level, but also the length of the CrIs becomes important. By definition, if the metric of the parameter of interest is the same as the metric used in the estimation then the highest posterior density interval is the shortest interval (and will be the smallest credible set provided the distribution is unimodal). Thus if satisfactory coverage is obtained for the 95%CrI on the same metric as that of the parameter of interest then, irrespective of how high the coverage is above 0.95, that CrI provides the shortest interval on that metric compared to 95% CrIs generated using different metrics. Thus if it gives acceptable coverage it is to be preferred based on the criteria of length. However, if a CrI estimated on the same metric as that of the parameter of interest does not give adequate coverage, then it may be possible to use a CrI generated on another metric which does (if it exists) and then transform the CrI to the desired metric in order to obtain the desired coverage. An example of this is seen for the UNI prior for scenario b3k2p2_0.5 assuming a CrI for σ_{tot} is required. The CrI directly obtained on the SD metric has poor coverage (0.885). However the CrI for σ^2_{tot} has adequate coverage (0.960). So by taking the CrI for σ^2_{tot} and square-rooting the limits of the CrI to construct a CrI for σ_{tot} a CrI is found that has adequate coverage, but will be longer than the hpd CrI obtained directly on the SD metric.

Another aspect that sometimes could be considered important is that there is probability outside both limits of the CrI. In that case the pct or hpdl CrIs would be preferred to hpdv or hpds CrIs. However, variance components are bounded by zero in the restricted analysis and for a variance component study it may well be that a potential source of variance actually does not add variability to the response. Thus if there is high density close to 0 it was considered that 0 should not be ruled out of the CrI obtained and that the shape of the posterior would determine whether 0 was in or out of the CrI. Thus no preference was given to the pct or hpdl CrIs in the evaluation of the priors.

Initially the results were summarised by calculating the minimum coverage of the credible intervals over the three true values for σ^2_{batch} (0.5, 6, 24), for example see Table 7.1. The

minimum coverage is coloured with values above 0.97 as yellow, 0.94-0.97 as green, 0.92-0.94 as white and those below 0.92 as red. Estimated coverages greater than or equal to 0.94 were sought as discussed above. However those above 0.97 were coloured differently as they are unlikely to occur if the true coverage is ≤ 0.96 and thus will result in CrIs too wide for the coverage probability aimed for. It should be remembered that the minimum coverage across the true batch variances is used to initially summarise the results and thus a situation where coverage is only low at one true batch variance value cannot be distinguished from one where it is low across all of those investigated in this initial summary. However, this is aligned with a frequentist approach where an exact 95% confidence interval implies coverage greater than or equal to 0.95 for all parameter values.

In approaching the issue of estimating the uncertainty of variance estimates in analytical and process variability studies neither the philosophy of a frequentist or Bayesian approach was considered to be preferred upfront. A Bayesian approach was employed to overcome the issue that a frequentist approach to providing confidence intervals for variance components and other quantities of interest would generally require some form of asymptotic approximation which would be unlikely to hold generally in this application where sample sizes are very small as discussed in Chapter 3. As well as not requiring asymptotic assumptions, an advantage of the Bayesian approach is that it allows easy estimation of quantities of interest in addition to the individual variance components e.g. sums of variance components, tolerance intervals. An additional benefit of the Bayesian approach is that the meaning of a 95% credible interval, that there is a 95% probability that the parameter lies within the interval, is likely to be more intuitive and meaningful to the user of the statistic than the frequentist construction of a confidence interval that under repeated sampling the confidence interval will contain the true value 95% of the time. A further advantage of a Bayesian approach is that it allows the potential for informative prior knowledge to be incorporated. However, a corresponding disadvantage is that a Bayesian approach requires the use of a prior, the choice of which may be considered subjective, and especially for small studies, the use of different priors may provide practically different answers. Without assessment (calibration), it will not be known how often the true parameter lies within the constructed 95% credible interval when the Bayesian methodology/analysis is used routinely. In the pharmaceutical industry these studies are often performed in an environment under scrutiny by regulatory authorities. Historically frequentist approaches have been used and the regulatory authorities are likely to be concerned with a move to a methodology seen as subjective.

Given the above disadvantages, the coverage of the CrIs obtained is to be evaluated. Whilst examining the proportion of CrIs containing the true value might be thought to be more in tune with a frequentist approach rather than a Bayesian approach it was considered appropriate as:

- If under repeated simulations the 95% credible intervals contain the parameter a substantially different proportion of the time than 95%, it is questionable whether the methodology is suitable.
- The credible intervals are only to be examined for a range of true values where the quality of estimation was of most interest, unlike a frequentist approach which would desire the correct coverage for all true values. Given the bottom level variance (between portion variance) is 6, the higher level variance components are considered likely to be within the range 0.5 to 24 corresponding to $\frac{1}{12}$ and 4 times the portion variance as discussed in Section 4.1.
- Without consideration of coverage, one might choose priors on the basis that the "width" of the intervals is attractively small, even though that decrease could just be a consequence of low coverage.

Sweeting (2008) states that "Prior distributions that match posterior predictive probabilities with the corresponding frequentist probabilities are attractive when a major goal of a statistical analysis is the construction of prediction regions. Such priors provide calibration of Bayesian prediction or may be viewed as a Bayesian mechanism for producing frequentist prediction intervals."

Once the coverage of the 95% CrIs is estimated, and (ignoring simulation error) unless the coverage is equal to 0.95 for all the situations analysed, a choice is to be made in evaluating which priors provide "better" coverage. There may be different perspectives on the utility/loss function of the actual coverage differing from the desired coverage of 0.95 for different scientific situations and by different people. However, the approach taken in this thesis is that the 95% CrIs should be consistent with a true coverage of at least 0.95 (subject to simulation error). This aligns with the frequentist framework that exact confidence intervals should have a coverage of at least $1 - \alpha$, though in that case it should apply across all parameter values whereas here it is restricted to parameter values of interest. It also enables a user of the statistic to be reasonably confident of the minimum coverage across the scenarios and parameter values of interest which is likely to be preferred by regulatory authorities. However, it is accepted that this may result in CrIs which are conservative. Alternative approaches may use utility/loss functions which minimise differences (on some distance metric e.g. mean squared error) or evaluate average coverage (Bayarri and Berger (2004) discuss local average and average coverage - as well as comparing aspects of Bayesian and frequentist analyses including the evaluation of frequentist coverage for Bayesian procedures). In the on-line supplementary material provided by Ionan et al. (2014) they discuss that a common Bayes loss function for an interval estimate is a trade-off between width and coverage and provide a nice visualisation of interval width versus coverage. However, as will be discussed in Section 8.1, the linear width of an interval for a variance component may not be the best metric to use for the width. If a quantitative assessment is required additional

complications would be: deciding on the loss function to be used as the actual coverage decreases below the desired coverage or increases above it, since a symmetric function is unlikely to be suitable; the metric to be used for the parameters to average over between the ranges of interest (linear may not be suitable); and given that the priors are intended to be used routinely for a variety of designs, how to quantify the characteristics of those designs in order to be able to average over numeric quantities. To do this would require an even more extensive simulation study than that performed for this thesis (probably impractically large) to enable a good model of coverage and if required, interval "width" over the parameter ranges and designs.

7.2 Summary of Coverage Achieved

7.2.1 Coverage for Inverse Gamma Prior Distributions for the Original Batch Sampling Design

Firstly the coverages of the CrIs for the informative inverse gamma prior distributions described in Table 5.2 together with the other inverse gamma priors described in Table 5.1 are examined for the original batch sampling design (6 batches, 2 kegs per batch, 16 portions per keg). A summary of the results is given in Table 7.1 where the minimum coverage over the three true values for σ^2_{batch} (0.5, 6, 24) is provided for a true $\sigma^2_{keg} = 0.5$.

Of the vague inverse gamma priors (IG0, IG1, IG2, IG3) the CrIs (of all types) for IG0 had low coverage for σ^2_{tot} (< 0.9), but the CrIs for the rest had minimum coverage close to 0.95, as did the CrIs for the IIN prior. None of the mildly informative inverse gamma priors (I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, IhE6) reached coverage close to 0.95 for σ^2_{tot}, with the hpd intervals for IhE6 being closest (minimum coverage = 0.925).

The inverse gamma priors with coverage approaching 0.95 (IG1, IG2, IG3, IIN) for σ^2_{tot} also had "good" coverage for the batch, keg and portion variance components. Of the remaining priors IhE6, which had coverage for total variance closer to 0.95 than the remainder, also had the highest minimum estimated coverage over all the individual variance components of 0.919 for the hpdv CrI type and 0.928 for the hpds CrI type. Priors I2O3, I2O6, I1O3, I1O6, I1E3 and I1E6 had poor coverage for either batch or keg variance, typically both. When "good" coverage was achieved for the batch or keg variance it was typically not achieved for all CrI types. "Good" coverage was achieved for the portion variance for all priors and CrI types except IG0 where it was only achieved for the pct CrI.

From the above it is concluded that all the inverse gamma priors intended to be mildly informative (I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, IhE6) were actually too informative to achieve the desired coverage, as was prior IG0. This is further investigated in Section 7.3.1. The other vague inverse gamma priors (IG0, IG1, IG2, IG3) did achieve

TABLE 7.1: **Minimum coverage over σ_b^2 values 0.5, 6, 24 for a range of inverse gamma priors for original batch sampling design**

Prior	Crl Type	Total Variance	Batch Variance	Keg Variance	Portion Variance
IG0	hpdv	0.834	0.831	0.899	0.936
	hpds	0.837	0.845	0.911	0.939
	pct	0.888	0.889	0.966	0.945
IG1	hpdv	0.946	0.944	0.988	0.945
	hpds	0.946	0.948	0.988	0.946
	pct	0.949	0.950	0.979	0.951
IG2	hpdv	0.953	0.951	0.991	0.945
	hpds	0.952	0.954	0.990	0.945
	pct	0.950	0.952	0.978	0.950
IG3	hpdv	0.954	0.951	0.991	0.945
	hpds	0.953	0.954	0.990	0.947
	pct	0.950	0.953	0.978	0.950
IIN	hpdv	0.953	0.951	0.991	0.953
	hpds	0.953	0.954	0.990	0.952
	pct	0.944	0.954	0.978	0.954
I2O3	hpdv	0.023	0.693	0.203	0.953
	hpds	0.013	0.735	0.006	0.954
	pct	0.001	0.792	0.000	0.952
I2O6	hpdv	0.000	0.875	0.000	0.953
	hpds	0.000	0.265	0.000	0.953
	pct	0.000	0.015	0.000	0.953
I1O3	hpdv	0.276	0.918	0.701	0.954
	hpds	0.210	0.939	0.377	0.952
	pct	0.043	0.958	0.108	0.953
I1O6	hpdv	0.000	0.977	0.000	0.955
	hpds	0.000	0.976	0.000	0.954
	pct	0.000	0.588	0.000	0.953
I1E3	hpdv	0.876	0.866	0.994	0.954
	hpds	0.879	0.882	0.978	0.953
	pct	0.795	0.915	0.943	0.953
I1E6	hpdv	0.690	0.894	0.952	0.954
	hpds	0.640	0.913	0.850	0.953
	pct	0.337	0.937	0.708	0.952
IhE6	hpdv	0.925	0.919	0.995	0.953
	hpds	0.925	0.928	0.983	0.953
	pct	0.845	0.945	0.954	0.953

Shading:	<0.92	0.92-0.94	0.94-0.97	>0.97

the desired coverage and thus were chosen to be evaluated in further scenarios in Section 7.2.2. If further investigation of mildly informative priors was desired then it is recommended to investigate inverse gamma priors based on the median but with a smaller shape parameter (α) than 0.5 since the mildly informative priors based on the median generally performed better than those based on the mode, and decreasing α_k and α_b from 1 to 0.5 improved the coverage.

7.2.2 Coverage for Various Prior Distributions for a Range of Scenarios

Total Variance The coverages for the priors investigated (listed in Table 5.1) for a range of scenarios are summarised in Table 7.2 for total variance and for SD for the various CrI types.

Priors were examined to see which achieved "good" coverage for at least one CrI type for total variance.

- Priors HCY, HT3, IG1, IG2, IG3, IIN and JEFF had estimated coverage ≥ 0.94 across all seven scenarios investigated for at least the CrI types hpdv and hpds.
- The UNI prior did not quite achieve "good" coverage for the scenarios k2p16_0.5 and k6p16_0.5 (estimated coverage in the range 0.92-0.94) for any interval type.
- The ICCU prior did not achieve coverage ≥ 0.94 for four scenarios for any interval type, though it is noted that the hpd intervals did achieve coverage > 0.9 for all scenarios.
- For the FLAT prior when the number of batches is only 3 (b3k2s2_0.5) the posterior is improper. The coverage obtained was < 0.04 and examination of example analyses showed chains of posterior samples with an accepted proposal value retained for many successive samples. For the other scenarios (where the posterior is proper) the hpdv CrI had "good" coverage.

Next the types of CrIs were assessed to see which gave "good" coverage.

- For the hpdv and hpds CrIs
  - For priors HCY, HT3, IG1, IG2, IG3, INN and JEFF both hpdv and hpds CrIs gave "good" coverage across all scenarios.
  - For the FLAT prior, excepting scenario b3k2p2_0.5 which has an improper posterior, the hpdv CrIs gave "good" coverage across all scenarios. In addition to scenario b3k2p2_0.5 the hpds CrIs gave poor coverage (0.895) for the next smallest design, scenario k2p2_0.5.
  - The UNI prior did not achieve "good" coverage but the hpdv CrIs had higher coverage than the hpds CrIs and achieved "good" coverage for more scenarios. For the smallest design the hpds interval had much poorer coverage (0.885) than the hpdv CrI (0.960).

TABLE 7.2: **Minimum coverage over σ_b^2 values 0.5, 6, 24 for a range of scenarios and CrI types for σ_{tot}^2**

Scenario		b3k2p2_0.5	k2p2_0.5	k2p16_0.5	k2p16_6	k2p16_24	k6p16_0.5	k6p2_0.5
No. of Batches		3	6	6	6	6	6	6
No. of kegs per batch		2	2	2	2	2	6	6
No. of portions per keg		2	2	16	16	16	16	2
Keg Var		0.5	0.5	0.5	6	24	0.5	0.5
Prior	CrI Type							
UNI	hpdv	0.960	0.964	0.932	0.969	0.965	0.939	0.943
	hpds	0.885	0.955	0.929	0.947	0.950	0.937	0.939
	hpdl	0.771	0.913	0.929	0.917	0.926	0.936	0.929
	pct	0.792	0.887	0.924	0.916	0.926	0.930	0.930
HCY	hpdv	0.993	0.973	0.965	0.986	0.986	0.971	0.973
	hpds	0.969	0.968	0.959	0.980	0.979	0.971	0.969
	hpdl	0.897	0.940	0.951	0.959	0.959	0.961	0.960
	pct	0.806	0.894	0.926	0.925	0.938	0.932	0.935
HT3	hpdv	0.988	0.977	0.966	0.989	0.986	0.972	0.974
	hpds	0.963	0.967	0.961	0.979	0.978	0.970	0.970
	hpdl	0.882	0.939	0.950	0.959	0.957	0.962	0.958
	pct	0.801	0.890	0.925	0.930	0.934	0.931	0.933
ICCU	hpdv	0.956	0.917	0.907	0.927	0.937	0.920	0.913
	hpds	0.956	0.919	0.907	0.930	0.940	0.922	0.916
	hpdl	0.926	0.931	0.907	0.933	0.947	0.923	0.921
	pct	0.887	0.915	0.909	0.946	0.962	0.923	0.937
IG1	hpdv	0.994	0.957	0.946	0.960	0.980	0.945	0.947
	hpds	0.992	0.956	0.946	0.958	0.977	0.943	0.946
	pct	0.917	0.941	0.949	0.951	0.953	0.950	0.946
IG2	hpdv	0.995	0.964	0.953	0.966	0.983	0.951	0.955
	hpds	0.989	0.962	0.952	0.964	0.978	0.950	0.952
	pct	0.894	0.930	0.950	0.945	0.946	0.953	0.947
IG3	hpdv	0.996	0.964	0.954	0.967	0.984	0.953	0.954
	hpds	0.990	0.961	0.953	0.965	0.979	0.952	0.952
	pct	0.892	0.929	0.950	0.945	0.947	0.952	0.947
IIN	hpdv	0.995	0.967	0.953	0.967	0.983	0.953	0.956
	hpds	0.984	0.964	0.953	0.965	0.979	0.952	0.954
	pct	0.820	0.896	0.944	0.941	0.946	0.950	0.935
JEFF	hpdv	0.996	0.965	0.954	0.967	0.984	0.952	0.955
	hpds	0.989	0.963	0.953	0.965	0.978	0.951	0.953
	pct	0.890	0.930	0.950	0.944	0.946	0.952	0.947
FLAT	hpdv	0.038	0.943	0.978	0.974	0.973	0.988	0.977
	hpds	0.038	0.895	0.967	0.950	0.941	0.982	0.962
	pct	0.009	0.627	0.804	0.734	0.732	0.873	0.814

Shading: <0.92 | 0.92-0.94 | 0.94-0.97 | >0.97

  - For the ICCU prior there was little difference in the coverage of the hpdv and hpds CrIs - none giving "good" coverage across all scenarios.

- The hpdl CrIs were only evaluated for priors UNI, HCY, HT3 and ICCU across all scenarios. There were no priors for which the hpdl CrIs gave "good" coverage across all scenarios. The hpdl CrIs had poorer coverage for the smallest design (b3k2p2_0.5) compared to the hpdv and hpds intervals and poorer coverage for the next smallest design k2p2_0.5 for the HT3 and UNI priors.
- There were no priors for which the pct CrIs gave "good" coverage across all scenarios.

Individual Variance Components For the priors shown in Table 7.2 the coverages for the individual variance components are summarised in Table 7.3 for the various CrI types.

Priors were examined to see which achieved "good" coverage for at least one CrI type for the individual variance components.

- Only priors HCY and HT3 had estimated coverage ≥ 0.94 across all seven scenarios investigated for at least one CrI type. However, this was for the hpdl CrI type which had not achieved coverage ≥ 0.94 for total variance. For the hpdv and hpds CrI types, "good" coverage was achieved except for scenario k6p16_0.5 where coverage was in the range 0.92-0.93 for the keg variance, and scenario k2p16_24 where coverage=0.939 for the hpdv CrIs for the keg variance. The priors HCY and HT3 had very similar coverages.
- Priors IG1, IG2, IG3, INN and JEFF had poor coverage (mostly < 0.9) for the keg variance for the scenarios with true σ_k^2 of 6 and 24 (k2p16_6 and k2p16_24). It is also interesting to note that, whilst the hpdv CrIs for σ_b^2 achieved good (or actually very high) coverage for the scenario k2p16_24, the coverage of the hpds CrIs was poor. Except for the IIN prior, the priors also had poor coverage (mostly < 0.9) for the portion variances for the three scenarios with the smallest number of portions per keg (b3k2p2_0.5, k2p2_0.5 and k6p2_0.5). The IIN prior had "good" coverage for the portion variance but this is expected as it was specifically chosen to be informative for σ_p^2 (aligned with its true value of 6). Given that the hpdl CrIs gave "good" coverage across the scenarios for the HCY and HT3 priors, analyses were performed for one situation (scenario k2p16_24, σ_b^2=0.5) where the coverage was poorest for the hpdv, hpds and pct CrIs to see whether the hpdl CrIs could give "good" coverage. As the inverse gamma and JEFF priors gave very similar coverages this was only performed for the JEFF prior. The coverage of the hpdl CrIs for the keg variance was 0.894 and thus was also poor. For the batch variance the coverage of the hpdl CrIs was very poor (0.513). Given that the coverage results for the JEFF prior were very similar to those of the inverse gamma priors

TABLE 7.3: **Minimum coverage over σ_b^2=0.5, 6, 24 for σ_b^2, σ_k^2, σ_p^2 for evaluated scenarios**

		Batch Variance / SD							Keg Variance / SD							Portion Variance / SD						
No. of Batches		3	6	6	6	6	6	6	3	6	6	6	6	6	6	3	6	6	6	6	6	6
No. of kegs per batch		2	2	2	2	2	6	6	2	2	2	2	2	6	6	2	2	2	2	2	6	6
No. of portions per keg		2	2	16	16	16	16	2	2	2	16	16	16	16	2	2	2	16	16	16	16	2
Keg Var		0.5	0.5	0.5	6	24	0.5	0.5	0.5	0.5	0.5	6	24	0.5	0.5	0.5	0.5	0.5	6	24	0.5	0.5
Prior	CI Type																					
UNI	hpdv	0.997	0.944	0.921	0.956	0.985	0.934	0.938	0.999	0.998	0.994	0.959	0.930	0.928	0.994	0.961	0.940	0.947	0.944	0.952	0.947	0.953
	hpds	0.996	0.944	0.928	0.956	0.950	0.936	0.940	0.978	0.986	0.989	0.954	0.930	0.922	0.984	0.958	0.942	0.946	0.943	0.950	0.944	0.951
	hpdl	0.994	0.979	0.940	0.961	0.922	0.933	0.930	0.943	0.962	0.966	0.941	0.936	0.959	0.978	0.958	0.953	0.947	0.944	0.946	0.945	0.953
	pct	0.999	0.979	0.956	0.984	0.976	0.947	0.948	0.979	0.978	0.967	0.941	0.966	0.933	0.988	0.957	0.961	0.945	0.943	0.946	0.946	0.955
HCY	hpdv	1.000	0.956	0.960	0.965	0.989	0.969	0.971	1.000	0.998	0.994	0.956	0.939	0.929	0.994	0.961	0.940	0.946	0.944	0.951	0.949	0.951
	hpds	1.000	0.953	0.957	0.961	0.987	0.971	0.966	0.995	0.987	0.988	0.960	0.947	0.922	0.985	0.958	0.943	0.945	0.942	0.950	0.945	0.951
	hpdl	0.998	0.987	0.972	0.974	0.947	0.968	0.973	0.967	0.969	0.967	0.962	0.959	0.958	0.978	0.958	0.954	0.948	0.942	0.946	0.944	0.956
	pct	0.996	0.964	0.954	0.975	0.978	0.952	0.950	0.980	0.980	0.968	0.946	0.960	0.933	0.989	0.957	0.961	0.947	0.943	0.945	0.945	0.957
HT3	hpdv	1.000	0.959	0.963	0.968	0.991	0.971	0.972	1.000	0.998	0.994	0.957	0.944	0.929	0.994	0.962	0.942	0.948	0.944	0.952	0.948	0.951
	hpds	1.000	0.955	0.960	0.964	0.989	0.969	0.966	0.995	0.987	0.989	0.960	0.949	0.922	0.985	0.961	0.943	0.948	0.943	0.948	0.946	0.952
	hpdl	0.997	0.987	0.973	0.974	0.945	0.966	0.969	0.964	0.967	0.966	0.960	0.962	0.958	0.978	0.958	0.953	0.947	0.943	0.945	0.945	0.955
	pct	0.997	0.968	0.955	0.976	0.975	0.950	0.949	0.981	0.979	0.968	0.944	0.959	0.933	0.989	0.957	0.960	0.946	0.942	0.945	0.945	0.955
ICCU	hpdv	0.892	0.867	0.896	0.875	0.878	0.918	0.905	0.999	0.998	0.996	0.939	0.874	0.956	0.990	0.911	0.885	0.948	0.942	0.952	0.945	0.915
	hpds	0.897	0.885	0.913	0.891	0.887	0.924	0.912	0.994	0.984	0.982	0.953	0.887	0.952	0.965	0.918	0.893	0.947	0.942	0.948	0.943	0.920
	hpdl	0.977	0.953	0.944	0.940	0.902	0.936	0.936	0.928	0.932	0.949	0.967	0.908	0.952	0.942	0.932	0.908	0.947	0.943	0.944	0.944	0.932
	pct	0.967	0.919	0.933	0.938	0.944	0.940	0.943	0.961	0.962	0.948	0.959	0.924	0.947	0.968	0.941	0.919	0.947	0.946	0.942	0.947	0.939
IG1	hpdv	0.994	0.942	0.944	0.950	0.990	0.944	0.944	0.996	0.995	0.988	0.855	0.835	0.941	0.978	0.818	0.835	0.945	0.949	0.947	0.945	0.893
	hpds	0.994	0.945	0.948	0.951	0.914	0.944	0.944	0.977	0.966	0.988	0.872	0.848	0.947	0.949	0.828	0.848	0.946	0.948	0.947	0.945	0.904
	pct	0.978	0.959	0.950	0.915	0.676	0.951	0.948	0.928	0.940	0.979	0.910	0.894	0.949	0.950	0.883	0.892	0.951	0.949	0.949	0.946	0.923
IG2	hpdv	0.999	0.951	0.951	0.958	0.992	0.950	0.951	0.997	0.995	0.991	0.871	0.855	0.943	0.979	0.838	0.848	0.945	0.948	0.947	0.944	0.897
	hpds	0.998	0.951	0.954	0.958	0.913	0.951	0.951	0.976	0.967	0.990	0.887	0.866	0.947	0.948	0.848	0.860	0.945	0.948	0.947	0.943	0.908
	pct	0.973	0.962	0.952	0.912	0.660	0.954	0.950	0.923	0.937	0.978	0.920	0.906	0.948	0.949	0.898	0.900	0.950	0.950	0.948	0.946	0.926
IG3	hpdv	0.999	0.952	0.951	0.959	0.992	0.951	0.951	0.997	0.995	0.991	0.873	0.857	0.943	0.980	0.839	0.849	0.945	0.949	0.947	0.944	0.899
	hpds	0.998	0.953	0.954	0.959	0.913	0.952	0.951	0.977	0.967	0.990	0.890	0.866	0.949	0.949	0.849	0.861	0.947	0.948	0.945	0.944	0.909
	pct	0.971	0.962	0.953	0.911	0.660	0.954	0.950	0.922	0.937	0.978	0.921	0.908	0.949	0.950	0.900	0.900	0.950	0.949	0.949	0.947	0.926
IIN	hpdv	1.000	0.953	0.951	0.959	0.991	0.951	0.951	1.000	0.999	0.991	0.874	0.856	0.942	0.989	0.996	0.962	0.953	0.953	0.952	0.945	0.944
	hpds	1.000	0.952	0.954	0.958	0.914	0.951	0.952	0.996	0.987	0.990	0.888	0.867	0.948	0.969	0.995	0.967	0.952	0.951	0.950	0.944	0.951
	pct	0.981	0.963	0.954	0.910	0.659	0.953	0.950	0.967	0.969	0.978	0.921	0.907	0.949	0.970	0.990	0.977	0.954	0.952	0.951	0.946	0.962
JEFF	hpdv	0.999	0.952	0.952	0.959	0.991	0.951	0.951	0.997	0.994	0.991	0.873	0.857	0.944	0.979	0.836	0.849	0.943	0.950	0.947	0.945	0.898
	hpds	0.999	0.953	0.955	0.959	0.912	0.951	0.952	0.978	0.966	0.990	0.890	0.868	0.948	0.949	0.848	0.860	0.946	0.947	0.947	0.945	0.907
	pct	0.971	0.964	0.953	0.911	0.658	0.954	0.949	0.923	0.938	0.978	0.921	0.907	0.949	0.950	0.898	0.902	0.951	0.948	0.949	0.947	0.927
FLAT	hpdv	0.056	0.999	0.995	0.999	0.997	0.993	0.993	0.706	0.996	0.993	0.980	0.978	0.957	0.978	0.952	0.959	0.951	0.950	0.949	0.944	0.945
	hpds	0.056	0.995	0.990	0.985	0.951	0.985	0.985	0.677	0.957	0.974	0.971	0.975	0.942	0.946	0.937	0.960	0.950	0.949	0.947	0.944	0.949
	pct	0.000	0.951	0.914	0.818	0.263	0.894	0.903	0.292	0.877	0.908	0.916	0.935	0.942	0.946	0.895	0.961	0.951	0.949	0.948	0.946	0.957

Shading: <0.92 | 0.92-0.94 | 0.94-0.97 | >0.97

it was not considered worthwhile to perform the simulations for those priors and the hpdl CrI type was not considered useful for these priors.

- The UNI prior did not quite achieve "good" coverage for both the scenarios k2p16_24 and k6p16_0.5 (estimated coverage in the range 0.92-0.94) for the keg variance for any interval type. For the batch variance each CrI type had poor coverage (in range 0.92-0.94) for at least 2 scenarios except for the pct CrI which achieved "good" coverage.
- The ICCU prior had coverage < 0.94 for at least three scenarios for each interval type for the batch variance, the hpdv and hpds CrIs being particularly poor. It also did not achieve the desired coverage for scenario k2p16_24 for any interval type. Coverage was also generally poor for the portion variance for the designs with 2 portions/keg (b3k2p2_0.5, k2p2_0.5, k6p2_0.5).
- For the FLAT prior when the number of batches is only 3 (b3k2s2_0.5) the posterior is improper. The coverage obtained was < 0.06 for the batch variance and poor for the keg variance (0.706 or less). For the other scenarios (where the posterior is proper) the hpdv and hpds CrIs had "good" coverage.

7.3 Further Evaluation of the Coverage

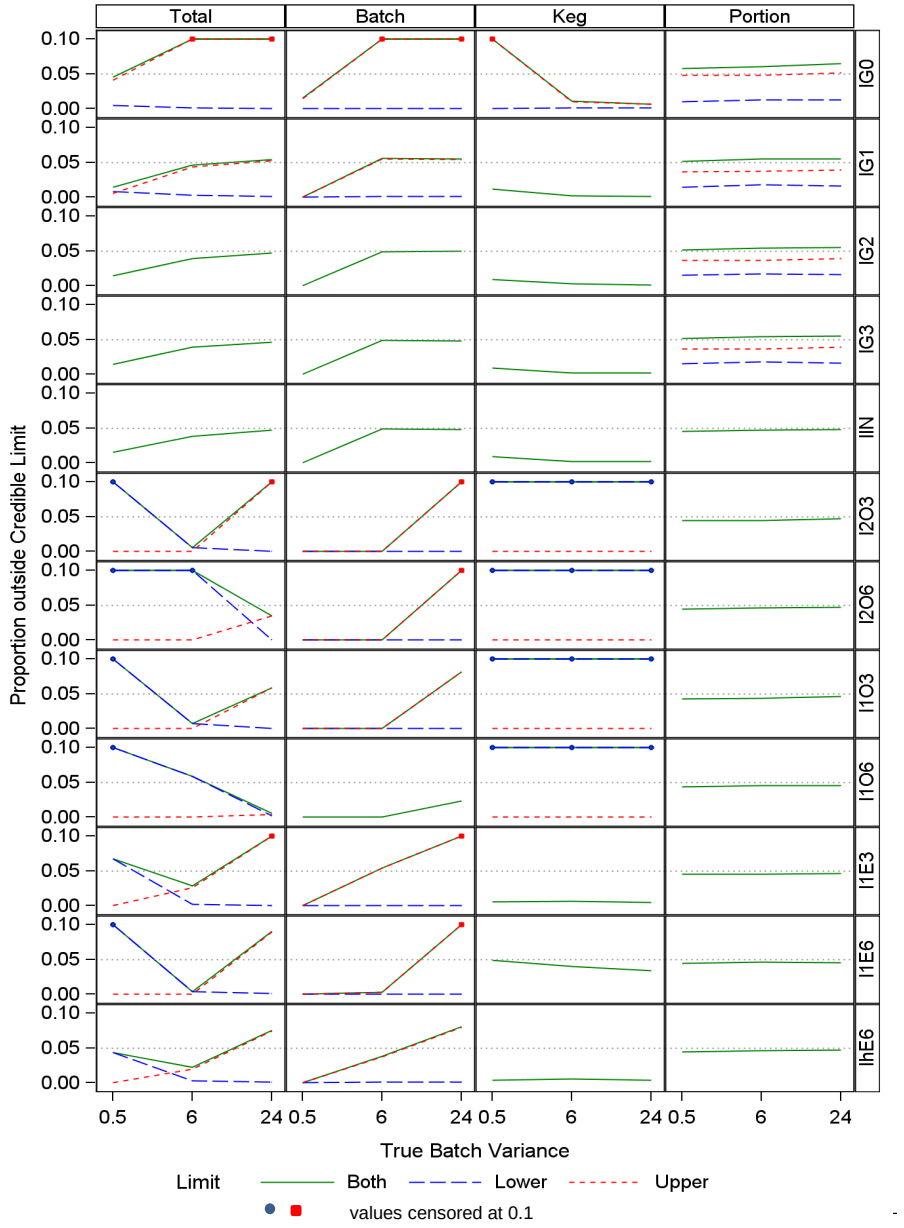
The priors are further evaluated in this section through examining the proportion lying outside the credible limits for scenarios where the coverage was low. Note that for the highest posterior density CrIs there is no requirement for them to have the same proportion of 0.025 outside each end of the interval unlike the pct CrIs. Thus for hpd CrIs where the proportion outside one limit exceeds 0.05 this indicates a problem at that end of the interval. Where the total proportion outside exceeds 0.05 but 0.05 is not exceeded for either limit the problem is less clear. The results for the inverse gamma priors evaluated for the original batch sampling design are provided in Section 7.3.1. The results for the priors evaluated over all scenarios are given in Section 7.3.2.

7.3.1 Inverse Gamma Priors Evaluated for Original Batch Sampling Design

Figure 7.1 shows the proportion of hpdv CrIs with true variance outside the limits (solid line) for σ^2_{tot} and for the variance components. A reference line is drawn at 0.05 illustrating the maximum desired proportion outside the limits (though noting that simulation error could see observed proportions up to 0.06). If the proportion outside both limits was greater than 0.05 the proportion outside the lower and upper limit are also shown. Proportions greater than 0.1 were censored for plotting purposes and are shown as filled circles (lower) or squares (upper). As discussed above, there is no

requirement for the hpd CrIs to have the same proportion of 0.025 outside each end of the interval.

FIGURE 7.1: **Proportion of hpdv CrIs where true variance is outside limits for inverse gamma priors for scenario k2p16_0.5**



None of the mildly informative inverse gamma priors (I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, IhE6) reached coverage close to 0.95 for σ^2_{tot}.

- For those based on the mode being set at the likely value (I2O3, I2O6, I103, I1O6) it was seen in Table 7.1 that the minimum coverage (over σ^2_b= 0.5, 6, 24) was extremely poor for the keg variance and similarly for the total variance for all interval types. It is seen from Figures 7.1, 7.2 and 7.3 that for the keg variance this occurs for all three values of σ^2_b and is due to the CrI for σ^2_k tending to be

FIGURE 7.2: **Proportion of hpds CrIs where true variance is outside limits for inverse gamma priors for scenario k2p16_0.5**

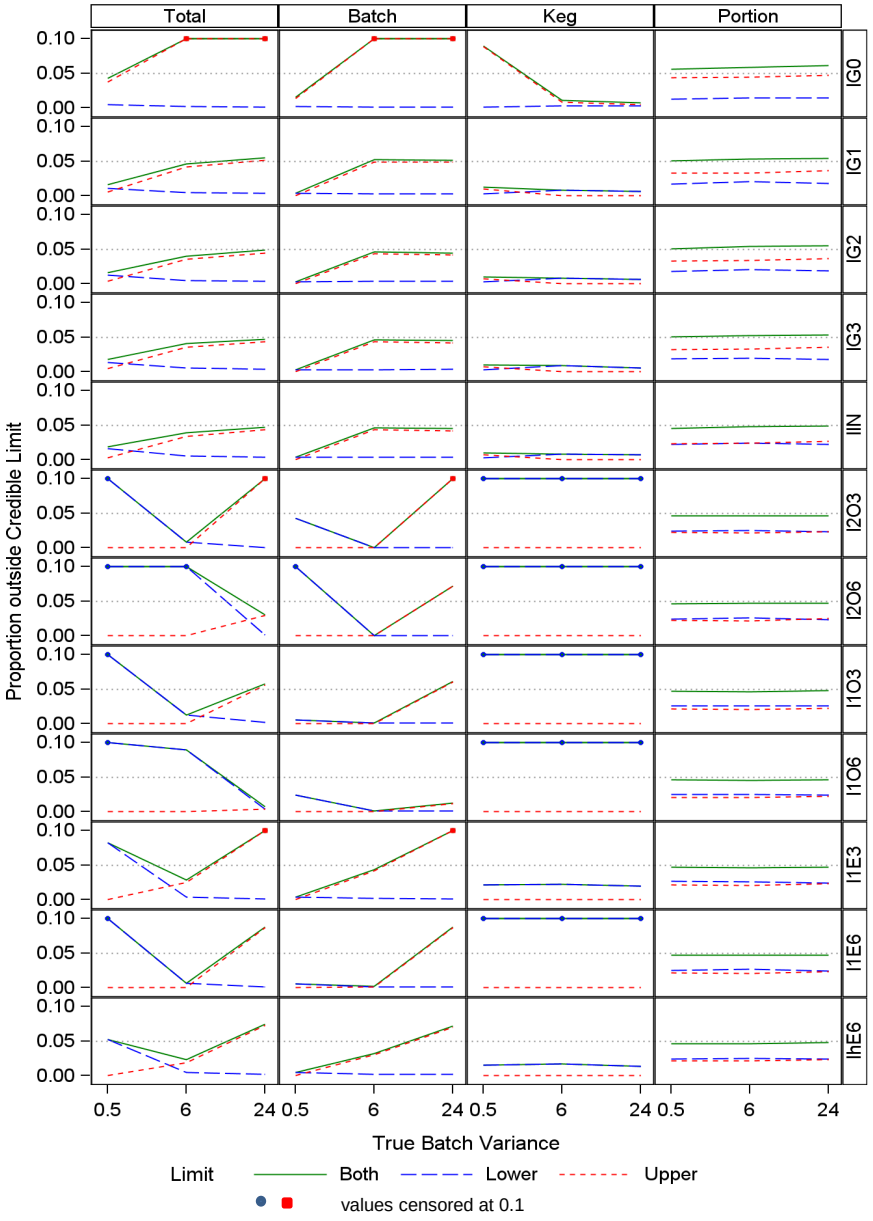
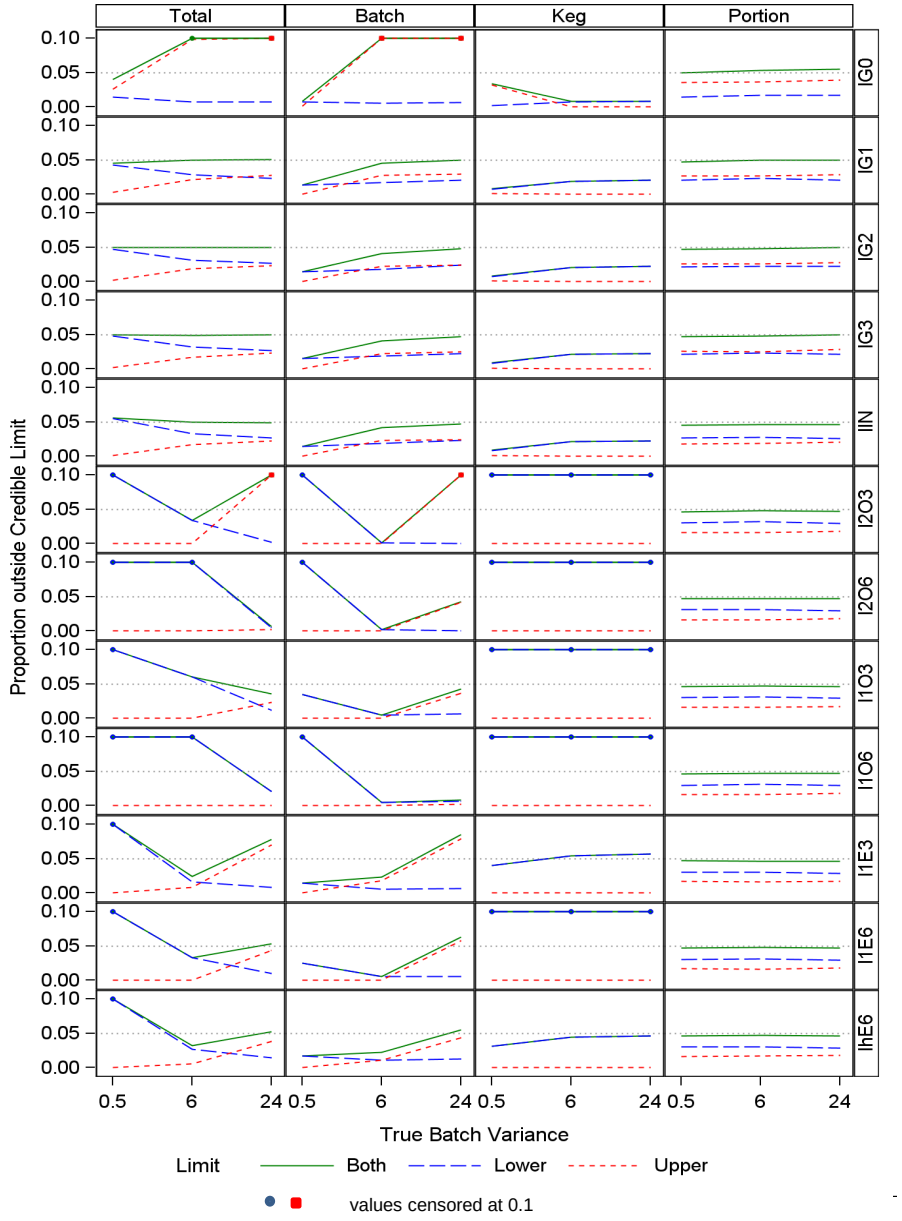


FIGURE 7.3: **Proportion of pct CrIs where true variance is outside limits for inverse gamma priors for scenario k2p16_0.5**



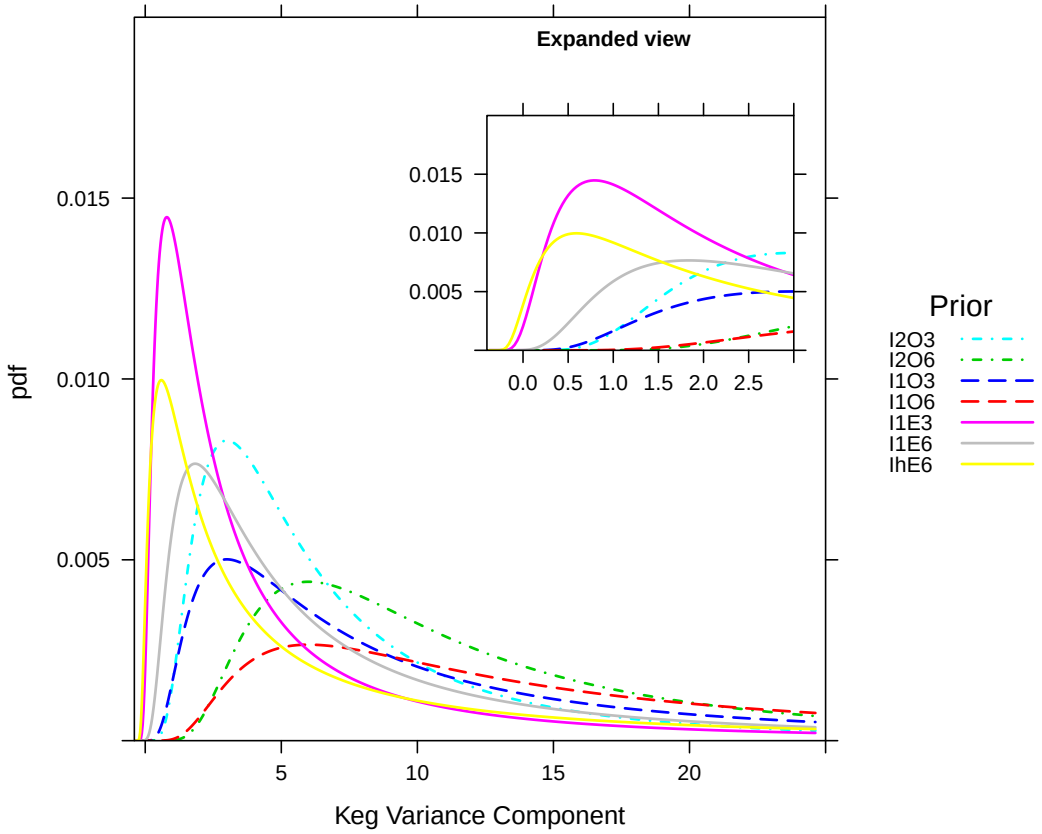
above the true value (0.5). This is then reflected in the coverage for the total variance, though as σ_b^2 increases to 24 the coverage for the total variance is then more affected by the coverage achieved for σ_b^2 where the CrI for σ_b^2 tends to be below the true value. Whilst Figure 5.11 showed non-negligible density for a true σ_k^2=0.5, this was based on the marginal distribution for σ_k^2 given a prior for σ_p^2 of IG(1,12). The effective priors on σ_k^2 are shown in Figure 7.4 when inverse gamma priors are placed on the keg stratum variance and a value of 6 is assumed for σ_p^2. It is seen from Figure 7.4 that for the priors based on the mode the density is extremely low at σ_k^2=0.5 (when σ_p^2=6).

- For the three priors based on the median (I1E3, I1E6 and IhE6) the coverage for σ_{tot}^2 did not reach the desired coverage of 0.95 for any hpd CrI type, though the minimum coverage summarised in Table 7.1 was generally better than for those priors based on the mode. For all three priors, hpd interval types and σ_b^2=24 the CrIs for the batch variance tended to be below the true value. The effect of this was also seen with the CrIs for the total variance being below the true value for σ_b^2=24. In addition the hpds and pct interval types for the I1E6 prior tended to have the CrI for the keg variance above the true value of 0.5 for all three values of σ_b^2. Again it is seen that whilst Figure 5.11 showed non-negligible density for a true σ_k^2=0.5 given a prior for σ_p^2 of IG(1,12), Figure 7.4 shows that when a value of 6 is assumed for σ_p^2 the density is low at σ_k^2=0.5 for I1E6, though not for I1E3 and IhE6. For prior I1E6 the poor coverage for the keg variance is then seen reflected in that for the total variance, where the CrIs tend to be above the true value of σ_{tot}^2 when σ_b^2=0.5. This is not seen for σ_b^2=6 or 24, presumably because the contribution of the keg variance to the total is swamped by the batch variance. Additionally when σ_b^2=0.5, the pct CrIs for the total variance tend to be above the true value for the I1E3 and IhE6 priors.

The desired coverage was not achieved for the mildly informative priors using independent inverse gamma distributions on the stratum variances. It is seen from the above that this was mainly due to the difficulties of achieving only a mildly informative resultant prior distribution on another variance component when the data provides good information in estimating one variance component. That is a problem here when a set of priors is required that is at most mildly informative. However, in an alternative situation where it is desired to make use of historical information it is also clear that the choice of parameters for an informative inverse gamma prior will be very important in order that the prior does not dominate the information from the study itself in an unintended manner.

For prior IG0 poor coverage was seen for σ_{tot}^2, with its CrI being too frequently below the true value for all interval types, for σ_b^2=6 or 24. A similar situation was seen for the coverage for σ_b^2. For σ_k^2 its hpd CrIs were too frequently below the true value for σ_b^2=0.5, though not for σ_b^2=6 or 24. This suggests that the IG0 prior is too informative

FIGURE 7.4: **Marginal distribution on σ_k^2 for mildly informative inverse gamma distributions on keg stratum variances when σ_p^2=6**



and the CrI will often underestimate σ_{tot}^2 and some of the variance components. The reasons for this can be seen in Figures 5.6, 5.7 and 5.8 where the marginal density for the batch variance (on variance or SD metric) is higher for small values than for IG1 or IG2.

Inverse gamma priors IG1, IG2, IG3 and IIN had better coverage and are discussed in Section 7.3.2.

7.3.2 Priors Evaluated for All Scenarios

Using the same approach to plotting as Figure 7.1, Figures 7.5 - 7.8 show the proportion of CrIs with true total variance outside the limits (solid line) for hpdv, hpds, hpdl and pct CrI types. Figures 7.9 - 7.20 show the proportion of CrIs with true batch, keg and portion variances outside the limits for the various CrI types.

FIGURE 7.5: **Proportion of hpdv CrIs where true σ^2_{tot} is outside limits**

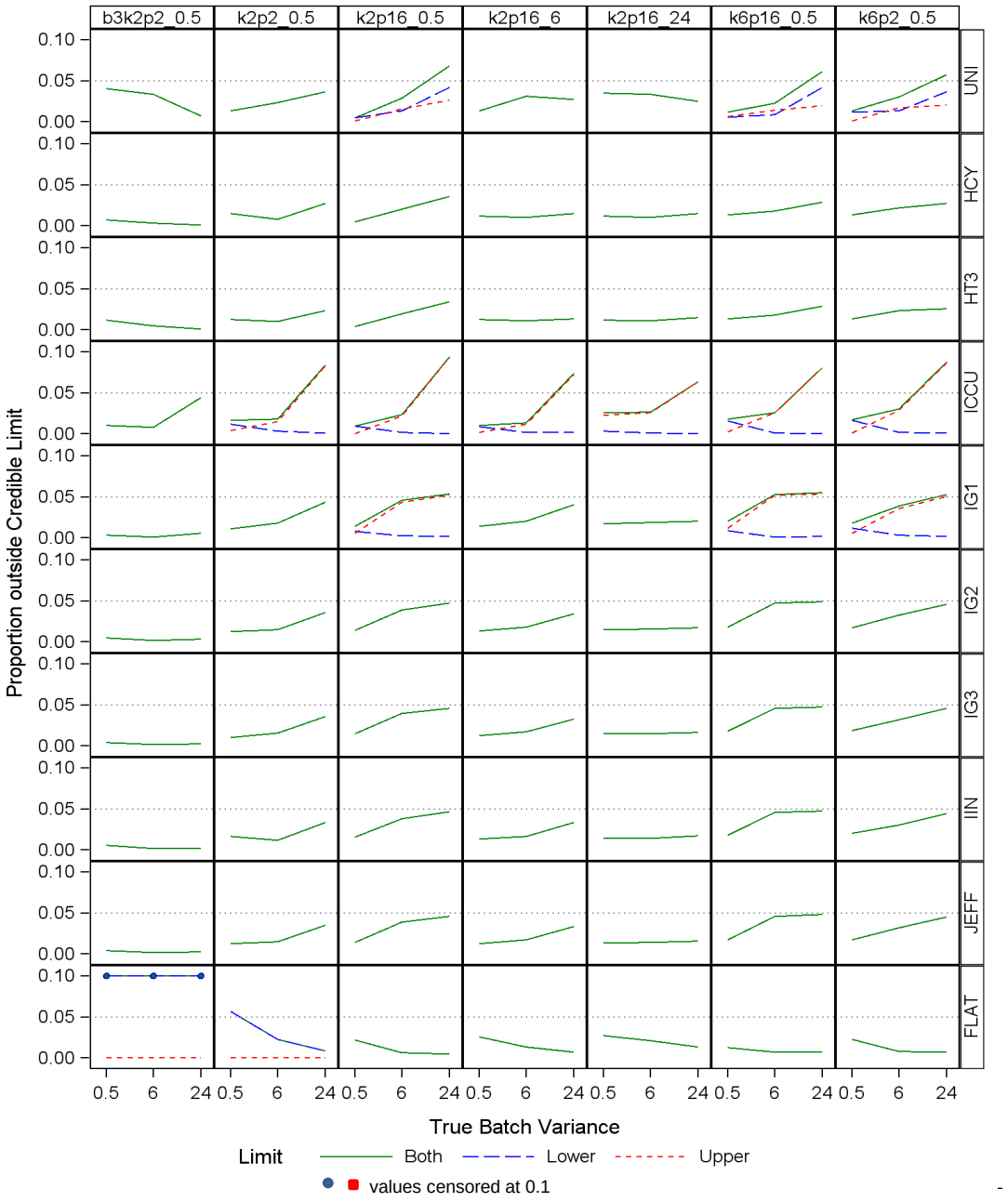


FIGURE 7.6: **Proportion of hpds CrIs where true σ^2_{tot} is outside limits**

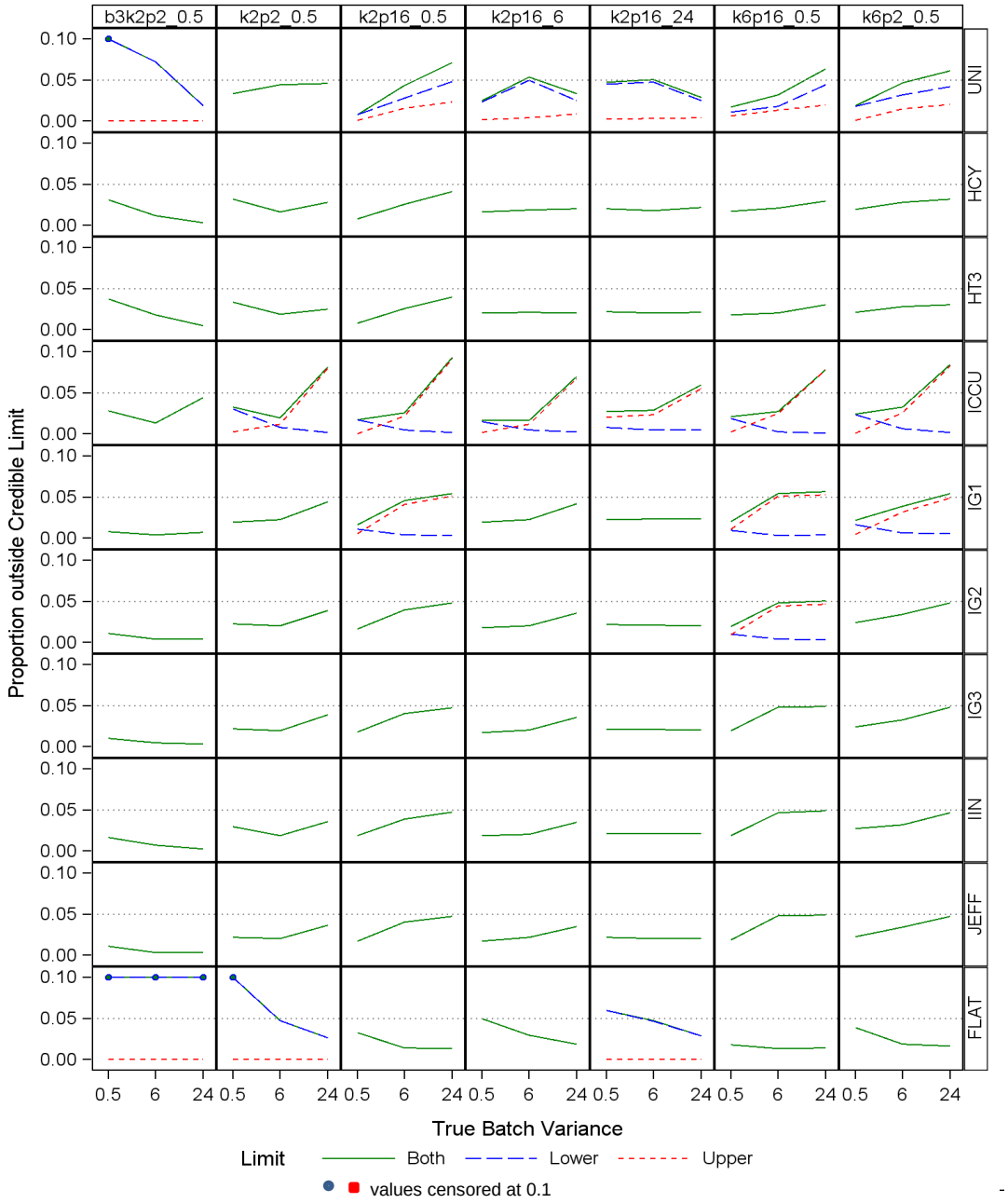


FIGURE 7.7: **Proportion of hpdl CrIs where true σ^2_{tot} is outside limits**

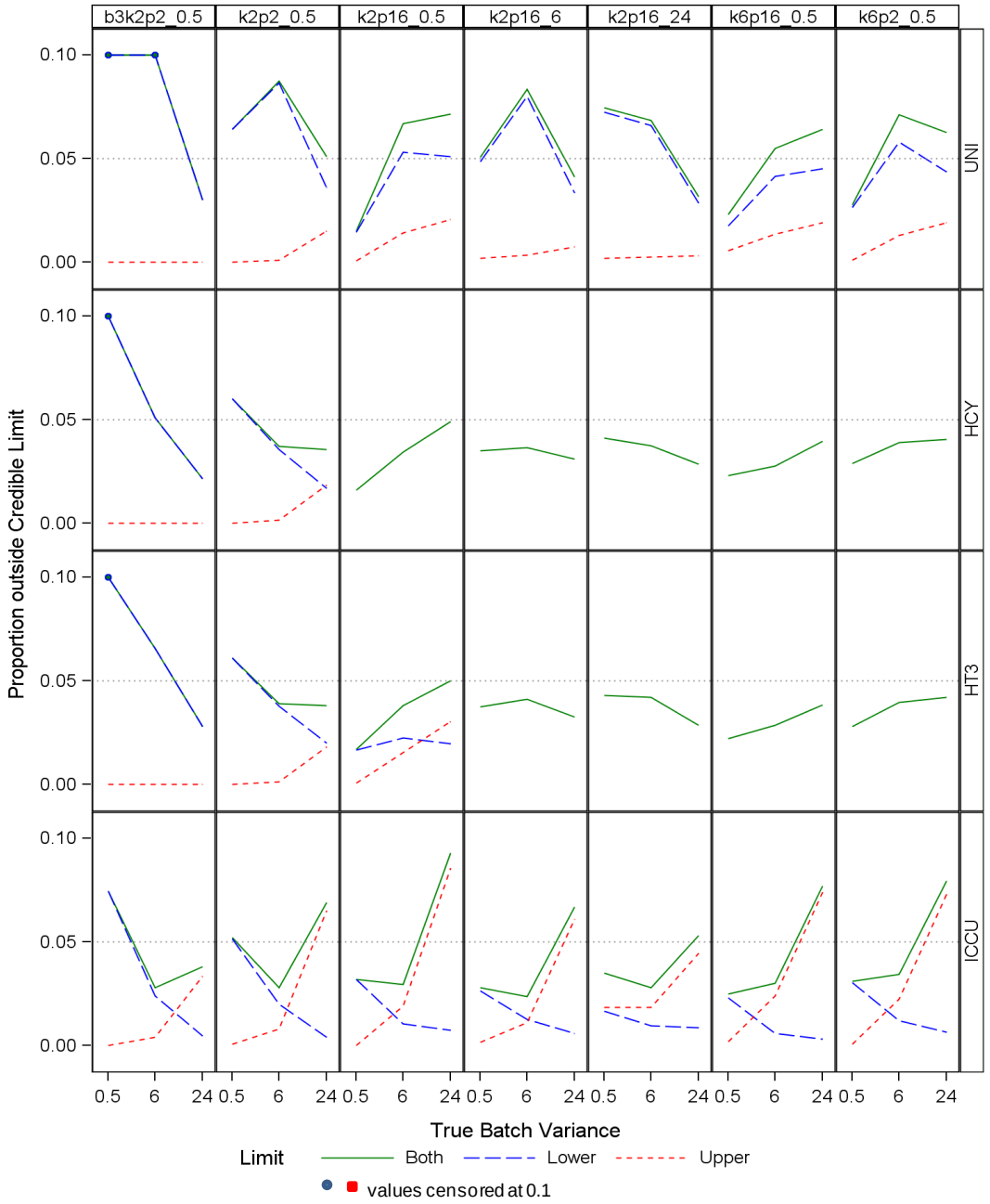


FIGURE 7.8: **Proportion of pct CrIs where true σ^2_{tot} is outside limits**

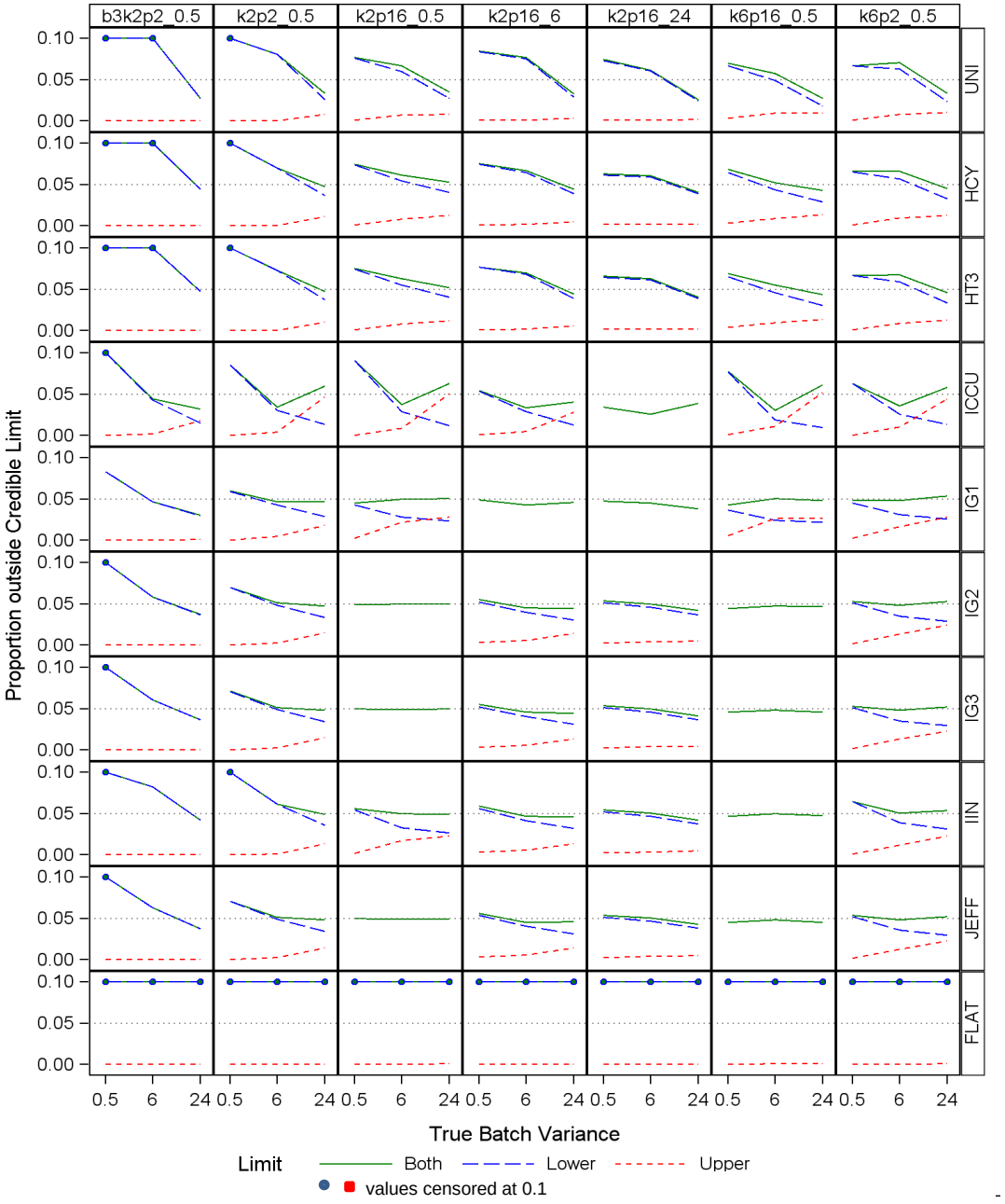


FIGURE 7.9: **Proportion of hpdv CrIs where true σ_b^2 is outside limits**

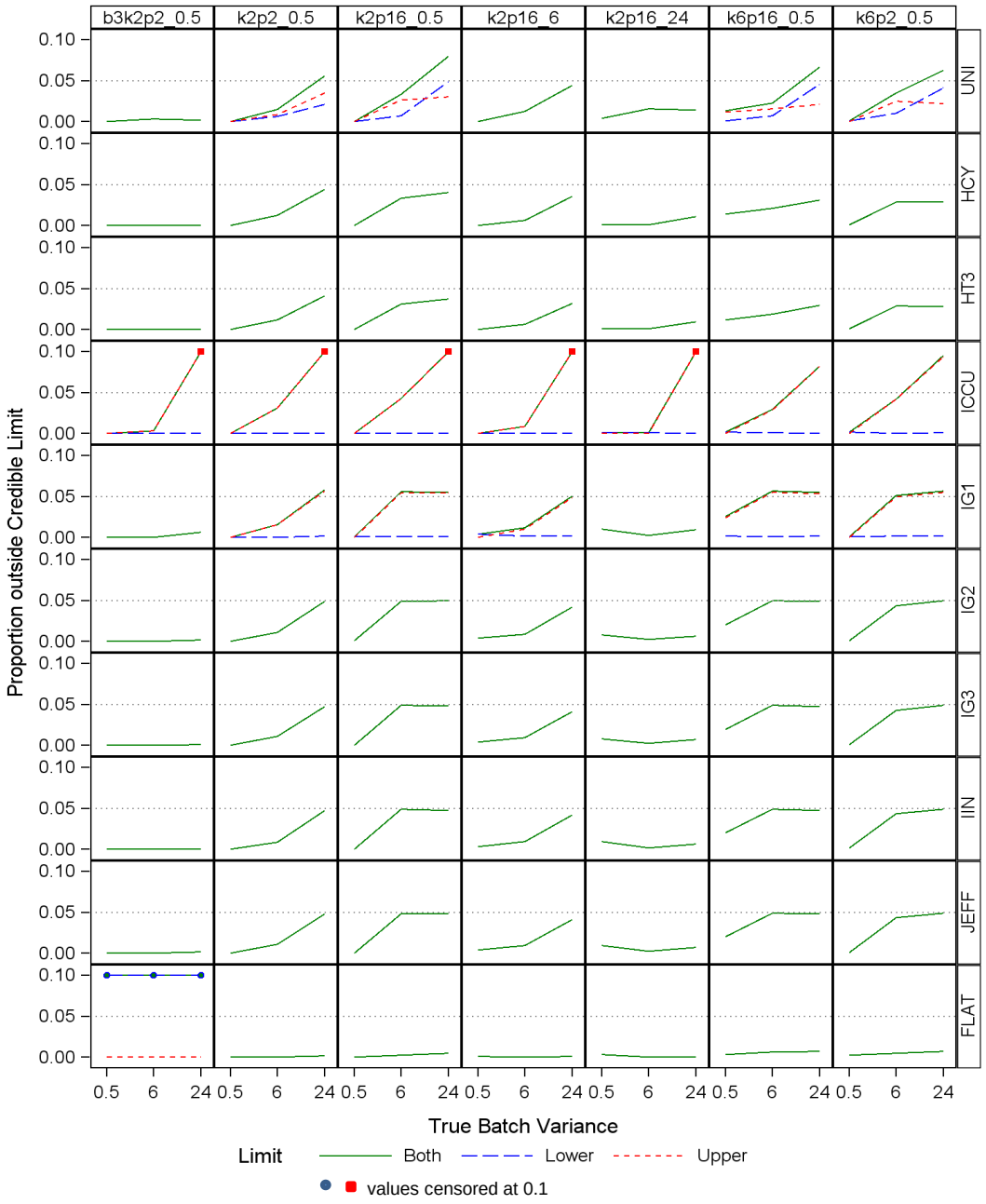


FIGURE 7.10: **Proportion of hpds CrIs where true σ_b^2 is outside limits**

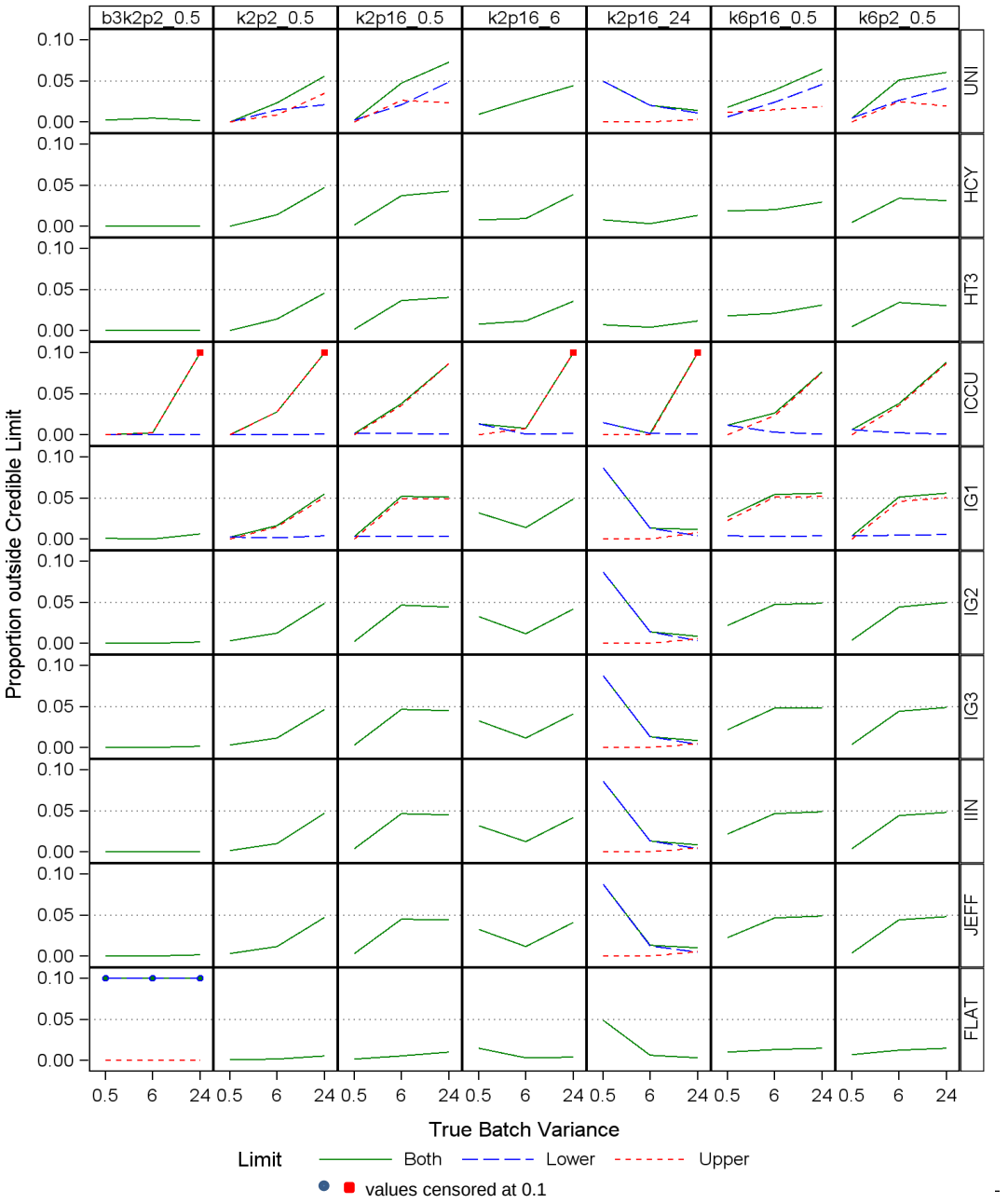


FIGURE 7.11: **Proportion of hpdl CrIs where true σ_b^2 is outside limits**

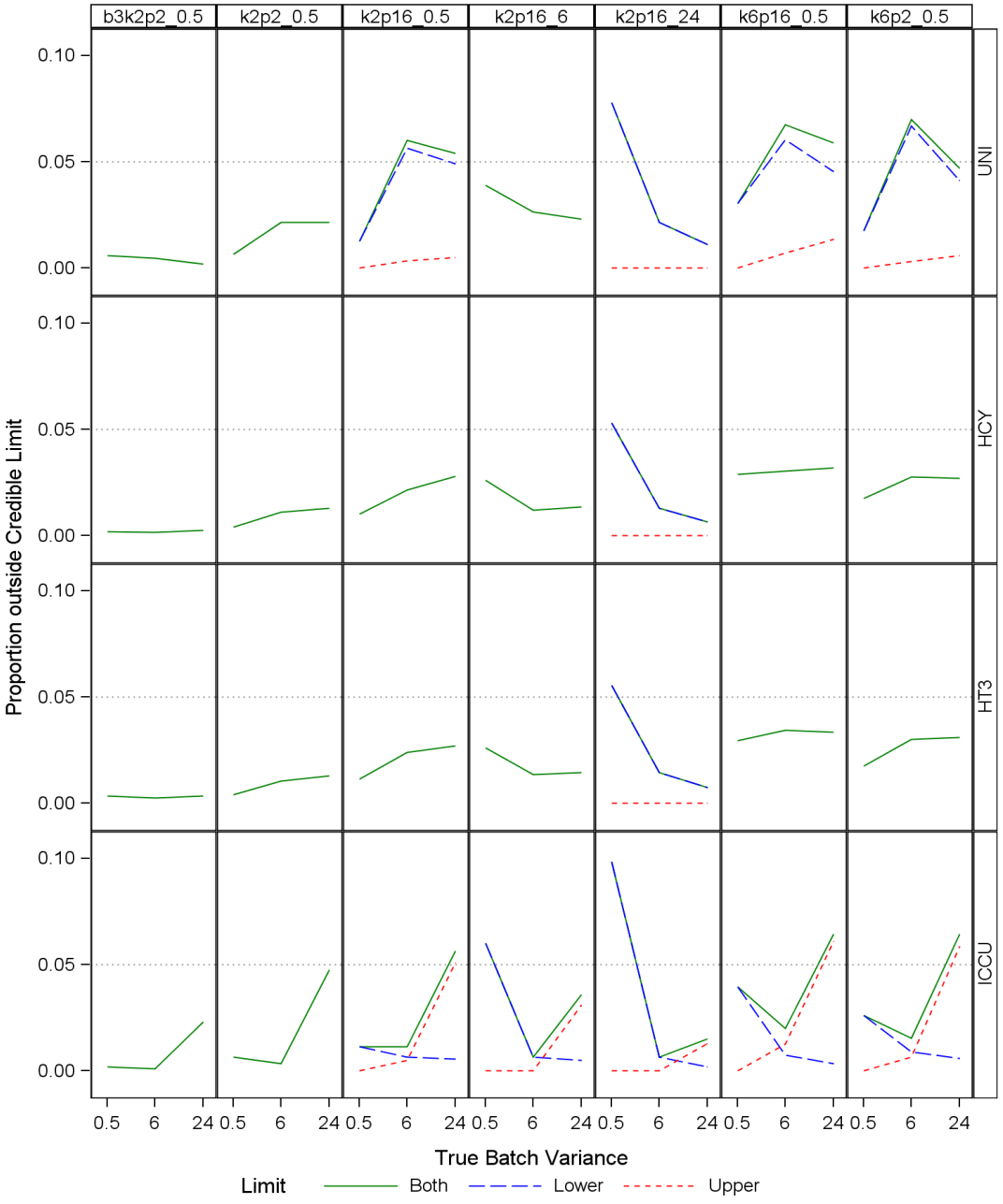


FIGURE 7.12: **Proportion of pct CrIs where true σ_b^2 is outside limits**

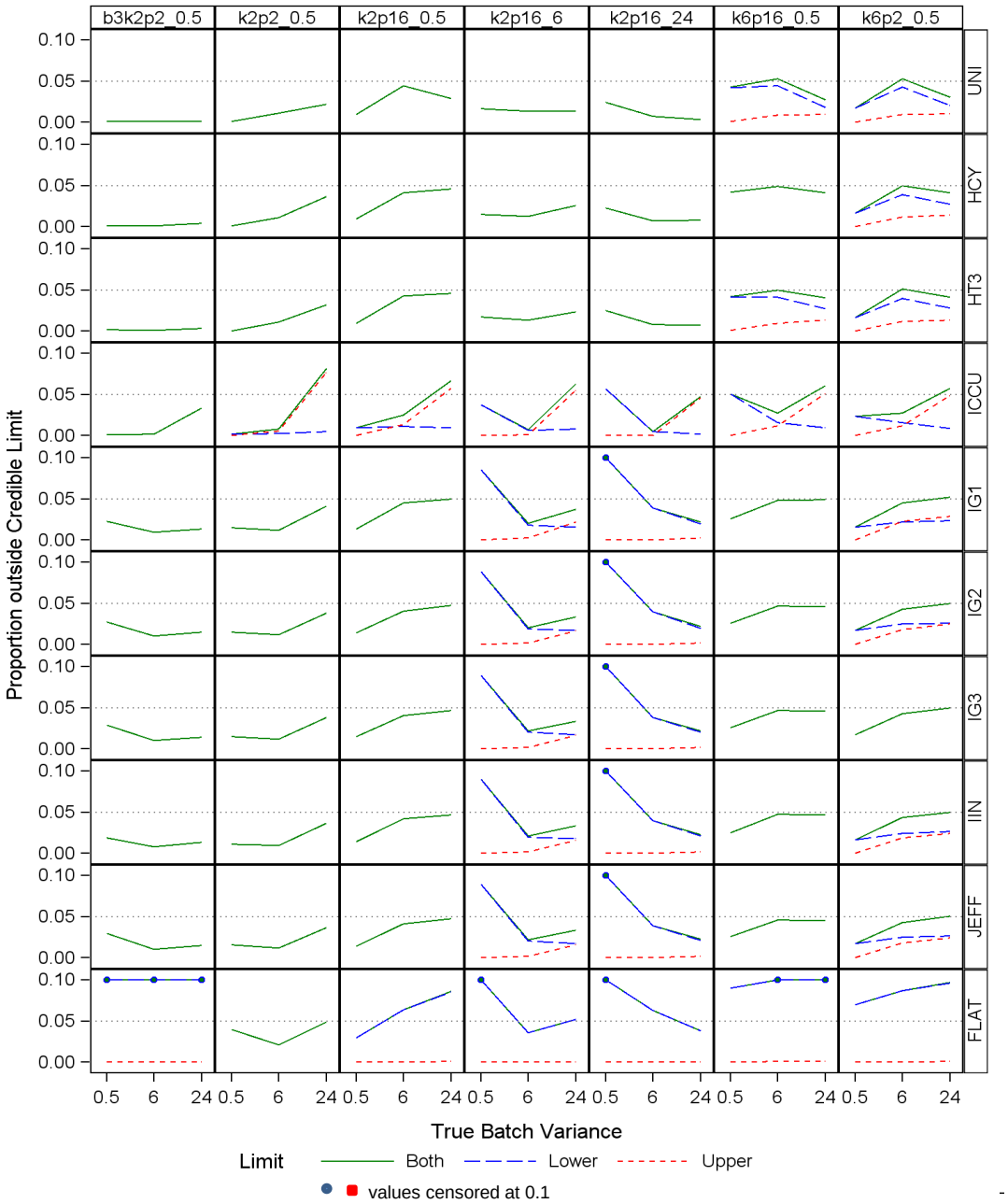


FIGURE 7.13: **Proportion of hpdv CrIs where true σ_k^2 is outside limits**

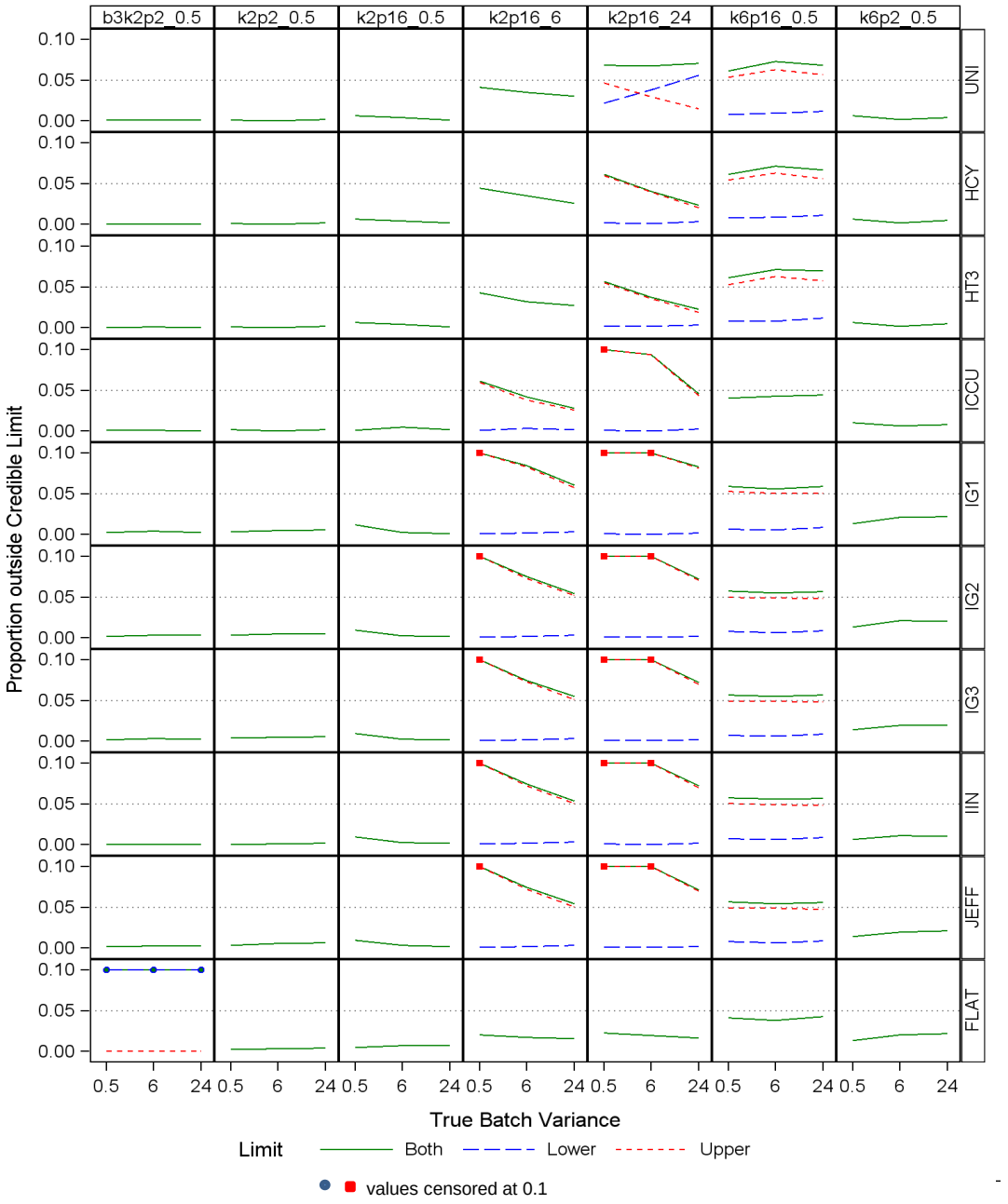


FIGURE 7.14: **Proportion of hpds CrIs where true σ_k^2 is outside limits**

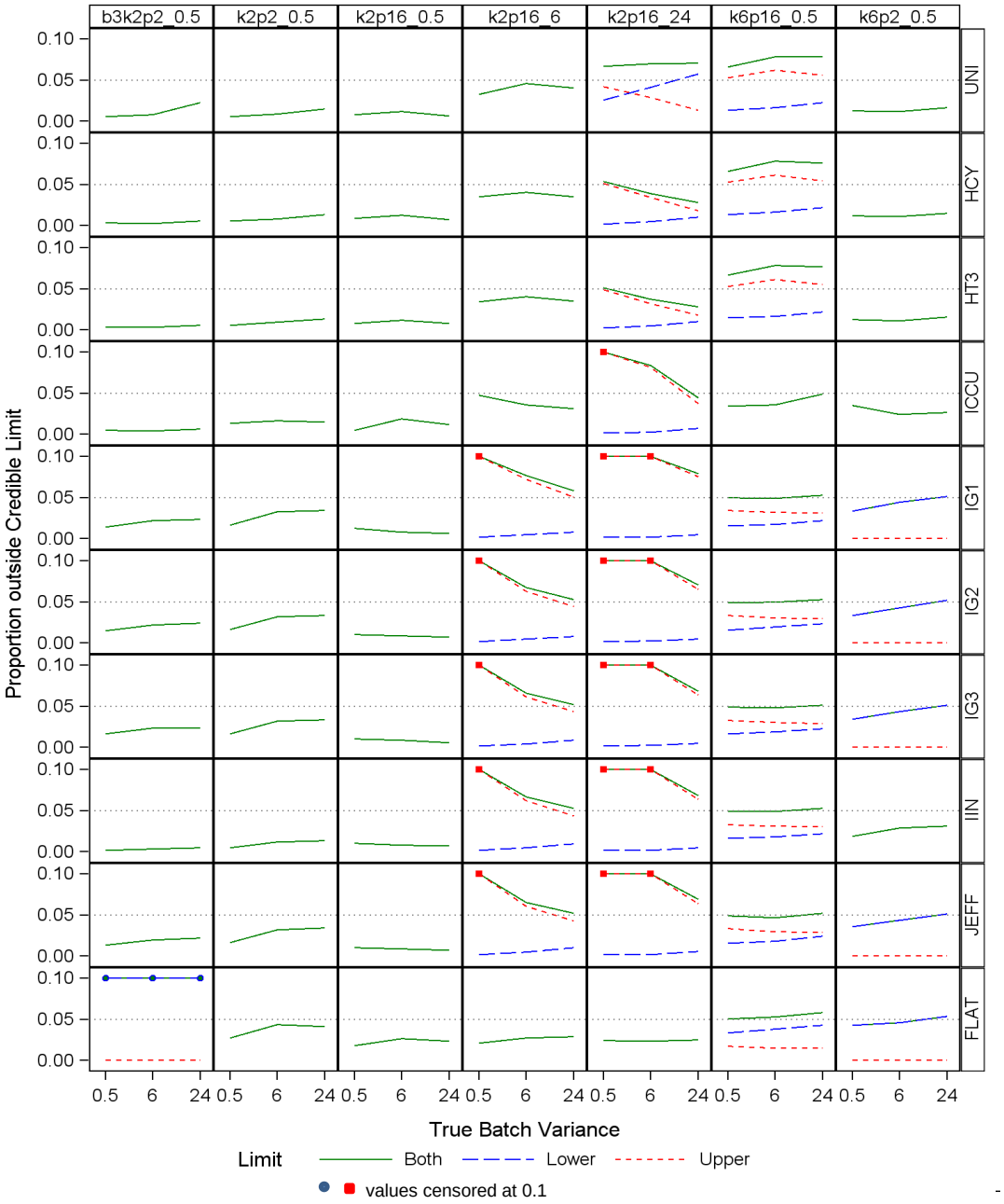


FIGURE 7.15: **Proportion of hpdl CrIs where true σ_k^2 is outside limits**

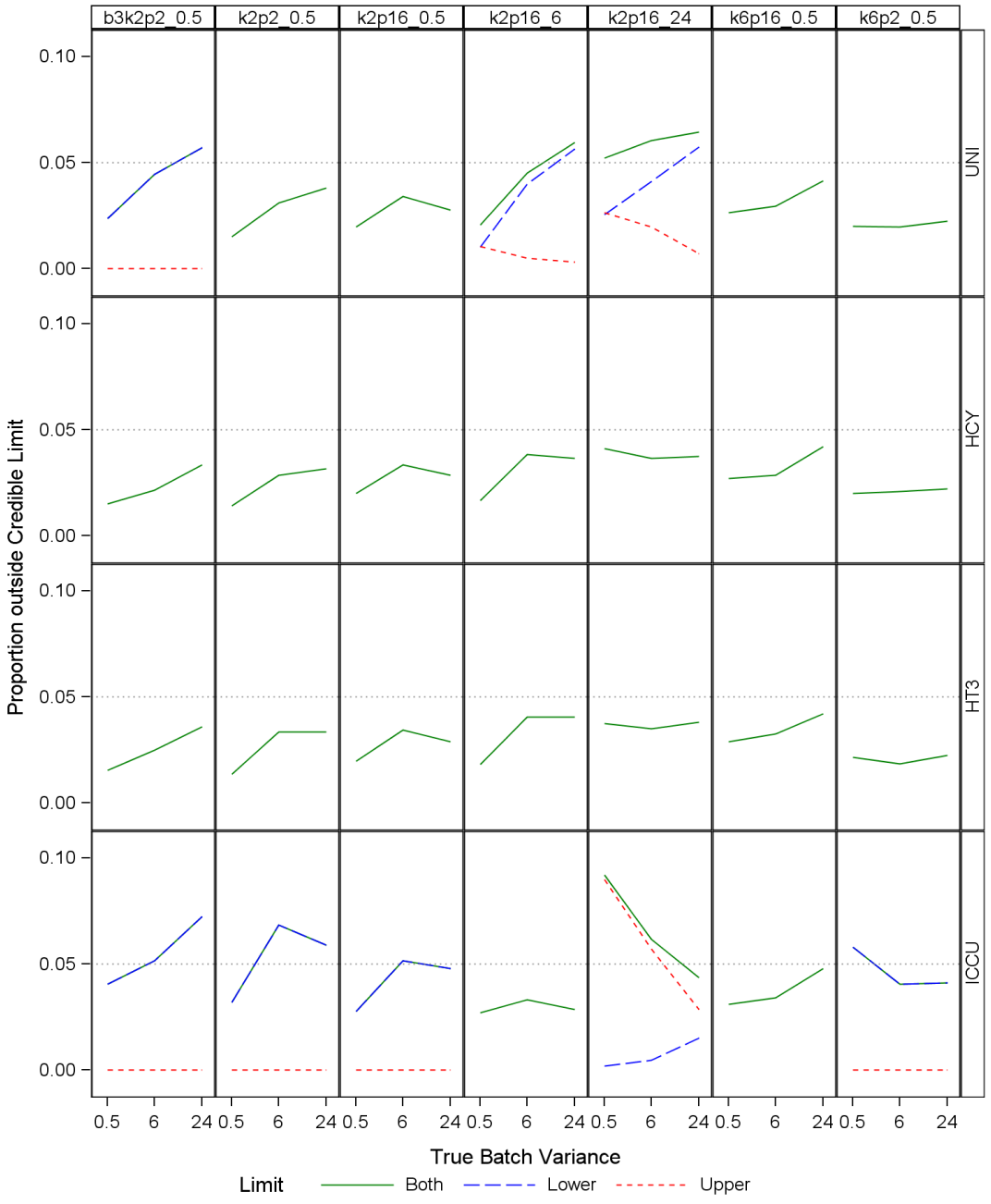


FIGURE 7.16: **Proportion of pct CrIs where true σ_k^2 is outside limits**

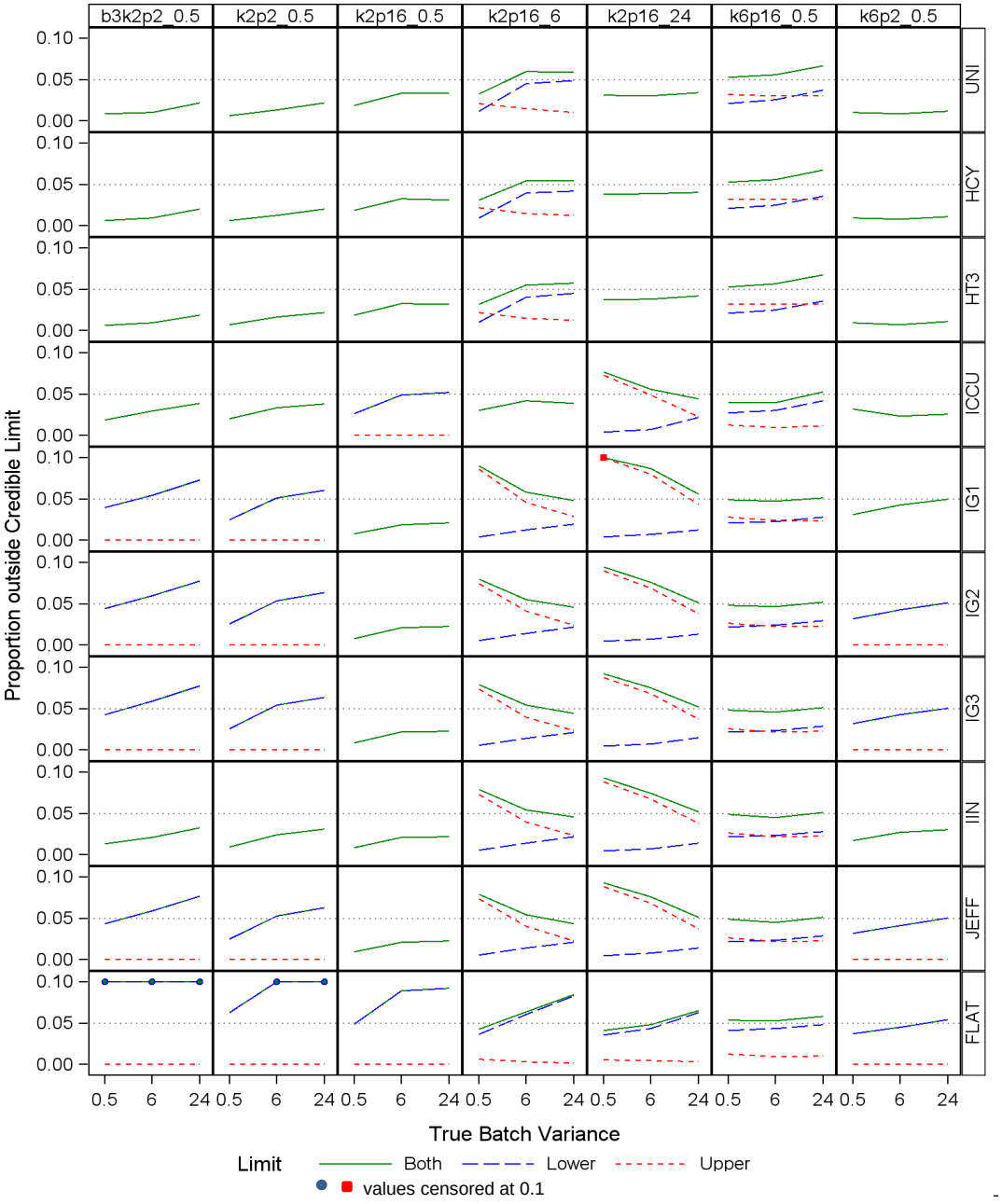


FIGURE 7.17: **Proportion of hpdv CrIs where true σ_p^2 is outside limits**

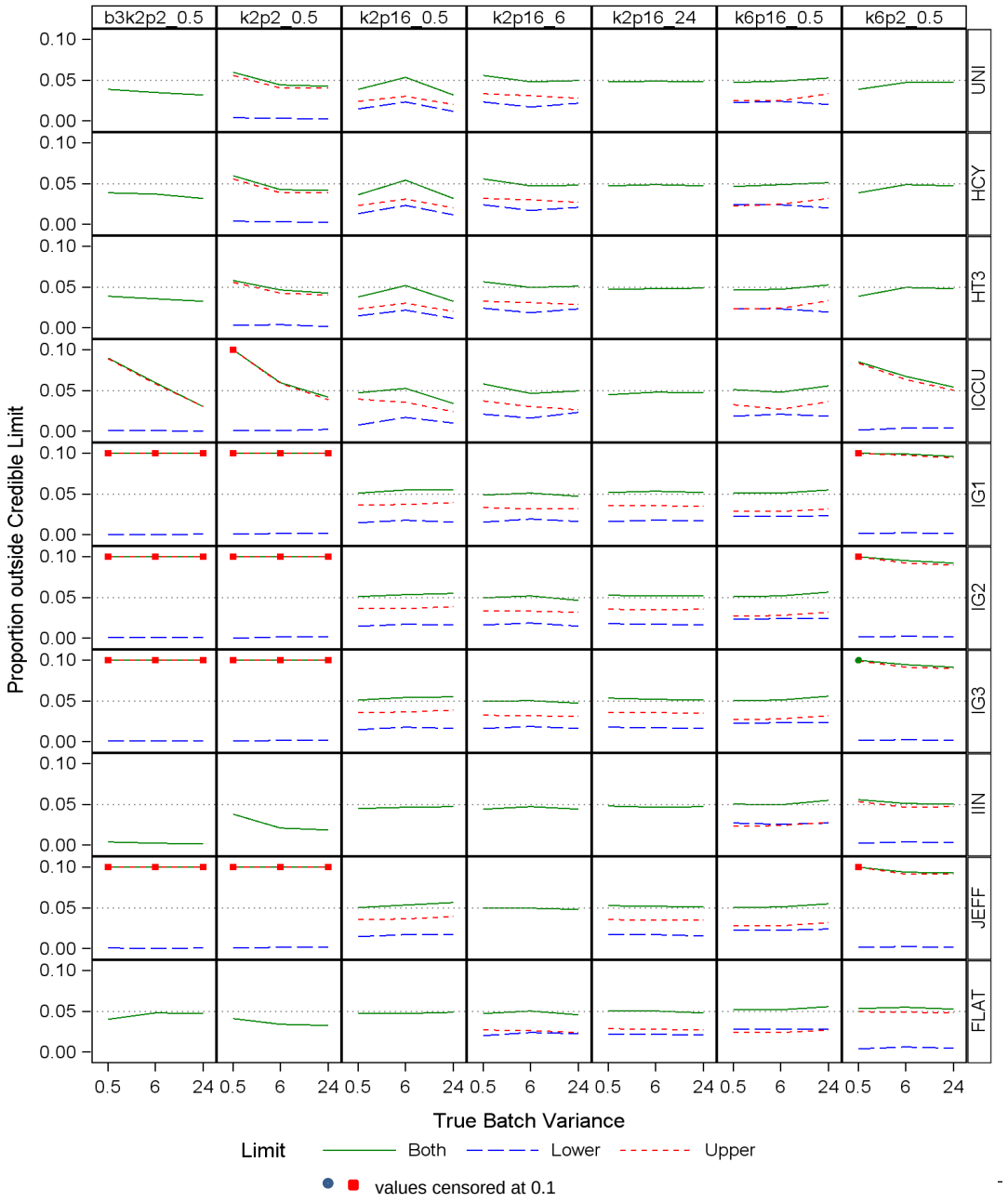


FIGURE 7.18: **Proportion of hpds CrIs where true σ_p^2 is outside limits**

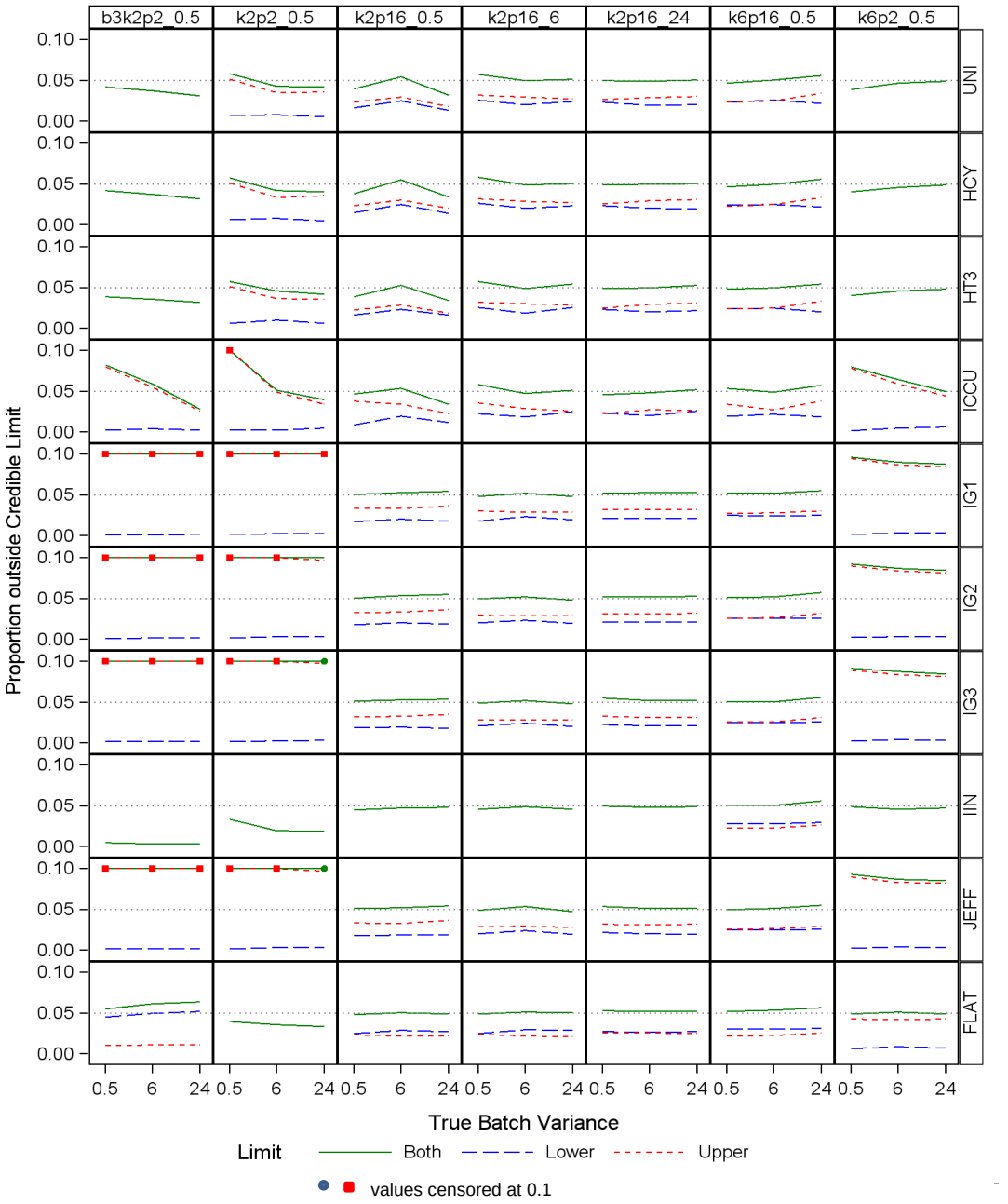


FIGURE 7.19: **Proportion of hpdl CrIs where true σ_p^2 is outside limits**

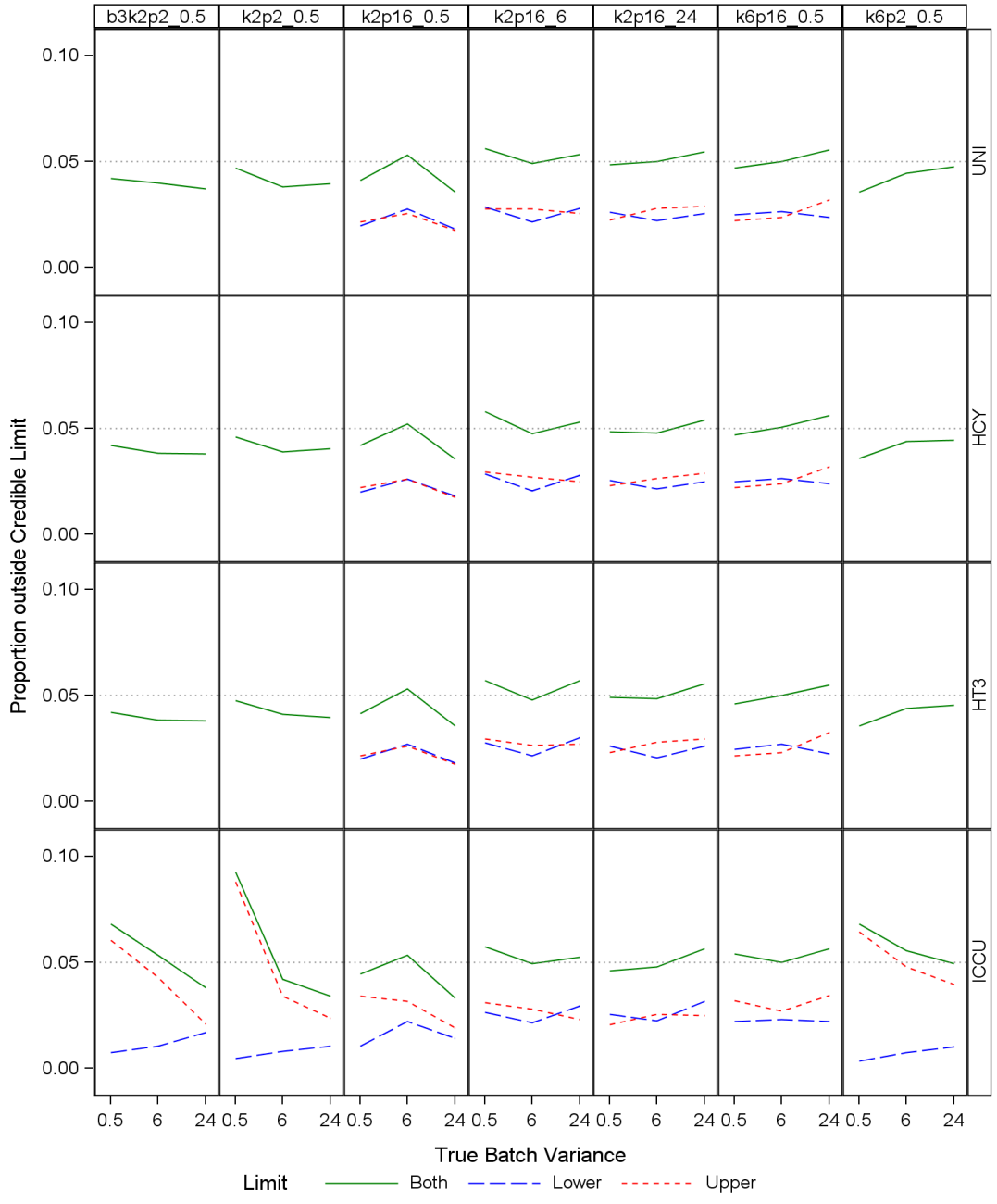
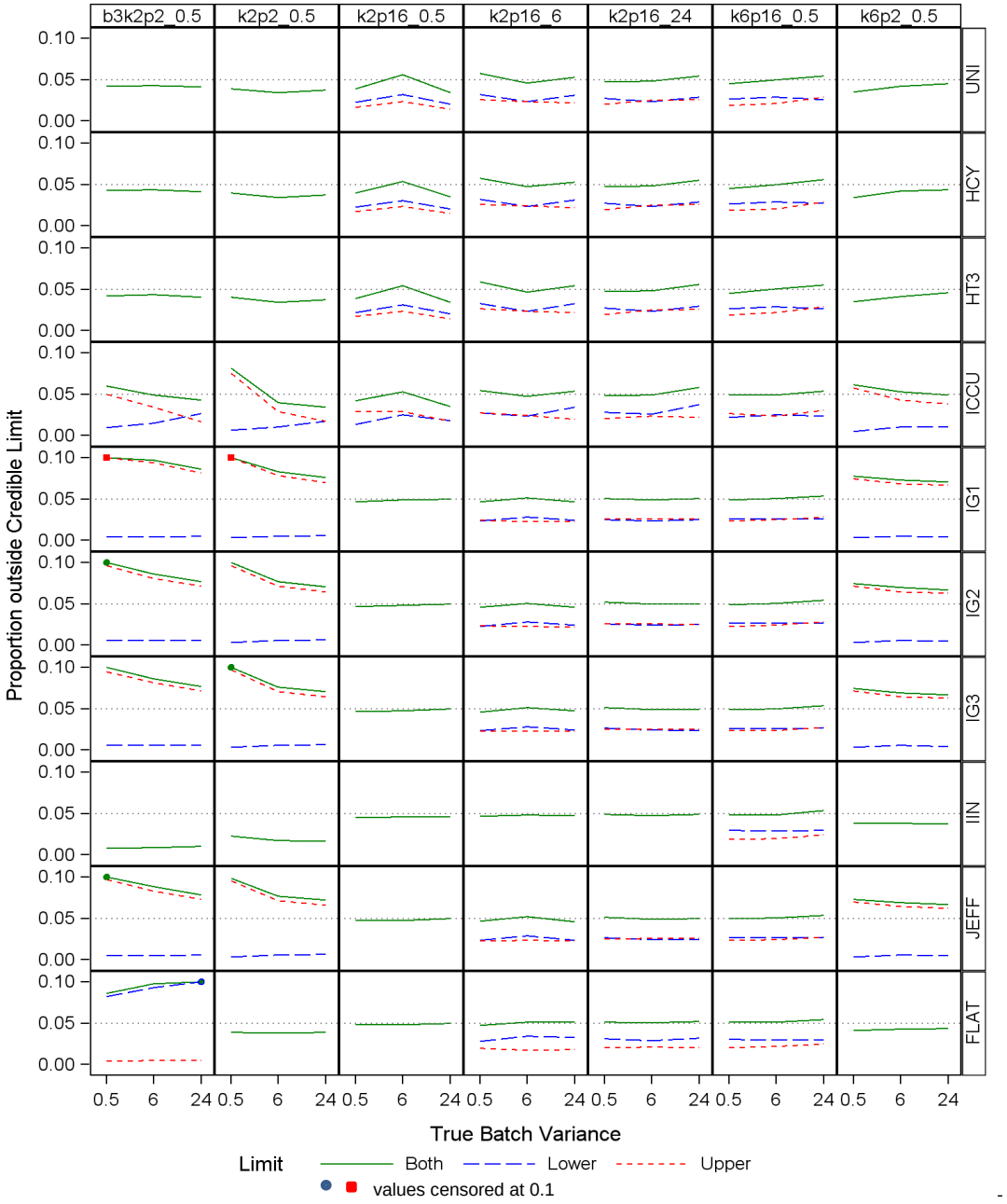


FIGURE 7.20: **Proportion of pct CrIs where true σ_p^2 is outside limits**



HCY and HT3

- The hpdl CrI type did not achieve coverage ≥ 0.94 for total variance for the smallest design b3k2s2_0.5. In Figure 7.7 it is seen that this occurs for a true σ_b^2=0.5 and despite the individual variance components having good coverage (see Figures 7.11, 7.15 and 7.19). A look at some individual analyses found that, for those where the CrI did not include the true value for total variance, the estimates for the individual variance components tended to be quite high but their CrIs were wide enough to encompass the true value, unlike the CrI for the total.
- The pct CrI had poorest coverage for total variance for the smallest designs. Like hpdl CrIs it is seen (see Figure 7.8) that this occurs for a true σ_b^2=0.5 and despite the individual variance components having good coverage (see Figures 7.12, 7.16 and 7.20).
- The hpdv and hpds CrIs for σ_k^2 for the largest design (k6p16_0.5) had a slight tendency to be below the true value (Figures 7.13 and 7.14), even though it will be seen in Figure 8.7 that the median was slightly above for all three values of σ_b^2. This also occurred for the scenario k2p16_24, but only for σ_b^2=0.5.

IG1, IG2, IG3, IIN and JEFF

- Whilst these priors had estimated coverage ≥ 0.94 across all seven scenarios investigated for at least the CrI types hpdv and hpds for total variance, they had poor coverage (mostly < 0.9) for the keg variance for the scenarios with true σ_k^2 of 6 and 24 (k2p16_6 and k2p16_24). In Figures 7.13, 7.14 and 7.16 it is seen that the hpdv, hpds and pct CrIs were increasingly below the true value of σ_k^2 (6 or 24) as σ_b^2 decreased (it will be seen in Figure 8.4 that the posterior median also falls below the true value for σ_b^2=0.5 and 6). Also, whilst the hpdv CrIs for σ_b^2 achieved "good" (or actually very high) coverage for the scenario k2p16_24, the coverage of the hpds and pct CrIs was poor. The hpds and pct CrIs for σ_b^2 tend to be above the true value for σ_b^2=0.5 (see Figures 7.10 and 7.12) but this was not the case for hpdv CrIs (Figure 7.9). It will be seen in Section 7.4.2 that if coverage is not similar, the proportion of true values falling below the lower CrI limit is generally larger for the SD metric than for the variance metric.
- Except for the IIN prior the CrIs also had poor coverage (mostly < 0.9) for the portion variances for the three scenarios with the smallest number of portions per keg (b3k2p2_0.5, k2p2_0.5 and k6p2_0.5). It is seen from Figures 7.17 - 7.20 that the CrIs for σ_p^2 tend to be below the true value and this occurs for all three true values of σ_b^2.
- For the smallest design (b3k2p2_0.5) whilst the hpdv and hpds CrIs had "good" coverage for total variance, the pct CrIs did not. The poor coverage occurred when σ_b^2=0.5 and the CrIs tended to be above the true value (Figure 7.8).

- The IG3 prior can be used as an approximation to Jeffreys' prior (Lunn et al., 2013) since for the IG distribution $f(\phi) \propto \phi^{-\alpha-1}\exp(-\frac{\beta}{\phi})$ and if $\alpha \to 0$ and $\beta \to 0$ then $\phi^{-\alpha-1}\exp(-\frac{\beta}{\phi}) \to \frac{1}{\phi}$ which is the JEFF prior. The results in fact are very similar across the IG1, IG2, IG3 and JEFF priors.

UNI

- The hpds CrI for σ^2_{tot} for the smallest design had much poorer minimum coverage over the three σ^2_b evaluated (0.885) than the hpdv CrI (0.960). Unlike the hpdv interval type, the hpds CrIs for σ^2_{tot} are too frequently above the true value for true σ^2_b=0.5 (Figure 7.6) and similarly for the hpdl and pct CrIs for true σ^2_b=0.5 and 6 (Figures 7.7 and 7.8). This did not occur for the hpds CrIs for HCY, HT3 or ICCU priors - perhaps because those priors have their highest value for the density function on the SD metric at an SD of 0 (see Figure 5.1) whereas the UNI prior is flat across all SDs for which it has positive density.
- For total variance "good" coverage was not quite achieved for the scenarios k2p16_0.5 and k6p16_0.5 (estimated coverage in the range 0.92-0.94) for any interval type (see Table 7.2). From Figures 7.5- 7.8 it is seen that the CrIs largely tended to be above the true value, though for the hpd CrIs there was still non-negligible probability for the CrIs to be below the true value. For example, for the hpdv CrIs the proportion of time the CrI for σ^2_{tot} was above the true value were 0.042 and 0.042 for the two scenarios respectively, and the proportions below were 0.0265 and 0.0195 respectively. This suggests that the intervals are too narrow, rather than too low. This poor coverage occurred when σ^2_b=24 for the hpd CrIs (and σ^2_b=6 for hpdl CrIs) but for σ^2_b=0.5 and 6 for pct CrIs.
- For the k6p16_0.5 scenario the coverage for σ^2_b and σ^2_k for the hpdv and hpds CrIs (and one or other of the hpdl and pct CrI types) did not achieve 0.94. For the batch variance a similar situation occurred to the total variance and the hpd CrIs largely tended to be above the true value, though there was still non-negligible probability for the CrIs to be below the true values (see Figures 7.9 - 7.11). For the keg variance (see Figures 7.13 and 7.14) the hpdv and hpds CrIs largely tended to be below the true value for all three true values for σ^2_b although there was still non-negligible proportion above the true value e.g. for hpdv CrI type and true σ^2_b=24 the proportion of times the CrI for σ^2_k did not include the true value was 0.0565 when it was below and 0.012 when it was above.
- For the k2p16_24 scenario and hpd CrI types the coverage for σ^2_k was slightly less than desired for true σ^2_b values of 0.5, 6 and 24 (0.930 - 0.936). For hpdv and hpds CrI types (see Figures 7.13 and 7.14), for a true σ^2_b=0.5 the CrI for σ^2_k tends to be below the true value a higher proportion of times than desired whilst for true σ^2_b=24 the CrI for σ^2_k tends to be above the true value. For true σ^2_b=6 the CrI for σ^2_k tends not to contain the true value at either end a slightly higher proportion

of times than desired. For the hpdl CrI type and true σ_b^2=6 and 24 the CrI for σ_k^2 tends to be above the true value a higher proportion of times than desired (see Figure 7.15).

ICCU

- The ICCU prior did not achieve coverage ≥ 0.94 for σ_{tot}^2 for five or more scenarios for each interval type, though it is noted that the hpd intervals did achieve coverage greater than 0.9 for all scenarios. The ICCU prior had lower than desired coverage for σ_{tot}^2 with minimum coverage for most of the scenarios around 0.91-0.92. This occurred for true σ_b^2=24 for the hpd CrIs with the CrIs being lower then the true value (see Figures 7.5 - 7.7). For the pct CrIs this occurred for true σ_b^2=0.5 with the CrIs being higher then the true value (Figure 7.8).
- The prior had coverage < 0.94 for at least three scenarios for each interval type for the batch variance, the hpdv and hpds CrIs being particularly poor in that coverage ≥ 0.94 was not achieved for any scenario. Similarly to the total variance for the hpdv and hpds CrI types (Figures 7.9 and 7.10) the poor coverage occurred for a true σ_b^2=24 where the CrI for σ_b^2 tended to be below the true value a higher proportion of times than desired. This was similar for the pct CrIs for the batch variance (Figure 7.12) though it had not been seen for the total variance. For the hpdl CrIs (Figure 7.11) this was also seen for two scenarios. However, for scenario k2p16_24 the CrI for σ_b^2 tended to be above the true value a higher proportion of times than desired for a true σ_b^2=0.5.
- The coverage was poor for the keg variance for scenario k2p16_24 for all CrI types. The poor coverage occurred for a true σ_b^2=0.5 or 6 where the CrI for σ_k^2 tended to be below the true value a higher proportion of times than desired (Figures 7.13 - 7.16).
- Coverage was generally poor for the portion variance for the designs with 2 portions/keg (b3k2p2_0.5, k2p2_0.5, k6p2_0.5). It is seen from Figures 7.17 - 7.20 that the CrIs for σ_p^2 tend to be below the true value like the IG1, IG2, IG3 and JEFF priors, though unlike them this occurs mainly for the true value $\sigma_b^2 = 0.5$.

FLAT

- When the number of batches is only 3 (b3k2s2_0.5) the posterior is improper. The coverage obtained for total variance was < 0.04. The coverage obtained was < 0.06 for the batch variance and poor for the keg variance (0.706 or less). For the total variance and batch and keg variance components the CrIs tended to be above the true value (see Figures 7.5 - 7.16). Thus if it is not recognised that the posterior is improper the CrIs obtained are likely to overestimate the variance.

- Excepting scenario b3k2s2_0.5, the hpdv CrIs had "good" coverage for total variance and the variance components. Similarly for the hpds CrIs, except for total variance for the next smallest design evaluated (scenario k2p2_0.5) where the minimum coverage was 0.895. In Figure 7.6 it is seen that for true σ_b^2=0.5 the hpds CrI for σ_{tot}^2 tended to be above the true value a higher proportion of times than desired.
- For the pct CrIs it is seen that the CrIs for σ_{tot}^2 tended to be above the true value a higher proportion of times than desired for all three true σ_b^2 values (Figure 7.8). Where coverage was poor, it is seen that the CrIs for σ_b^2 and σ_k^2 tended to be above the true value a higher proportion of times than desired (Figures 7.12 and 7.16).

7.4 Summary

7.4.1 Choice of Prior for Routine Use

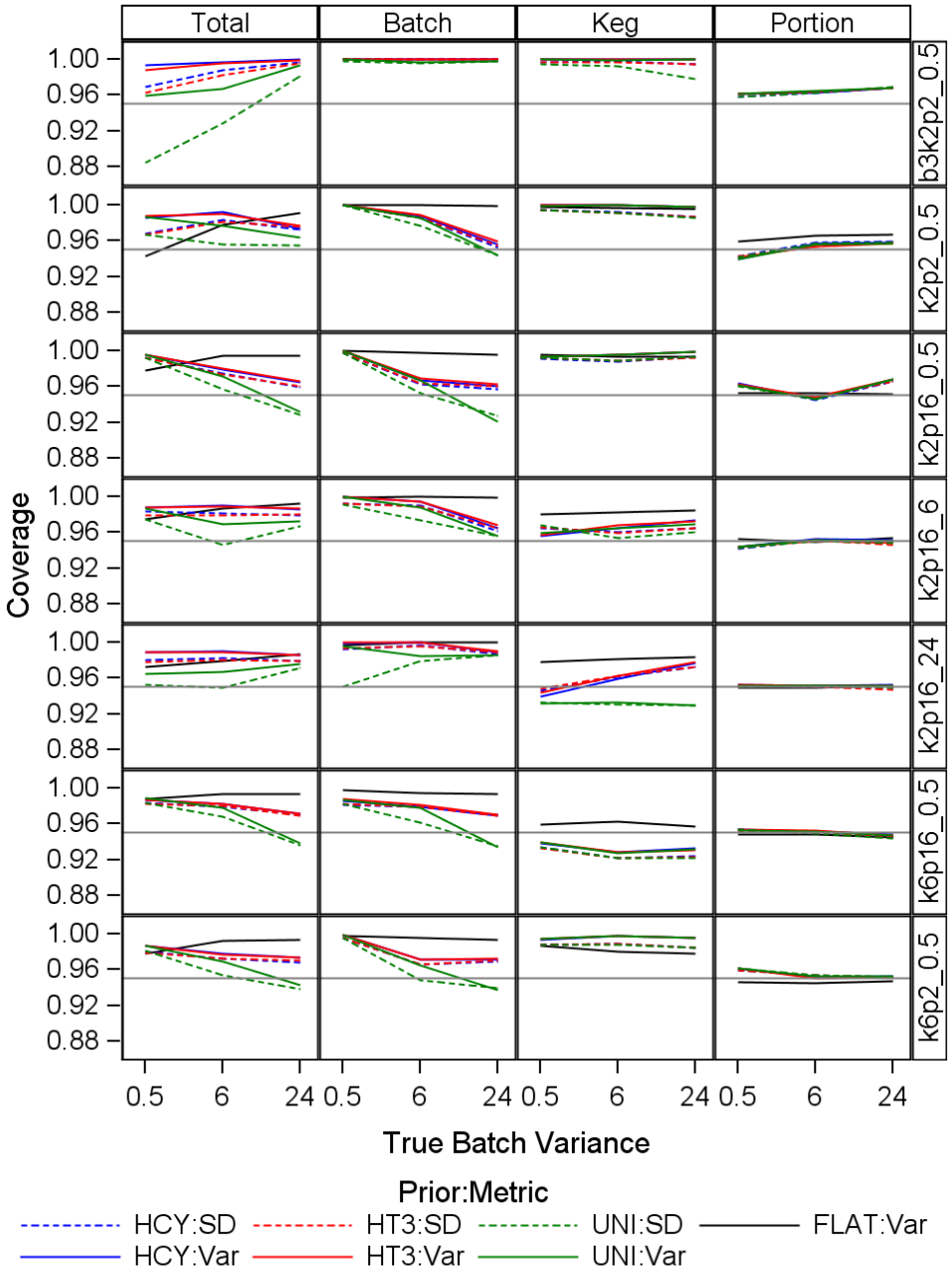
For the original batch sampling design (k2p16_0.5) the mildly informative inverse gamma priors on the stratum variances (I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, IhE6) where the mode or median of the prior was set to 3 or 6, the IG0 prior (IG(1,1)) which was intended to be vague and the ICCU prior which placed a uniform prior on the ratio of the variances, gave poor coverage for the total variance. The mildly informative inverse gamma priors and the IG0 prior had poor coverage for either the batch or keg variance, often both. The ICCU prior had coverage < 0.94 for the batch variance for hpdv, hpds and pct CrIs. The CrIs for priors I2O3, I2O6, I1O3, I1O6, I1E6 tended to overestimate the total variance when the true value was small for all CrI types and additionally I1E3, IhE6, ICCU for pct CrIs. The CrIs for priors I2O3, I1E3, IG0 tended to underestimate the total variance when the true value was large for all CrI types and additionally I1E6, IhE6, ICCU for hpd CrIs.

The priors I2O3, I2O6, I1O3, I1O6, I1E3, I1E6, IhE6, IG0, ICCU were thus considered too informative to be useful for routine use in credible interval estimation in small variance component studies.

Of the remaining priors HCY and HT3 were considered the best for routine use in credible interval estimation in small variance component studies. Whilst no prior gave "good" coverage in all cases, the HCY and HT3 priors seemed most acceptable of those investigated. With hpdv or hpds CrIs they gave "good" coverage for the total, batch and portion variances, and for the keg variance coverage was greater than 0.92. Where the coverage was less than 0.94 (scenarios k2p16_24 and k6p16_0.5) the hpdv or hpds CrIs for σ_k^2 tended to underestimate the true value. The hpdv or hpds CrIs should be used for total variance depending on whether one wishes to estimate variance or SD as using the CrI type based on the same metric as the quantity of interest will give the shortest CrIs. Using this rationale the hpdv CrIs would usually be used for the variance components, though using hpds CrIs to be consistent with hpds CrIs used for total variance would still have similar coverage. However, an alternative possibility is that hpdl CrIs are used for the variance components, as this CrI type gave "good" coverage for all the scenarios investigated. It is not recommended that the hpdl CrIs are used for total variance since for the smallest design and batch variance = 0.5 the coverage was poor and the CrI overestimates the total variance. The log scale also perhaps has some additional relevance as a metric. It was discussed in Section 5.3.2 that variance components are usually referred to on the variance metric probably because the components are additive on this scale and can be represented as a % of the total variation. However, variance components may also be referred to as a % of the intrinsic measurement variation - often the bottom level variance. In this case it will be the ratios that are of interest and

then perhaps the log variance scale has practical meaning. Though the use of hpdl CrIs for variance components is mentioned as a possibility, it is not considered a desirable solution. If a user required, for example, to estimate a sum of small variance components they would be unsure whether to use the hpdv or hpdl metric. The priors HCY and HT3 had very similar coverages. A summary of the coverage across the total variance and variance components is given in Figure 7.21.

FIGURE 7.21: **Coverage of credible intervals for variance components and total variance by selected priors and metrics**



Whilst HCY and HT3 were considered better and are recommended, the UNI prior did not perform too poorly. For total variance its shortcomings were that it did not quite achieve "good" coverage for the scenarios k2p16_0.5 and k6p16_0.5 (estimated coverage in the range 0.92-0.94) for any interval type. The CrIs more often tended to overestimate the total variance, but with a non-negligible proportion of true values above the CrIs this suggests that the intervals are too narrow rather than too low. The hpdv CrIs had higher coverage than the hpds CrIs and achieved "good" coverage for more scenarios. For the smallest design the hpds interval had much poorer coverage (0.885) than the hpdv CrI which was satisfactory (0.960), with the CrIs for σ^2_{tot} overestimating the total variance when σ^2_b=0.5. For the batch variance component the hpd CrIs did not achieve "good" coverage for 2 or more scenarios compared to none for HCY and HT3 (though the coverage was in the range 0.92-0.94 and the pct CrIs had coverage $\geq$0.94). There was a slight tendency for the CrIs to overestimate the batch variance but again there was a non-negligible proportion of true values above the CrIs. For the keg variance components there were more situations (compared to HCY and HT3 priors) where it did not quite achieve "good" coverage compared to HCY and HT3 priors, though again the coverage was in the range 0.92-0.94 for the various CrI types. A summary of the coverage across the total variance and variance components is seen in Figure 7.21.

If software was not a factor for consideration, then none of the priors available in PROC MIXED evaluated across the scenarios (IG1, IG2, IG3, INN, JEFF, FLAT) would be recommended for routine use in preference to HCY and HT3 for credible interval estimation in small variance component studies when both total variance and individual variance components are of interest. However as now discussed the FLAT prior gives "good coverage" except when the posterior is improper and the other priors give "good" coverage for total variance.

For the FLAT prior the scenario with only 3 batches (b3k2p2_0.5) has an improper posterior. When analysed the coverage obtained for total variance was < 0.04 and < 0.06 for the batch variance. Excluding the scenario with only 3 batches (b3k2p2_0.5) which has an improper posterior, the FLAT prior with an hpdv CrI type gave "good" coverage for total variance and all the variance components. Thus it is a contender. However, if the FLAT prior was to be used routinely, it would require a knowledge of when the posterior is improper in order that misleading results aren't obtained (since the software does not provide this). It also would not provide an answer for designs where the posterior is improper with a FLAT prior, even though designs with only three levels of a factor are not unusual in variation studies. A summary of the coverage across the total variance and variance components is given in Figure 7.21, for the hpdv CrIs only.

The IG1, IG2, IG3, INN and JEFF priors had estimated coverage ≥ 0.94 across all seven scenarios investigated for at least the CrI types hpdv and hpds for total variance. Thus for total variance these priors are acceptable. However, they had poor coverage

(mostly < 0.9) for the keg variance for the scenarios with true σ_k^2 of 6 and 24 (k2p16_6 and k2p16_24). The hpdv, hpds and pct CrIs were increasingly below the true value of σ_k^2 (6 or 24) as σ_b^2 decreased. It is also interesting to note that, whilst the hpdv CrIs for σ_b^2 achieved "good" (or actually very high) coverage for the scenario k2p16_24, the coverage of the hpds and pct CrIs was poor with the CrIs for σ_b^2 tending to be above the true value for σ_b^2=0.5. Except for IIN the priors also had poor coverage (mostly < 0.9) for the portion variances for the three scenarios with the smallest number of portions per keg (b3k2p2_0.5, k2p2_0.5 and k6p2_0.5) with the CrIs tending to be below the true value. IIN had "good" coverage for the portion variance but it was specifically chosen to be informative for σ_p^2 (aligned with its true value of 6). For the smallest design (b3k2p2_0.5) whilst the hpdv and hpds CrIs had "good" coverage for total variance, the pct CrIs did not. The poor coverage for total variance occurred when σ_b^2=0.5 and the pct CrIs tended to be above the true value. The results were seen to be very similar across IG1, IG2, IG3 and JEFF (note IG3 is an approximation to JEFF).

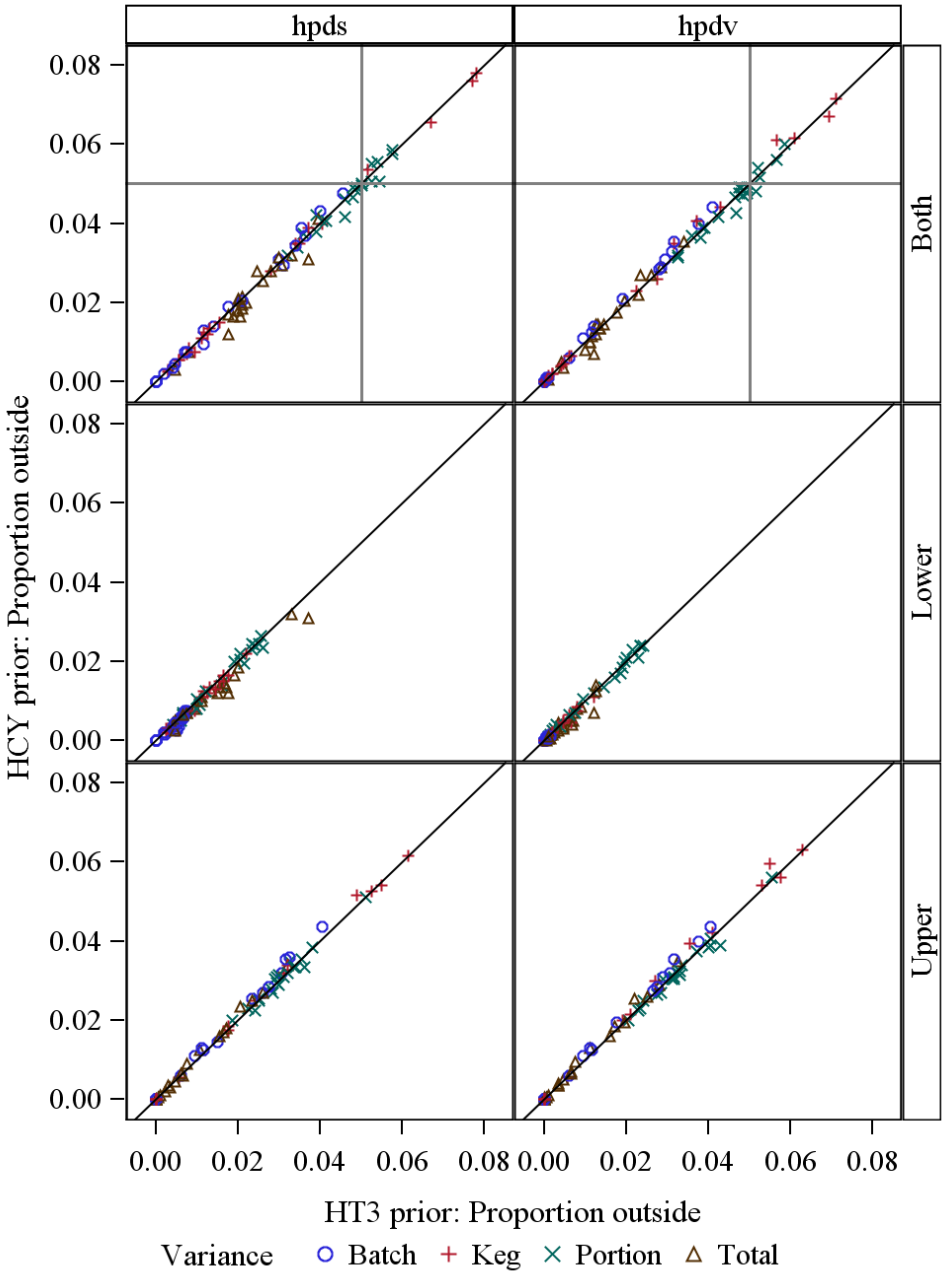
7.4.2 Comparison of the Priors and CrI types Most Likely to be Useful

The priors and CrI types most likely to be useful for routine use (HCY and HT3 priors with hpdv and hpds CrIs and FLAT prior with hpdv CrI) are now further compared. The UNI prior is also compared with the HT3 prior. The hpdv and hpds CrIs were also compared across a range of priors.

The proportion outside the hpds and hpdv CrIs is compared in Figure 7.22 for the HCY and HT3 priors. The proportion outside the lower limit, the proportion outside the upper limit and the sum of both the proportions outside the lower and the upper limits (where lines are drawn at 0.05 corresponding to coverage of 0.95) are plotted. The line of equality is also drawn on the graph and it is seen that the HCY and HT3 priors give very similar proportions outside the CrIs.

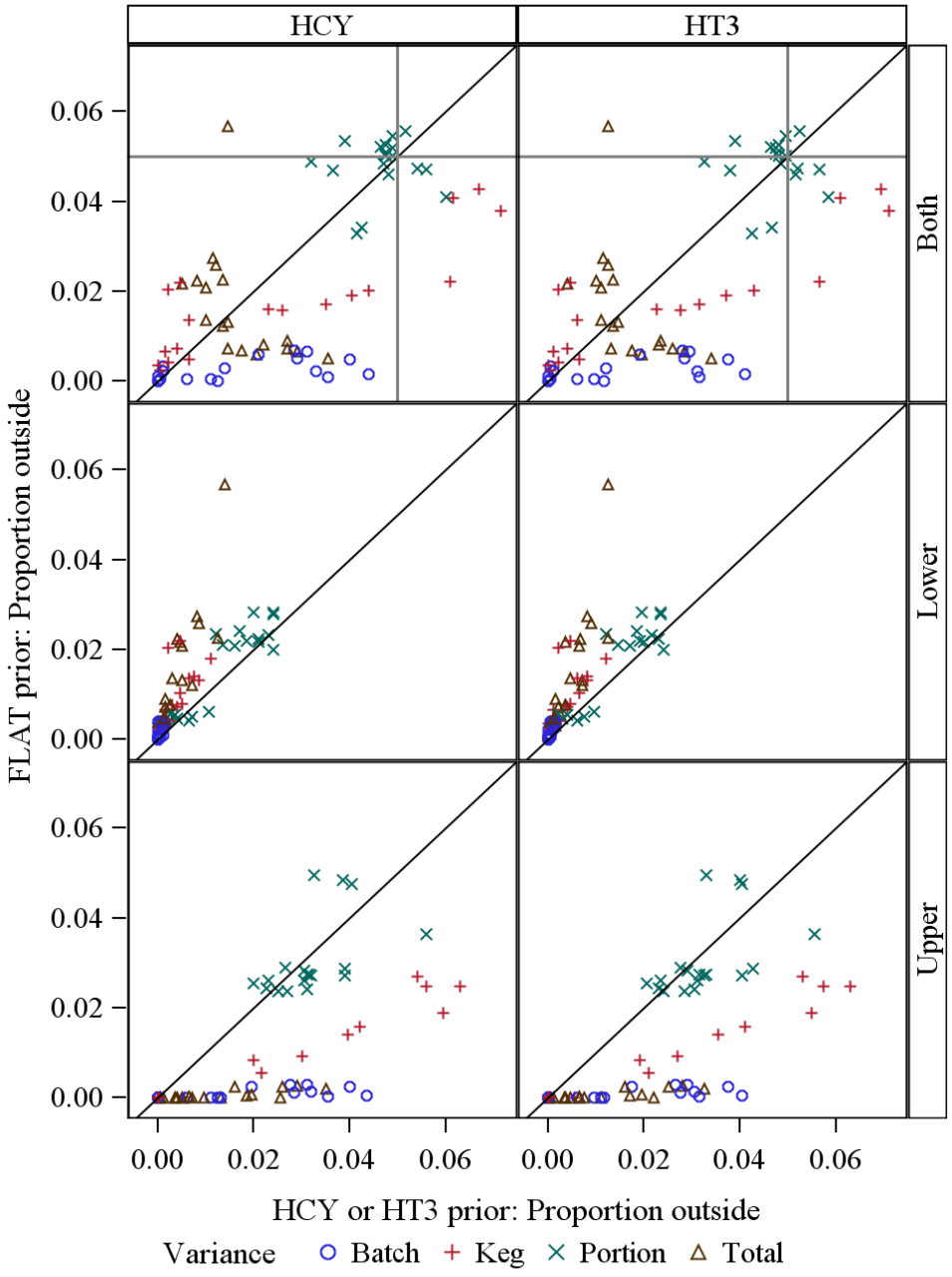
The proportion outside the hpdv CrIs for the FLAT prior is compared against those for the HCY and HT3 priors in Figure 7.23. Scenario b3k2p2_0.5 was excluded because the posterior is improper and it was seen in Section 7.2.2 that the coverage was extremely low (< 0.06) for the total and batch variances. It is seen that the FLAT prior does not give similar proportions outside the CrIs to those for the HCY and HT3 priors. For total variance there is a correlation between the proportion outside the lower limit for the FLAT prior and those for the HCY and HT3 priors, with the FLAT prior having a higher proportion (points lie to the left of the line of equality). For the upper limit there may be some association but the proportion for the FLAT prior was very small (points lie far to the right of the line of equality when the proportion was not small for the HCY or HT3 priors). When looking at the sum of both proportions outside for the total variance there is little positive correlation between those for the FLAT prior and those for the HCY and HT3 priors and the proportions fall either side of the line of

FIGURE 7.22: **Comparison of HCY and HT3 priors for proportion of analyses where true σ^2 is outside limits for hpdv and hpds CrIs. (Line of equality is plotted and for proportion outside both limits a line is drawn at 0.05)**



equality. Generally for the batch and keg variance components, like the total variance, the FLAT prior had a larger proportion outside the lower limit and a smaller proportion outside the upper limit than the HCY and HT3 priors, though for the batch variance the proportions outside the lower limit were very small even for the HCY and HT3 priors.

FIGURE 7.23: **Comparison of HCY and HT3 with FLAT prior for proportion of analyses where true σ^2 is outside limits for hpdv CrIs, excluding scenario b3k2p2_0.5. (Line of equality is plotted and for proportion outside both limits a line is drawn at 0.05)**



The proportion outside the hpds and hpdv CrIs for the UNI and HT3 priors are compared in Figure 7.24. The HCY prior was not plotted as it was seen in Figure 7.22 that the HCY

and HT3 priors gave very similar results. It is seen that the UNI prior generally gives similar but slightly smaller proportions outside the upper limit of the CrIs compared to the HT3 prior (points generally lie slightly to the right of the line of equality). However for the lower limit the proportion outside for the UNI prior is considerably higher than the HT3 prior for many situations for the total and batch variance and some situations for the keg variance (many points lie to the left of the line of equality). This observation for the lower limit is reflected in the sum of both proportions outside the limits and there are a number of situations where the proportion exceeds 0.06 for the UNI prior. For the portion variance there is a good relationship between the proportions outside for the UNI and HT3 priors.

The proportion outside the limits for hpdv and hpds CrIs are compared in Figures 7.25 and 7.26 for the HCY and HT3 priors. In Figure 7.25 the points are identified by the variance and in Figure 7.26 they are identified by the scenario. It is seen that there is a strong correlation between the two CrI types and whilst the points are reasonably close to the line of equality they mainly lie to the right and thus the proportion outside is generally larger for the SD metric than for the variance metric.

The proportions outside the lower and upper limits (respectively) for hpdv and hpds CrIs are compared in Figures 7.27 and 7.28 for a range of priors. Note for clarity for the smaller proportions, each figure has two sub plots with the plotting range split: plot a) where the proportions outside for the hpdv and hpds CrIs are not both less than 0.08 and plot b) where they are. It is seen that generally across the priors there is a reasonable relationship between the proportions outside the lower limit for the two CrI types (strongest for HCY and HT3). Generally if the points don't lie on the line of equality they tend to lie to the right and thus the proportion outside is generally larger for the SD metric than for the variance metric. For the upper limit the correlation is stronger but here if the points don't lie on the line of equality they tend to lie to the left and thus the proportion outside is generally smaller for the SD metric than for the variance metric.

To summarise, the HCY prior gave very similar results to the HT3 prior when examining the proportion of analyses where the true value was outside the CrI limits (lower or upper) whereas the FLAT and UNI gave fairly different results from the HT3 prior. For the UNI prior this difference was particularly seen at the lower limit whereas for the FLAT prior it was seen for both upper and lower limit. The proportion outside the lower limit of the CrIs is generally larger for the SD metric than for the variance metric across priors, noting that the difference is smallest for the portion variance. For the upper limit the converse applied, though in general the differences between the proportions outside for hpdv and hpds CrI types were smaller than those seen for the lower limit.

FIGURE 7.24: **Comparison of UNI and HT3 priors for proportion of analyses where true σ^2 is outside limits for hpdv and hpds CrIs. (Line of equality is plotted and for proportion outside both limits a line is drawn at 0.05)**

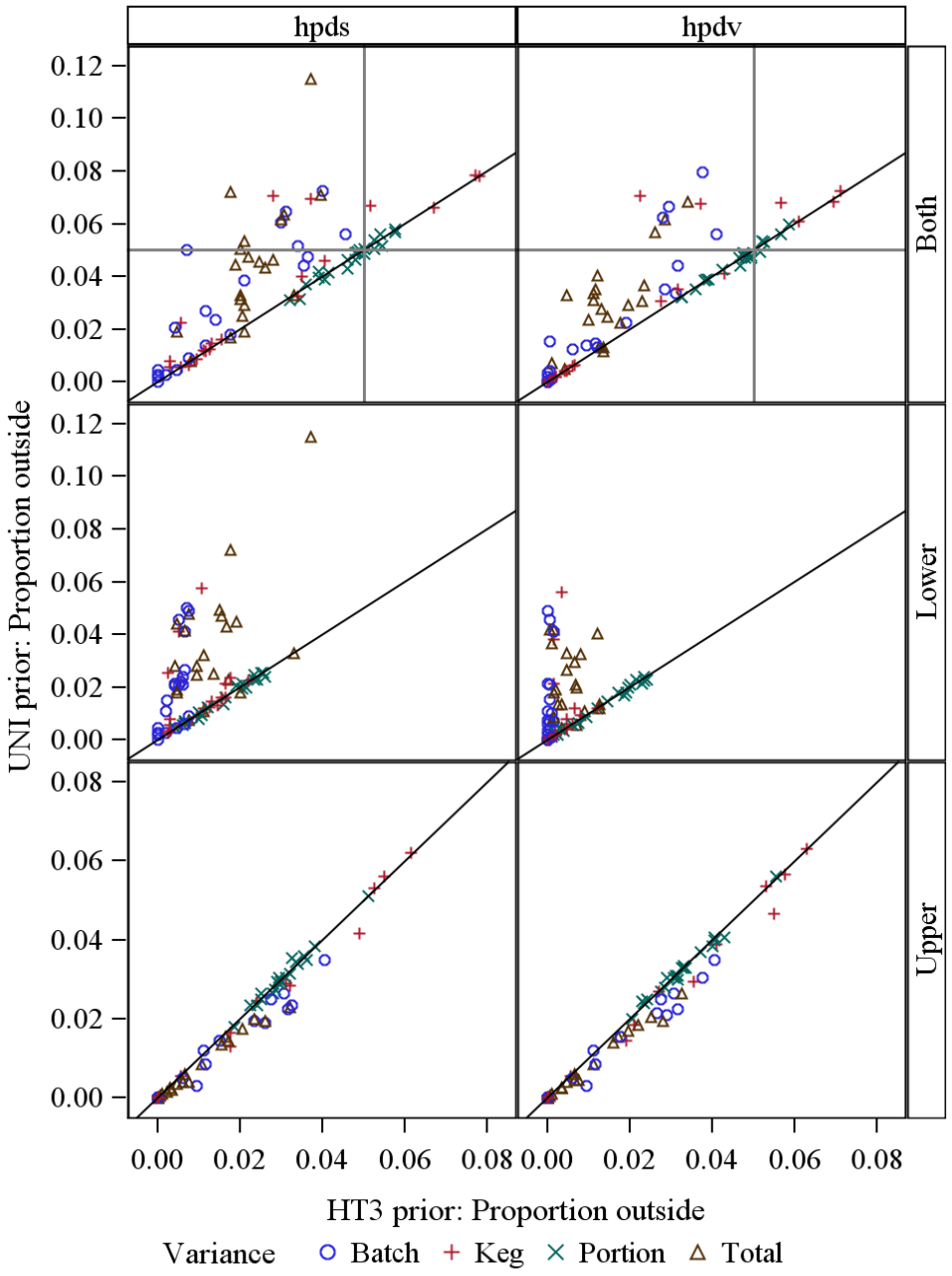


FIGURE 7.25: **Comparison of hpdv and hpds CrIs for proportion of analyses where true σ^2 is outside limits for HCY and HT3 priors by variance. (Line of equality is plotted and lines indicating proportion outside =0.05)**

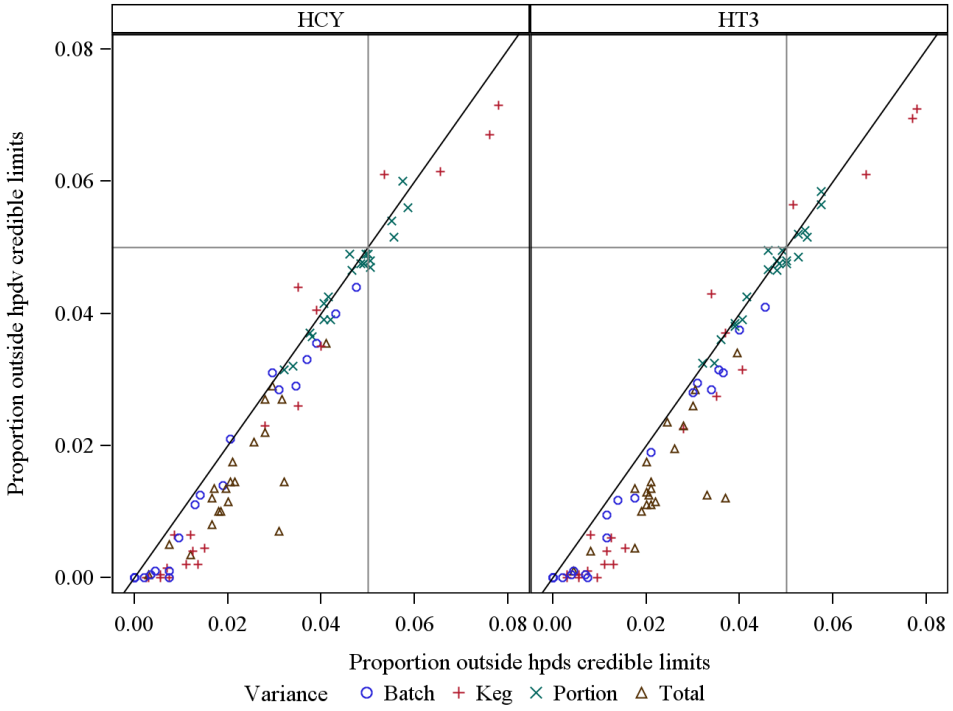


FIGURE 7.26: **Comparison of hpdv and hpds CrIs for proportion of analyses where true σ^2 is outside limits for HCY and HT3 priors by scenario. (Line of equality is plotted and lines indicating proportion outside =0.05)**

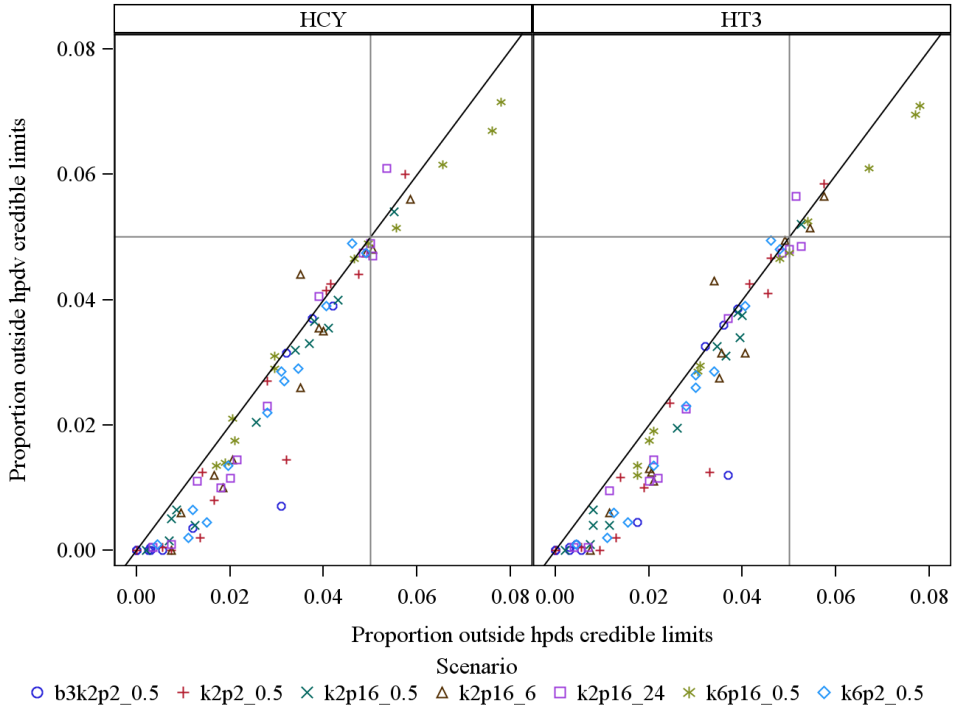


FIGURE 7.27: **Comparison of hpdv and hpds CrIs for proportion of analyses where true σ^2 is outside lower CrI limit for priors**

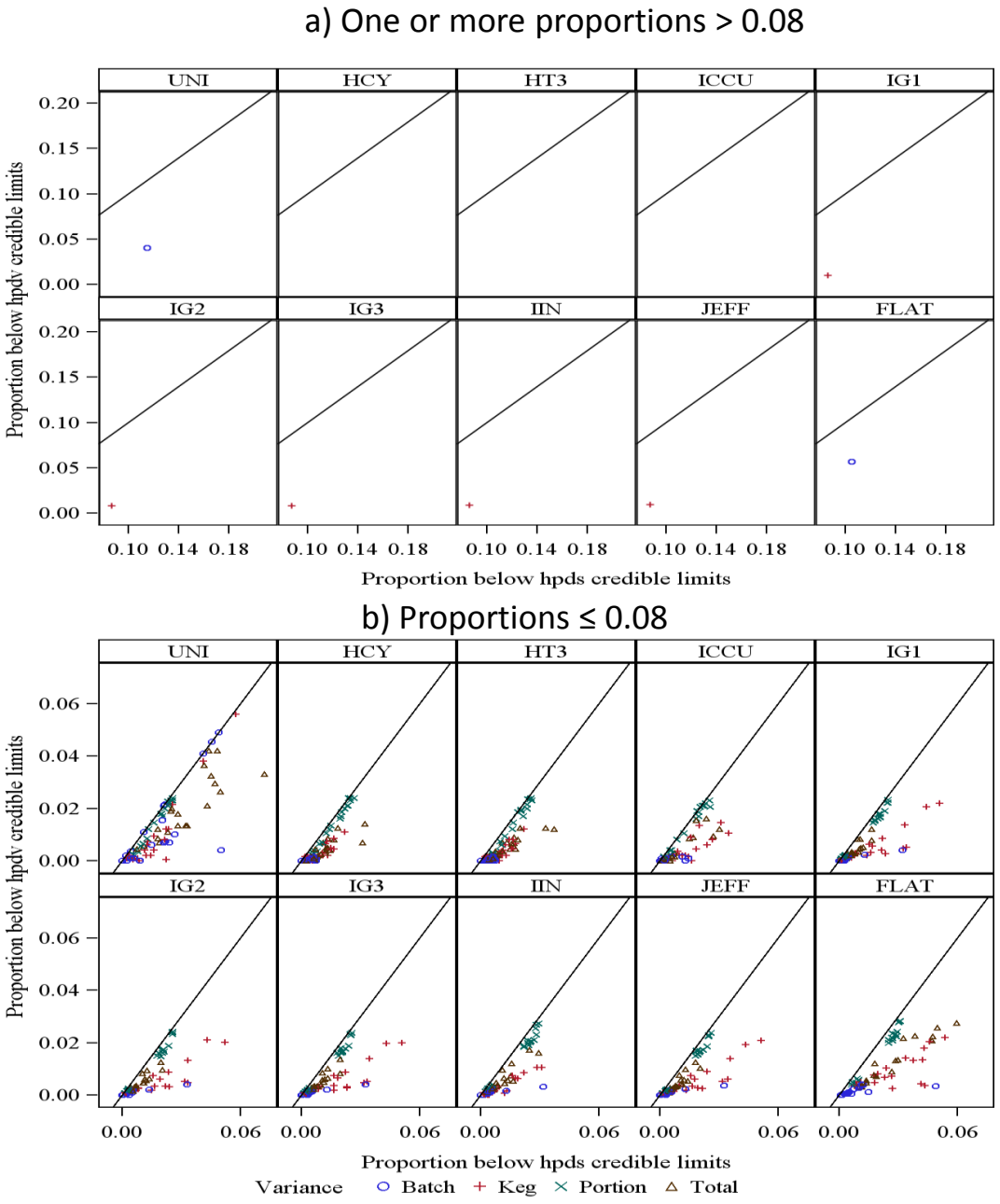
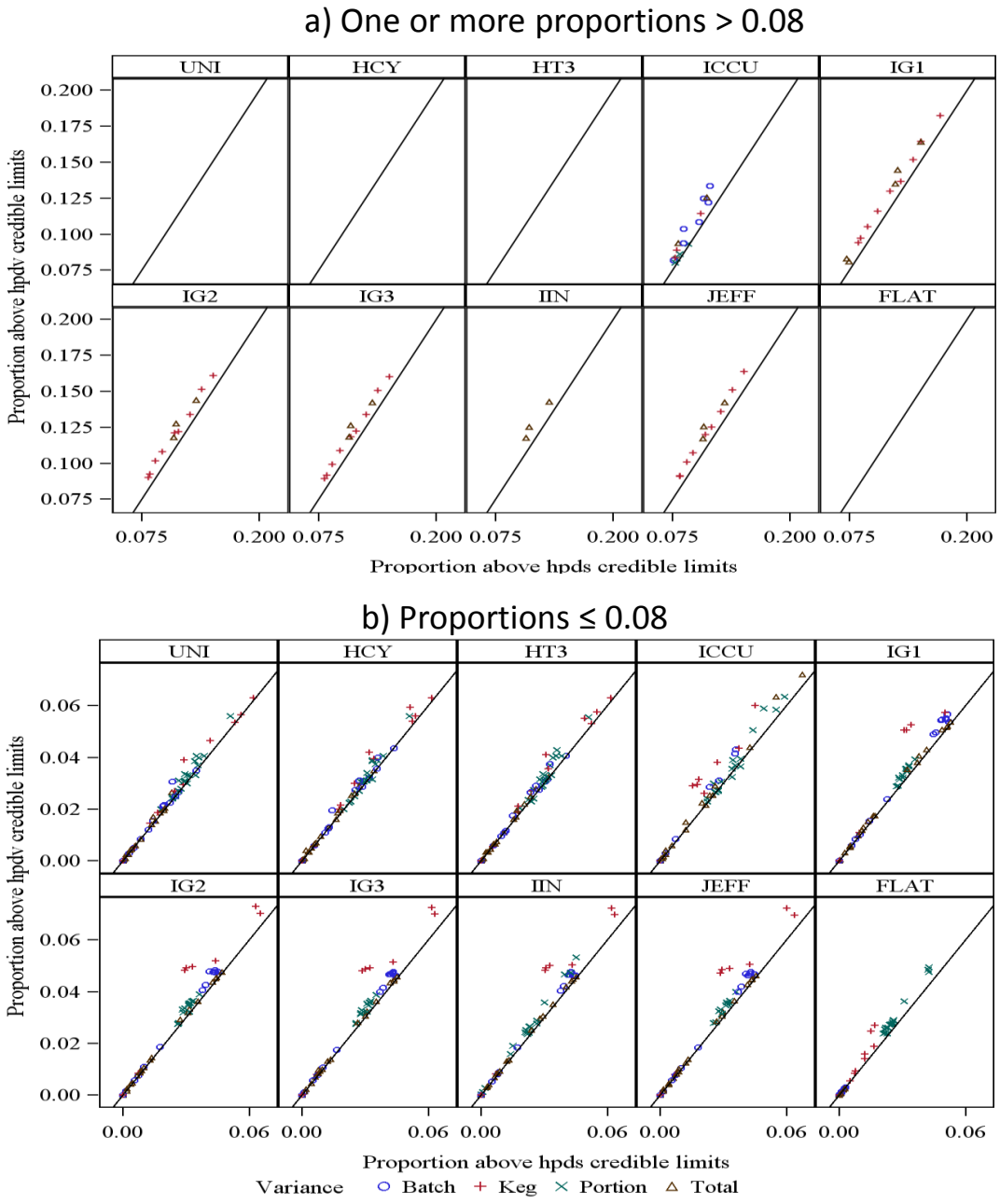


FIGURE 7.28: **Comparison of hpdv and hpds CrIs for proportion of analyses where true σ^2 is outside upper CrI limit for priors**



7.4.3 Other Comments and Conclusions

The mildly informative inverse gamma priors on stratum variances were found to be too informative to be useful for routine use. If further investigation of mildly informative priors is desired then it is recommended to investigate inverse gamma priors based on the median taking the most likely value but with a smaller shape parameter (α) than 0.5. Achieving only a mildly informative resultant prior distribution on another variance component when the data provides good information in estimating one variance component was seen to be a key difficulty for the mildly informative priors using independent inverse gamma distributions on the stratum variances. That is a problem here when a set of priors is required that is at most mildly informative. However, it is also likely to be a problem in an alternative situation where it is desired to make use of historical information in inverse gamma priors. The choice of parameters for an informative inverse gamma prior will be very important in order that the prior does not dominate the information from the study itself in an unintended manner.

For the design with 6 batches, 2 kegs/batch and 16 portions/keg (scenarios k2p16_0.5, k2p16_6, and k2p16_24), it was seen that the coverage of the CrIs obtained for the IG1-IG3, IIN and JEFF priors for the keg variance are strongly influenced by the true batch variance. As the batch variance decreases the CrIs for keg variance for true σ_k^2= 6 or 24 increasingly fall below the true values. This dependence was less strong for the HCY, HT3 and FLAT priors. A criticism of the JEFF prior and similarly applies to the inverse gamma priors are that they are not independent of the design. However another issue seen for these priors is that for the coverage for the CrIs for one variance component can be strongly dependent on the true value of another variance component.

There were no priors for which the pct CrIs gave "good" coverage for total variance across all scenarios. The coverage was poorest (no prior gave "good" coverage) for the scenario with the smallest design (b3k2s2_0.5) and most priors had coverage <0.94 for the next smallest design (k2s2_0.5). The poor coverage particularly occurred for σ_b^2=0.5. Given that the posterior distribution is asymmetric and bounded by zero the hpd intervals are likely to be more appropriate than pct intervals as they cover the highest density part of the distribution and also the mode. Due to the asymmetry the limits of the pct CrI will be higher than those of the hpd CrI. In the case of the small designs where poor coverage was obtained, the priors have strong influence on the posterior as there is little data and for a true value of σ_b^2=0.5 this tends to pull the lower limit of the pct interval above the true value. The pct CrIs also did not offer advantages for the individual variance components.

Chapter 8

Credible Intervals and Posterior Medians

Contents

This chapter examines the credible interval limits and posterior medians produced for the individual variance components and the total variance from the Bayesian analyses performed. It also further investigates the low coverage for scenario k6p16_0.5 seen for HCY, HT3 and UNI priors.

The main criterion for the choice of priors which can be used routinely in credible interval estimation in small variance component studies is the coverage of the CrIs as discussed in Section 7.1. Chapter 7 provided details of the coverage of credible intervals produced for the individual variance components and the total variance from Bayesian analyses using the various priors selected for evaluation in Chapter 5. The choice of priors based on coverage was discussed in Section 7.4.1 and the coverage of the better priors was then compared further in Section 7.4.2. If the coverage is considered acceptable then

the limits and length of the CrIs becomes relevant to the choice and are thus examined in this chapter. Additionally examination of the CrI limits provides additional insight into why some priors in some situations did not provide "good" coverage.

In order to provide a summary view of the credible interval limits, the median limits over the simulated datasets will be plotted for the various scenarios, priors etc. that have been explored. However, given that the intervals are asymmetric around the true value and are wide (due to the small design sizes), visualising them on a linear scale will obscure details of the lower limits. In addition the appearance of interval length on a linear scale will be dominated by the upper limits. An alternative scale - the icc (intraclass correlation coefficient) scale - is developed in Section 8.1 and is proposed as a useful method of visualising credible intervals for variance components. Section 8.2 examines the interval limits and posterior median across a range of priors. This includes some priors which did not provide "good" coverage (>0.94) in many situations but it is useful to see how the CrIs compare with those priors which gave better coverage. Whilst no prior gave "good" coverage in all cases, as discussed in Section 7.4, the HCY and HT3 priors were considered the best of those evaluated for routine use in credible interval estimation in small variance component studies. With hpdv or hpds CrIs they gave "good" coverage for the total, batch and portion variances, and for the keg variance coverage was greater than 0.92 for scenario k6p16_0.5 and "good" for the remaining scenarios. The CrIs for the UNI, FLAT, IIN, IG2 and JEFF prior are also compared with those of the HCY and HT3 priors. To gain further understanding, the individual CrI limits are investigated in Section 8.3 for some situations where the coverage was lower than desired, especially those for σ_k^2 for the HCY and HT3 priors for scenario k6p16_0.5. Then in Section 8.4 the reason for the low coverage for scenario k6p16_0.5 for σ_k^2 is further explored, to try to understand whether it is an underestimate of the true coverage due to an aspect of the simulations or MCMC procedure, or whether the true coverage is really low. It is found to be a true result and the sensitivity to the choice of the HT3 prior scale parameter, the σ_k^2 value and size of design is then explored. Though not the prime objective of the thesis, the medians are compared with the REML estimates in Section 8.5. A summary of the findings in this chapter is given in Section 8.6.

8.1 Scales for Visualising Credible Intervals

When simulation studies evaluate CrIs for parameters, the ends of the CrIs can be summarised in a table. However to aid the reader to assess the results in the light of the objective of the study, the ends are often plotted and/or the length of the CrI is summarised - for example see Figures 8 and 11 in Liu et al. (2015) (sample intervals plotted) and Tables 6 and 7 of Browne and Draper (2006) (interval lengths reported). However, when the sampling distribution is not symmetric, as is the case for variance components, use of a linear scale may be misleading or obscure aspects of interest. For

example, in Figure 11 in Liu et al. (2015) it is not possible to see any variation in the lower end of the CrIs for the group-level standard deviation parameter nor visually see the result that the smoothing of Spin "tends to (correctly) pull the lower bound of the interval all the way to the boundary". When viewing the CrIs plotted, the only differences which can be seen are those at the upper bound. Another illustration of why the linear scale seems less useful is the three credible intervals (0.01,22.49), (1.5,24) and (2.99,25.49). They all have the same length but for a variance or SD the differences between the lower end will usually be more important than those at the upper end. An alternative scale that might be used is a log scale. However, for small studies and small variance components this can have the opposite effect and result in too much focus on the lower limits. Figure 8.1 shows the hpdv CrIs (and posterior median shown as a cross bar) for the σ_k^2 for design b3k2s2_0.5 for selected priors plotted using various scales. The median values across all the simulated datasets of the CrI limits and the median are plotted. The true value (of 0.5) is shown by a dashed line. The CrIs plotted on the linear and log scales are shown in sub plots a) and b). It is seen that the CrIs plotted on the linear scale are dominated by the larger values and thus any assessment of length will focus on the upper values and is likely to favour priors with higher density on lower values which have a smaller upper value. In contrast, those plotted on the log scale are dominated by the small values and thus any assessment of length will focus on the lower values. In the context of the application and a true value of 0.5, a variance of less than 0.00001 is not practically different from 0.001, yet that range is the visual focus of the graph. Ionan et al. (2014) also reported widths of intervals for the intra-class correlation and noted that Bayesian intervals despite being wider than those for the other methods, generally had lower coverage than nominal. This is another situation where the sampling distribution would be expected to be asymmetric and thus an alternative scale to linear is desirable.

Though the credible interval results produced for this thesis are tabulated in Appendix G, it was considered desirable to provide a visual summary given the large number of results. Since neither the linear nor log scales are satisfactory an alternative scale is developed. This was motivated by consideration of the frequentist CI for a simple variance estimate. A frequentist CI would be given by $CI_l = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}} = CI_u$, where the variance is estimated from n data points and the χ^2 has n-1 degrees of freedom. The linear difference $CI_u - CI_l$ is dependent on S^2 whereas considering the length of the CI as a ratio $\frac{CI_u}{CI_l}$ is only dependent on the degrees of freedom (amount of information) not S^2 itself. Thus the ratio could be considered an alternative measure of length. However, the ratio is very dependent on the value of the denominator and related to considering the length on a log scale. An alternative to a simple ratio is proposed, which is a transformation akin to an intraclass correlation, giving a CrI length of $\frac{CI_u}{CI_u+constant} - \frac{CI_l}{CI_l+constant}$. If the constant is based on S^2 then again this width is not dependent on S^2. The estimation of CIs for variance components from complex study

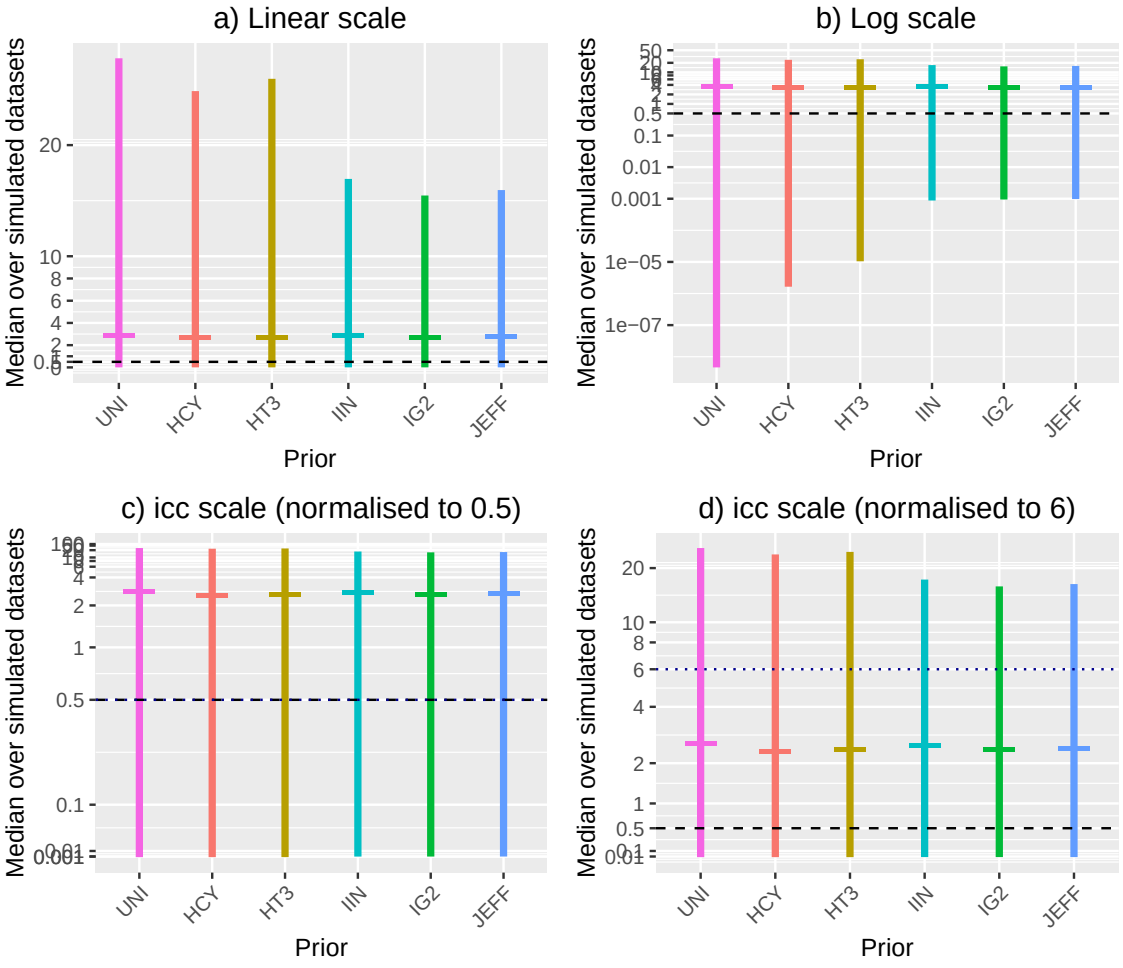
designs does not have the simple analytical formula of a CI for a simple variance estimate but it seems reasonable to apply similar principles. Scales based on the intraclass correlation (icc) are also shown in Figure 8.1. The two subplots labelled c) and d) show the CrIs transformed using the equation $\frac{\text{value}}{\text{value}+constant}$ (denoted the icc scale). For c) the variance component is scaled to its true value 0.5 (since all the analyses using the different priors aim to estimate the same true value this seems more appropriate than using the estimated value S^2 which varies according to the prior). This scale gives similar focus to both ends of the CrIs and also shows, in the context of the true value of 0.5, the CrIs seem reasonably similar. In the second case the constant is chosen to be 6 (plotted with a dotted line), the true value of σ_p^2. Here the CrIs for σ_k^2 are scaled according to the intrinsic variance. Visually this now concentrates more attention to values of practical importance i.e. those around 6. When the length of the CrIs is narrow compared to the values the choice of scale will make little difference. However, in these small studies the CrIs are wide compared to the values and thus an appropriate choice of scale is important. In Section 8.2 the icc scale normalised to 6 (the true value of the intrinsic variance) was chosen for general plotting purposes for variance components as this best allows the viewer to see what is of interest without misleading them. Sometimes the icc scale was normalised to the true value of the variance component itself for a closer look at the results. For the total variance and total SD the icc scale was normalised to 18 and 4.24 respectively representing the variance associated with batch, keg and portion variances of 6.

Example R code which produces plots using the icc scale is given in Appendix G, Section G.1. This makes use of the functionality in the ggplot2 and scales packages in R (Wickham, 2009, 2016).

8.2 Credible Interval Summary for Selected Priors

For the credible interval summary a number of, but not all, priors were selected as follows. Whilst no prior gave "good" coverage in all cases, the HCY and HT3 priors seemed most acceptable of those investigated. With hpdv or hpds CrIs they gave "good" coverage for the total variance which was considered most important. They also gave "good" coverage for the batch and portion variances, and for the keg variance, coverage was greater than 0.92. The CrIs for the UNI prior were also compared with those of the HCY and HT3 priors, though the UNI prior did not quite achieve "good" coverage for total variance for the scenarios k2p16_0.5 and k6p16_0.5 for any interval type (but coverage was greater than 0.93 for the hpdv CrI). Excluding the scenario with only 3 batches which has an improper posterior, the FLAT prior with an hpdv CrI type gave "good" coverage for total variance and all the variance components and thus the CrIs are assessed. The CrIs for three other priors available in PROC MIXED are also examined (JEFF, IIN and IG2). The CrIs for the ICCU prior are not plotted as the coverage was

FIGURE 8.1: **Example credible intervals with median visualised on various scales (dashed line = true value = 0.5)**



less than 0.92 for total variance for several scenarios (see Table 7.2). Nor are those of the IG1 and IG3 priors plotted as they are likely to be very similar to the IG2 or JEFF priors.

Plots for the credible interval summary are constructed as now described and are provided in Figures 8.2 to 8.6. The median lower and upper limits of the hpd intervals of all the simulated datasets (together with a horizontal bar for the median of the posterior median) are plotted using icc scales as described in Section 8.1. The credible intervals are summarised for $\sigma_b^2 = 0.5$, 6, 24 for all scenarios for the UNI, HCY, HT3, IIN, IG2, JEFF priors and for all scenarios except for b3k2p16_0.5 (where the posterior is improper) for the FLAT prior. Situations where the estimated coverage is less than 0.94 are identified by an asterisk above the median upper CrI limit. The true values are identified by a black dashed line and, where different, the icc constant is identified by a dotted dark blue line. For variance components hpdv intervals are plotted whilst for total variance both hpdv and hpds intervals are plotted. Figures 8.2 to 8.4 for the individual variance

components were plotted with an icc scale using a normalising constant of 6, the value of the true portion variance which is considered to be the intrinsic variance. Figures 8.5 and 8.6 for the total variance were plotted with an icc scale using normalising constants of 18 and 4.24 ($\sqrt{18}$) respectively (the values of σ^2_{tot} and σ_{tot} if batch, keg and portion variance were all 6). To provide better comparisons between priors for the portion variance, the subplots (for the scenarios) in Figure 8.2 are split into two rows in the figure (with different y-axis scales) according to whether there are 2 (p2) or 16 (p16) portions per keg (where # in the scenario label represents 2 or 16 accordingly). The other graphs are categorised by scenario and true batch variance. The median values from which the graphs are derived are given in Appendix G, Tables G.1 - G.4.

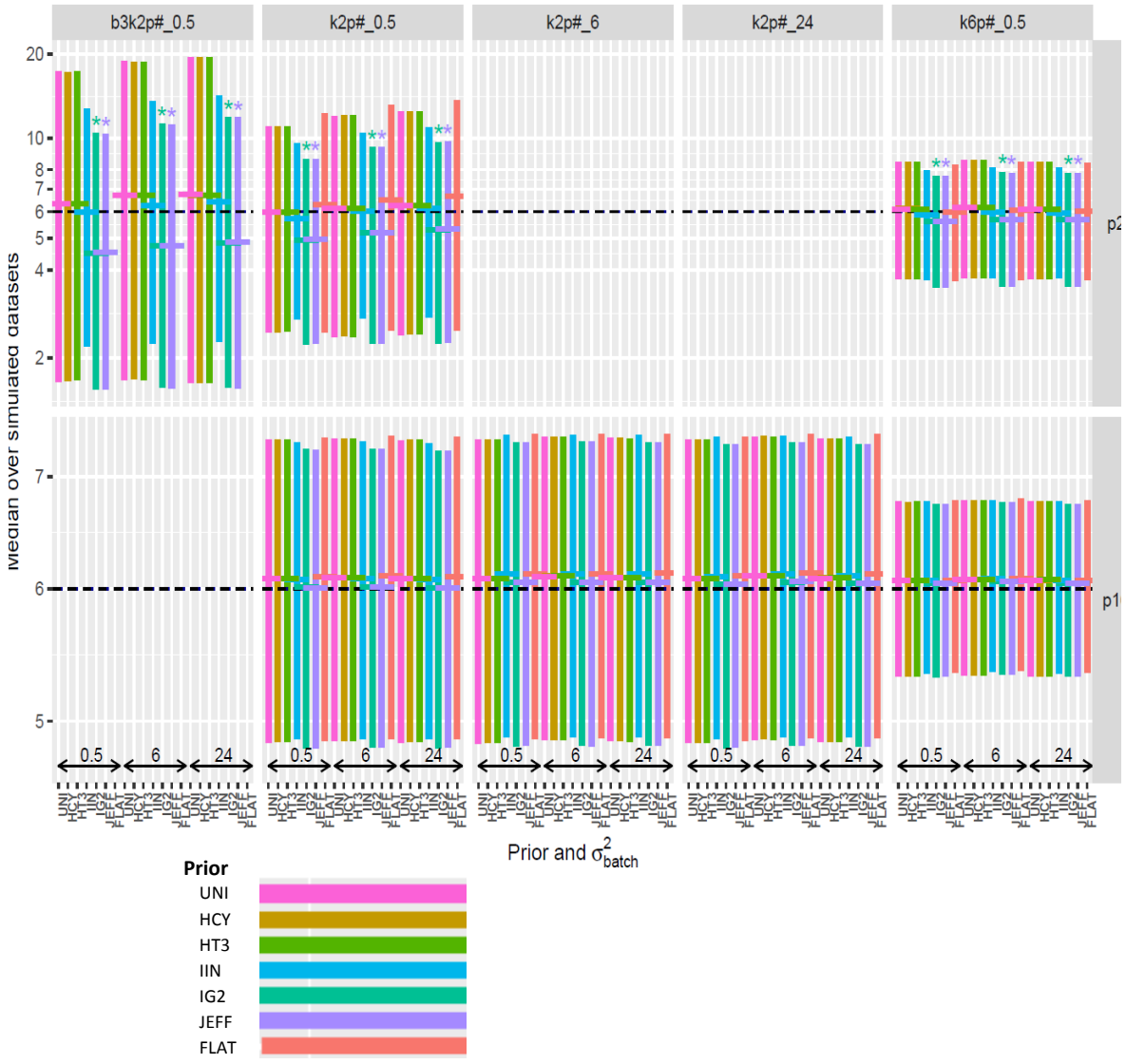
For selected situations where the coverage is poor, the lower and upper CrI limits for the individual simulated datasets are investigated in Section 8.3.

Graphs (similar to Figures 8.2 to 8.6) which additionally include the median hpdl CrIs for the HT3 prior are given in Appendix G, Section G.3. In Section 7.4.1 it was concluded that the use of hpdl CrIs is not considered desirable. Though they provide good coverage for individual variance components, they do not for total variance (see Section 7.3.2) for which hpdv or hpds CrIs would be required and thus would leave a user uncertain which metric to use in other situations e.g. estimating a sum of small variance components. It is seen in Section G.3 that the median hpdl CrIs are higher than those for other priors, especially the upper limit (for example, see Figure G.3). For the individual variance components this resulted in "good" coverage but for total variance the higher lower limit resulted in poor coverage for the smaller designs combined with smaller σ^2_b values. Given the size of the difference seen between the median hpdv and hpdl CrIs, this confirms that a strategy of using both metrics dependent on what is to be estimated is not desirable. In situations where the choice of metric is not clear the CrIs are not likely to be similar. Since the hpdv CrIs for HT3 gave "good" coverage for all variance components and scenarios except k6p16_0.5 for the keg variance, the hpdl CrIs would be unnecessarily giving much higher upper limits in most situations.

General observations

- For the HCY and HT3 priors the median of the CrI limits and posterior median are very similar.
- The UNI prior is more similar to HCY and HT3 priors than the other priors but there are some differences for batch, keg and total variance.
- For the IG2 and JEFF priors the median of the CrI limits and posterior median are very similar. They are also similar for the IIN prior for the batch and keg variance components. The IIN prior was designed to be informative for the portion variance and this is reflected in that it has lower median upper CrI limit, and similar or higher median lower CrI limit, than other priors also having "good" coverage

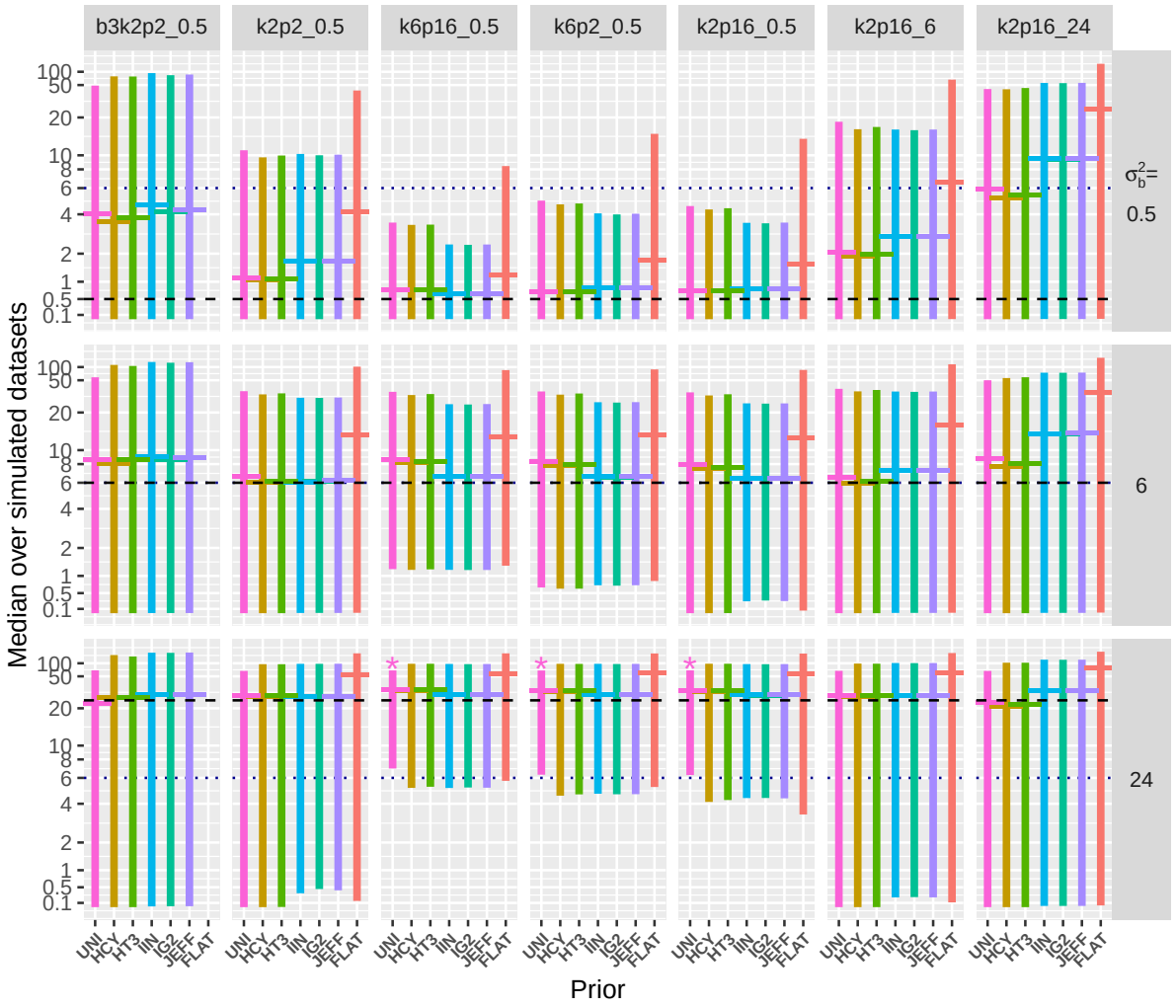
FIGURE 8.2: **Limits of hpdv CrIs (and posterior median) for σ_p^2 for selected priors plotted on icc scale (constant=6) for 2 portions/keg (top row) and 16 portions/keg (bottom row). Dashed lines represent the true value of 6.**



(> 0.94). Note the median CrI limits and posterior median were above those of the IG2 and JEFF priors which had poor coverage. For total variance the IIN prior also has similar medians of the CrI limits and posterior median to the IG2 and JEFF priors except for the scenarios with the two smallest designs and σ_b^2 = 0.5 or 6. Here the median CrI limits and posterior median are slightly higher which probably reflects the higher values seen for the portion variance.

- The FLAT prior gives higher medians of the upper CrI limit and posterior median for the keg, batch and total variance and SD than the other priors in all situations and almost all situations for the portion variance. They are considerably higher for total variance and SD and batch variance, even for the largest design (kp16_0.5)

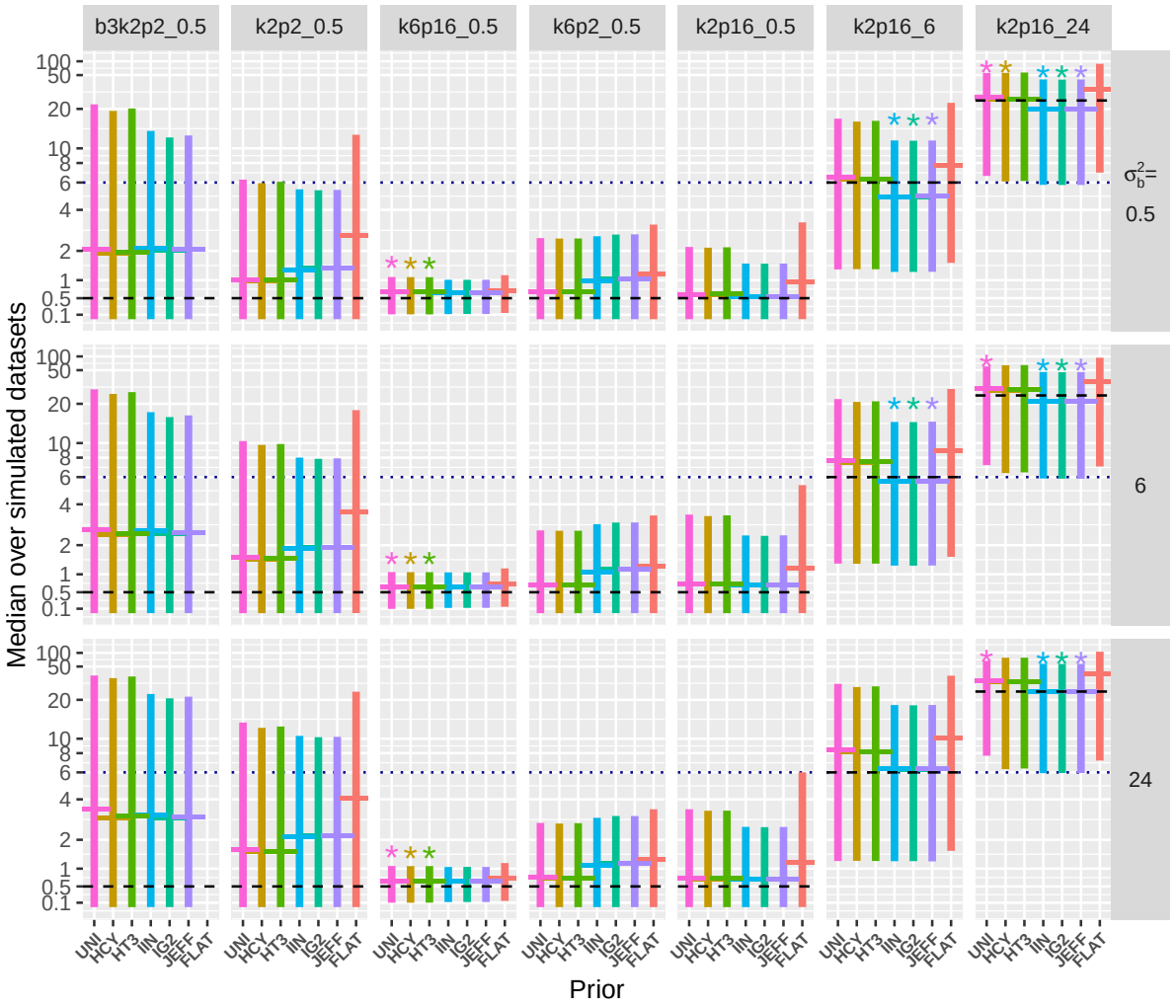
FIGURE 8.3: **Limits of hpdv CrIs (and posterior median) for σ_b^2 for selected priors plotted on icc scale (constant=6). Dashed lines represent the true values.**



and for keg variances for designs with 2 kegs per batch. For example, for scenario k6p16_0.5 and σ_b^2=6 the median of the posterior median and the CrI limits for σ_b^2 are 11.46 (1.34 , 82.0) for the FLAT prior compared to 7.59 (1.21 , 31.5) for the HT3 prior. For scenario k2p2_0.5 and $\sigma_b^2 = 0.5$ the coverage was less than 0.94 for total SD (but not total variance) where the true value was too frequently below the lower limit. For total variance the median of the lower CrI limit was also slightly higher than the other priors, except for the UNI prior for $\sigma_b^2 = 24$ but there was not a consistent pattern for the variance components. Note the median CrI limits were not plotted for b3k2p2_0.5 as the posterior is improper but were usually higher than other priors e.g. (lower=2.1,upper=28.1 for σ_p^2 when σ_b^2=6)

A more detailed comparison of the priors is now made for each variance separately.

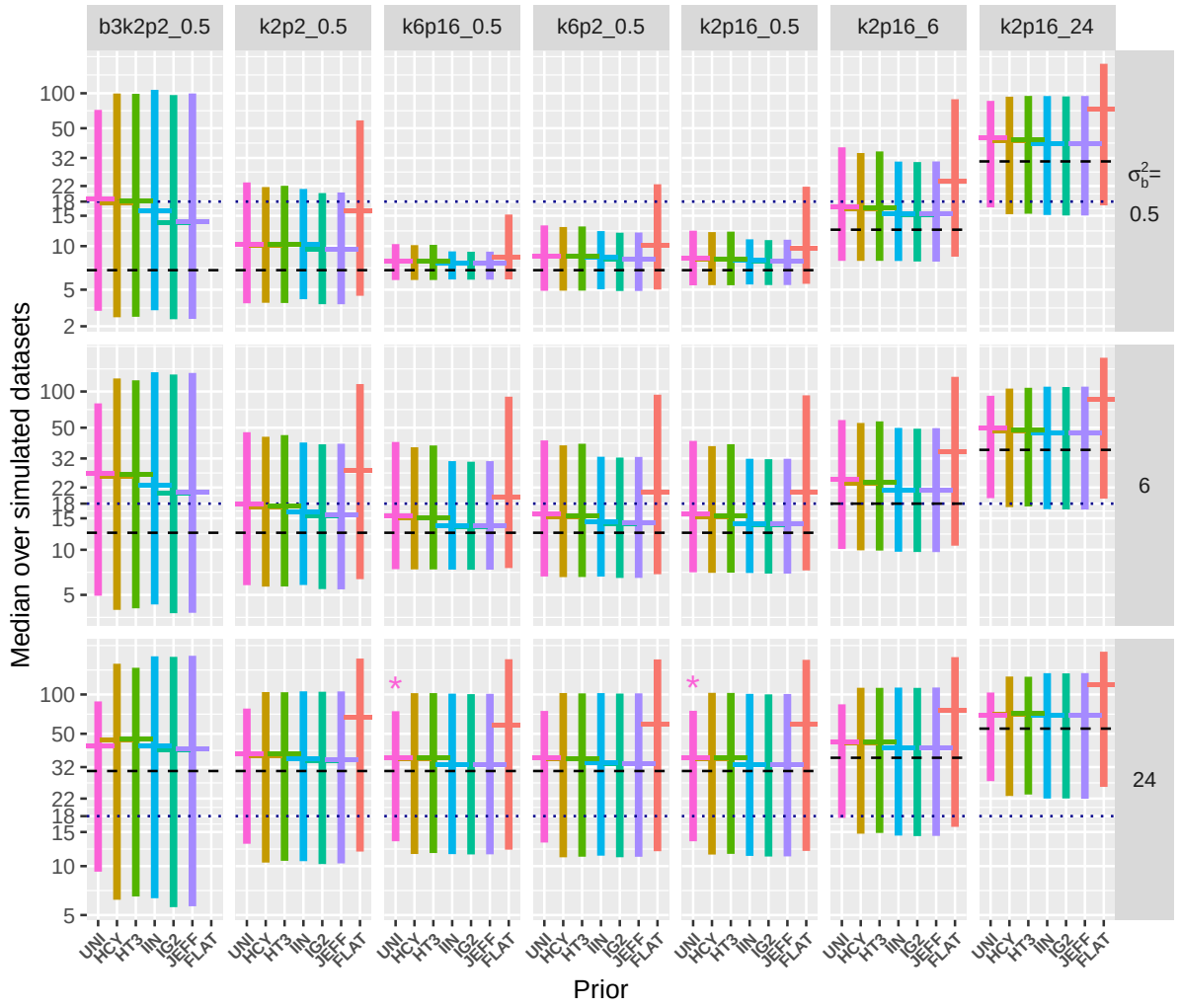
FIGURE 8.4: **Limits of hpdv CrIs (and posterior median) for σ_k^2 for selected priors plotted on icc scale (constant=6). Dashed lines represent the true values.**



Portion Variance In Figure 8.2 it is seen that for the scenarios with 16 portions per keg (p16) the priors give fairly similar results - perhaps as expected given the amount of data to estimate this variance component. The subtle differences between priors follow similar patterns to the larger differences seen for scenarios with only 2 portions per keg and thus are discussed in that context. The median of the posterior medians are generally above the true value for all priors. For the scenarios with 2 portions per keg (p2) the following observations are made:

- The coverage for priors IG2 and JEFF is less than 0.94 for the scenarios with only 2 portions per keg (as was seen in Table 7.3). In Figure 7.17 it was seen that for all three σ_b^2 values for those scenarios, when the true value was outside the CrI limits it was usually above the upper CrI limit. It is seen in Figure 8.2 that the median upper limit of the CrI is considerably lower for IG2 and JEFF prior compared to

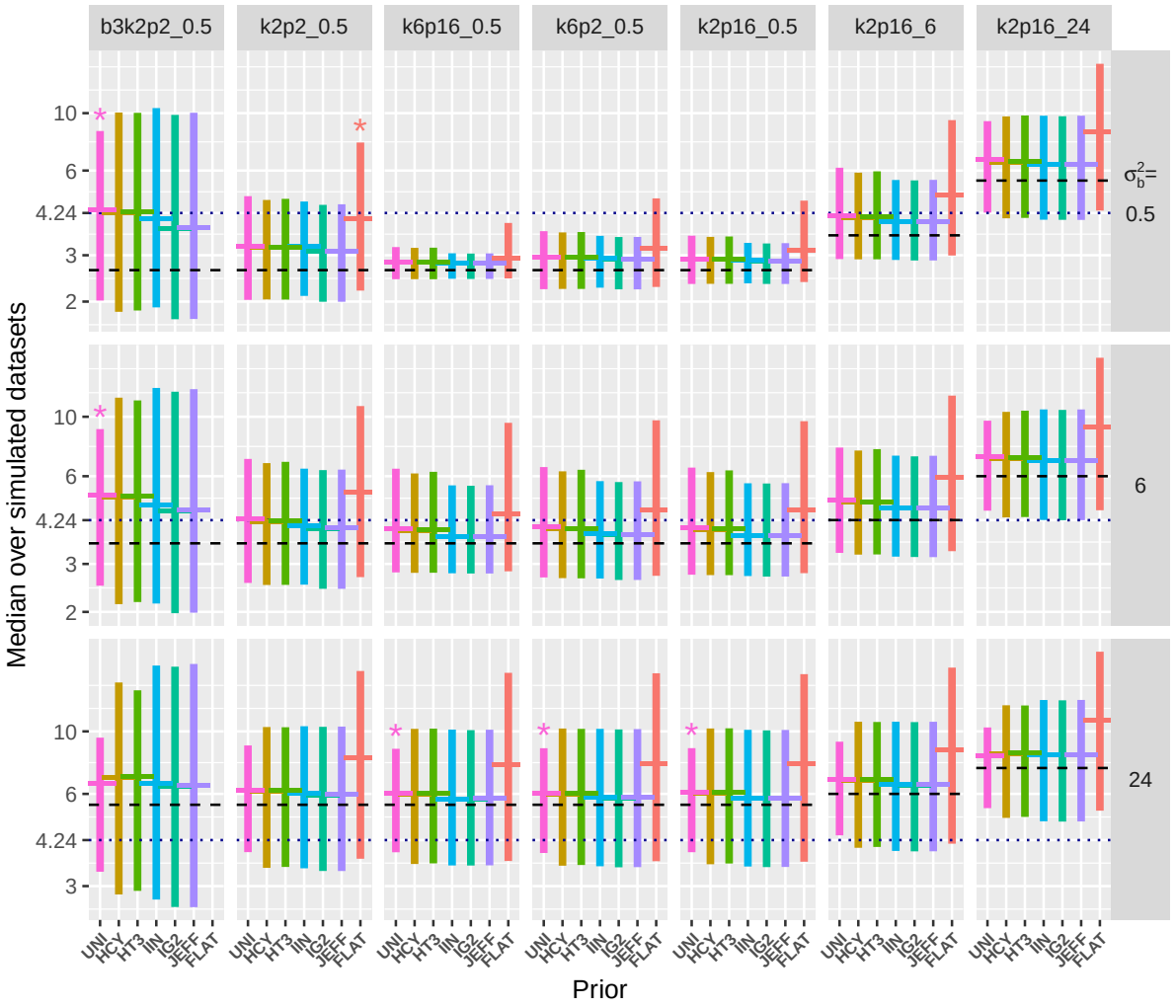
FIGURE 8.5: **Limits of hpdv CrIs (and posterior median) for σ^2_{tot} for selected priors plotted on icc scale (constant=18). Dashed lines represent the true values.**



that of the UNI, HCY and HT3 priors. The median of the posterior median is similarly lower (especially for the smaller designs) compared to those priors and also compared to the true value. The median lower limits are also below those of the other priors. For example, for scenario k2p2_0.5 and σ^2_b=6 the medians of the posterior median (CrI limits) are 4.97 (2.27,9.37) for the JEFF prior, compared to 5.87 (2.40,11.86) for the HT3 prior.

Batch Variance Figure 8.3 uses an icc scale with a constant of 6. As an illustration of how the icc scale can be used when there is not an intrinsic variance of interest, Figure G.5 in Appendix G shows the icc scale where the constant is the true batch variance.

FIGURE 8.6: **Limits of hpds CrIs (and posterior median) for σ^2_{tot} for selected priors plotted on icc scale (constant=4.24). Dashed lines represent the true values.**



The following observations are made:

- The medians of the lower CrI limit for the IIN, IG2 and JEFF priors are similar to, or in some situations higher than, those of the HCY and HT3 priors. However, the relative positions of the medians of the upper CrI limit and posterior median compared to the HCY and HT3 priors is very situation dependent.
- The UNI prior is fairly similar to the HCY and HT3 priors for true $\sigma^2_b = 0.5$ and 6, though there are small differences e.g. the median upper CrI limit is generally slightly higher but for the smallest design it is lower. However, for true $\sigma^2_b = 24$ there are larger differences. The median upper CrI limit is less than those for the other priors and for scenarios with more data and $\sigma^2_k = 0.5$ (k6p16_0.5, k6p2_0.5, k2p16_0.5) the median lower CrI limit is higher than those for the other priors.

For these three scenarios the shorter CrIs result in a coverage which is less than 0.94.

- The effect of the size of σ_k^2 on the estimation of σ_b^2 is seen by comparing scenarios k2p16_0.5, k2p16_6 and k2p16_24. As σ_k^2 increases the median upper CrI limits increase, especially for $\sigma_b^2 = 0.5$ and 6 (for σ_b^2 =24 it increases except for the UNI prior). There is also an effect on the median lower CrI limits. For $\sigma_b^2 = 6$ the median lower CrI limits when σ_k^2=0.5 are higher than those when σ_k^2=6 or 24. For $\sigma_b^2 = 24$ as σ_k^2 increases the median lower CrI limits decrease. This could be expected as given the nested design, the larger the keg variance the less well estimated each batch effect will be and hence the less well estimated the batch variance. The effect of the size of σ_k^2 on the median is dependent on the true value of σ_b^2 and in some cases the prior. For $\sigma_b^2 = 0.5$ the median of the posterior median for σ_b^2 is above its true value in all cases and increases as σ_k^2 increases. The extent of this positive bias is such that, whilst for σ_k^2=0.5 the median upper CrI limit for σ_b^2 is below 6 for all but the FLAT prior, for σ_k^2=24 the median of the posterior median is above 6 for 5 of the 7 priors (HCY and HT3 being the exceptions). For the other values of σ_b^2 investigated, the increase in the median of the posterior median above the true value is mainly seen as σ_k^2 increases from 6 to 24 for σ_b^2 =6. For the other situations the changes as σ_k^2 increases are prior dependent.
- The effect of the size of the design can also be seen by comparing scenarios b3k2p2_0.5, k2p2_0.5, k6p2_0.5, k2p16_0.5, k6p16_0.5. Increasing the number of batches from 3 (b3k2p2_0.5) to 6 reduces the median upper CrI limit (except for UNI prior, σ_b^2=24 and scenario k2p16_0.5 where the median is the same) as could be expected and increases the lower limit for σ_b^2= 6 and 24. In all cases for σ_k^2=0.5 the median of the posterior median for σ_b^2 is above the true value and the positive bias does not necessarily reduce with design size e.g. compare scenario k2p2_0.5 with scenario k6p16_0.5 for $\sigma_b^2 = 6$ and 24.

Keg Variance From Figure 8.4 it is seen that:

- The medians of the lower CrI limits are similar across the five priors HCY, HT3, IIN, IG2 and JEFF for most scenarios with some small differences for scenario k6p16_0.5 and scenarios k6p16_6 and k6p16_24. However, the relative positions of the medians of the upper CrI limit and posterior median for the IIN, IG2 and JEFF priors compared to the HCY and HT3 priors are very situation dependent.
- The UNI prior is more similar to the HCY and HT3 priors than the other priors but there are some differences. The median of the upper CrI limit is slightly higher in most situations, except for $\sigma_k^2 = 24$ and $\sigma_b^2 = 6$ or 24 where it is slightly lower.
- Increasing the batch variance has some effect in increasing the median upper CrI limit and posterior median for the keg variance, though the increase is much smaller

than the effect of increasing the keg variance on the upper CrI limit and posterior median for the batch variance. It is to be expected that the batch variance would have less effect on the keg variance than the keg variance on the batch variance as it is a nested design and thus σ_k^2 can be estimated within each batch.

- The effect of the size of the design can also be seen by comparing scenarios k2p16_0.5 with k6p16_0.5, and k2p2_0.5 with k6p2_0.5. Increasing the number of kegs from 2 per batch to 6 per batch for the design with 2 portions per keg provides fairly similar median lower CrI limits, but smaller median upper CrI limits and posterior median, though the median posterior medians are still above the true values. Increasing the number of kegs from 2 per batch to 6 per batch for the design with 16 portions per keg provides increased median lower CrI limits and decreased median upper CrI limits. For $\sigma_b^2 = 6$ or 24 the median of the posterior median is closer to the true value but this is not necessarily the case for $\sigma_b^2 = 0.5$.
- In Figure 8.4 three of the scenarios had some priors with "poor" coverage (< 0.94).
  - For scenario k2p16_6 the IIN, IG2 and JEFF priors had poor coverage for $\sigma_b^2 = 0.5$ and 6. It is seen that the median CrIs are below those of the other priors especially at the upper limit.
  - For scenario k2p16_24 the IIN, IG2 and JEFF priors had poor coverage for all three values of σ_b^2 and the median CrIs are below those of the other priors especially at the upper limit, above which the true value often laid (see Figure 7.13). The UNI prior also had poor coverage for all three values of σ_b^2. However here, for $\sigma_b^2 = 6$ and 24 the median lower CrI limit was above those of the other priors as well as the median upper CrI limit being below those of the HCY and HT3 priors i.e. the intervals seem too short. In addition, the HCY prior had poor coverage for $\sigma_b^2 = 0.5$, the median CrI limits (6.10 , 54.6) being similar but slightly lower than those of the HT3 prior (6.13 , 55.5).
  - For scenario k6p16_0.5 the UNI, HCY and HT3 priors had coverage less than 0.94 for all three values of σ_b^2 though the median upper CrI limits were higher than those for the IIN, IG2 and JEFF priors and the median lower CrI limits were lower than those for the IIN, IG2 and JEFF priors. This will be investigated further in Section 8.3.1.

Total Variance It is seen from Figures 8.5 and 8.6 that:

- Over all scenarios and σ_b^2 values the median of the posterior median overestimates the true value. As the design size increases for $\sigma_k^2 = 0.5$ the median becomes closer to the true value.
- The differences between the median CrI limits for IG2 and JEFF priors compared to HCY and HT3 priors were small and varied according to the situation. For example for $\sigma_b^2 = 6$ and scenarios k6p16_0.5, k6p2_0.5 and k2p16_0.5 the median

upper CrI limits for IG2 and JEFF priors are lower than those for HCY and HT3 priors for total variance and total SD (and the median lower CrI limits are similar), whereas for scenario k2p16_24 the median CrIs for IG2 and JEFF priors are wider than those for HCY and HT3 priors for $\sigma_b^2 = 6$ and 24 for total variance and SD. The median of the posterior medians were similar but those of the IG2 and JEFF priors were generally slightly closer to the true value.

- Though the median CrI limits for the UNI prior were more similar to the HCY and HT3 priors than the other priors for the individual variance components there were some small differences. For total variance in many cases the median CrI limits for the UNI prior were again similar to the HCY and HT3 priors. However, there were some situations where the median credible intervals were much shorter, the b3k2p2_0.5 scenario, scenario k2p16_24 and all scenarios where $\sigma_b^2 = 24$. Coverage was less than 0.94 for two of these situations for total variance and for five of these situations for total SD.

8.3 Investigation of Individual Credible Intervals for Situations where Coverage is Low

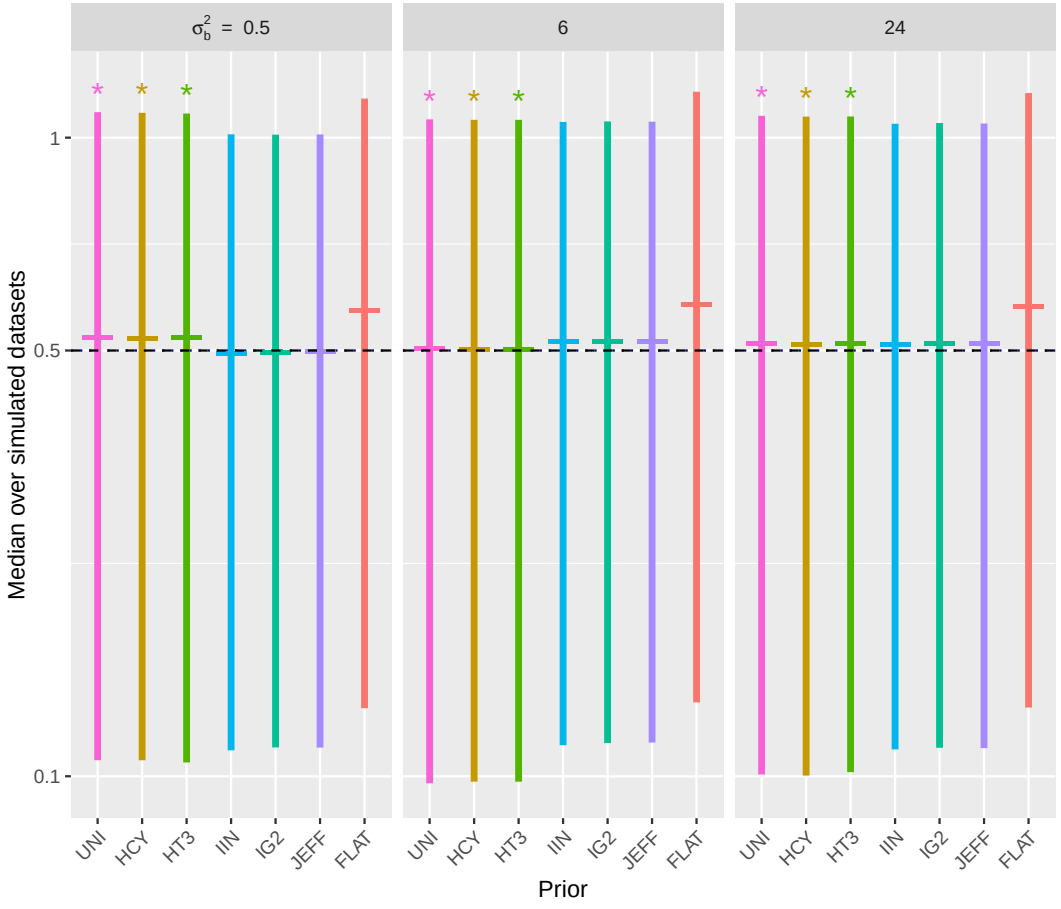
8.3.1 Scenario k6p16_0.5 - Keg Variance

In Figure 8.4 it was seen for scenario k6p16_0.5 that, despite the coverage of the hpdv CrIs for UNI, HCY and HT3 priors being less than 0.94 and the coverage of the hpdv CrIs for IIN, IG2 and JEFF priors being greater than 0.94, the median CrIs seemed wider for the former priors compared to those for the latter. This at first seems counter-intuitive. Figure 8.7 shows a closer view of the median CrI limits for the keg variance for scenario k6p16_0.5 than that provided in Figure 8.4 and confirms those observations for all three values of σ_b^2.

To investigate this further, the CrI limits and median for individual datasets are compared for the HCY, HT3, UNI and JEFF priors. Figure 8.8 shows the lower limits and Figure 8.9 shows the upper limits for HCY, UNI and JEFF priors plotted against the HT3 prior. In each plot the 2.5, 50 and 97.5 percentiles for the statistic are also plotted, together with the line of equality. Dotted grey lines indicate the true keg variance value of 0.5.

In Figures 8.8 and 8.9 the similarity of the CrI limits between the HCY and UNI priors with the HT3 prior is seen. The main observation being that when the lower CrI limit is at or very close to zero is observed for one of these three priors it does not necessarily have a value at or very close to zero on another. An expanded view of this is seen in Figure 8.10. The reason for this will be explored in Section 8.4.1 but as the values are

FIGURE 8.7: **Limits of hpdv CrIs (and posterior median) for σ_k^2 for selected priors plotted on icc scale (constant=0.5) for scenario k6p16_0.5. Dashed lines represent the true values.**



far below the true value of 0.5 it does not affect whether the true value is above the CrI lower limit. It is also noted that the JEFF prior has fewer values at or very close to zero compared to the other priors. This may be a consequence of the smaller lower CrI limits for the JEFF prior often being above those for the HT3 prior, but it could be that the fewer very small values is due to the type of MCMC sampling used.

FIGURE 8.8: **Lower limit of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ_k^2 for scenario k6p16_0.5.**

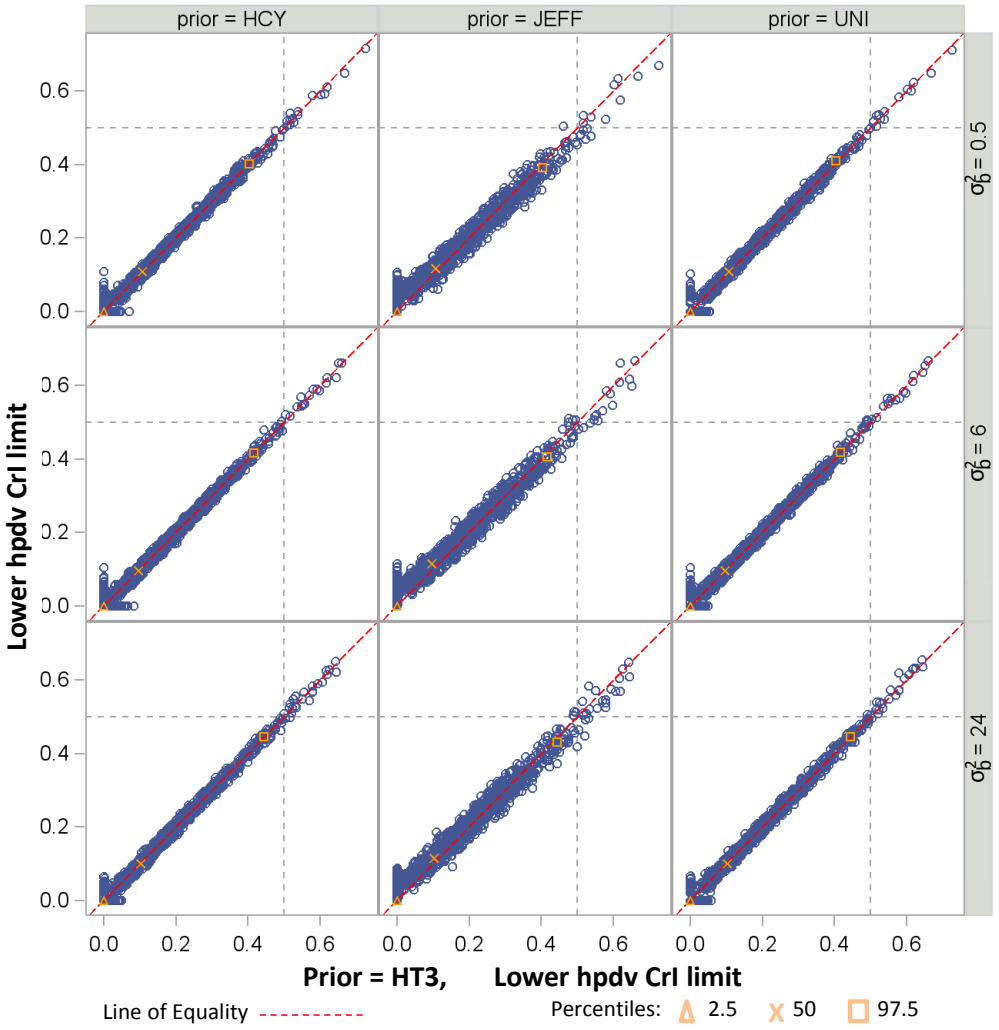


FIGURE 8.9: **Upper limit of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ_k^2 for scenario k6p16_0.5.**

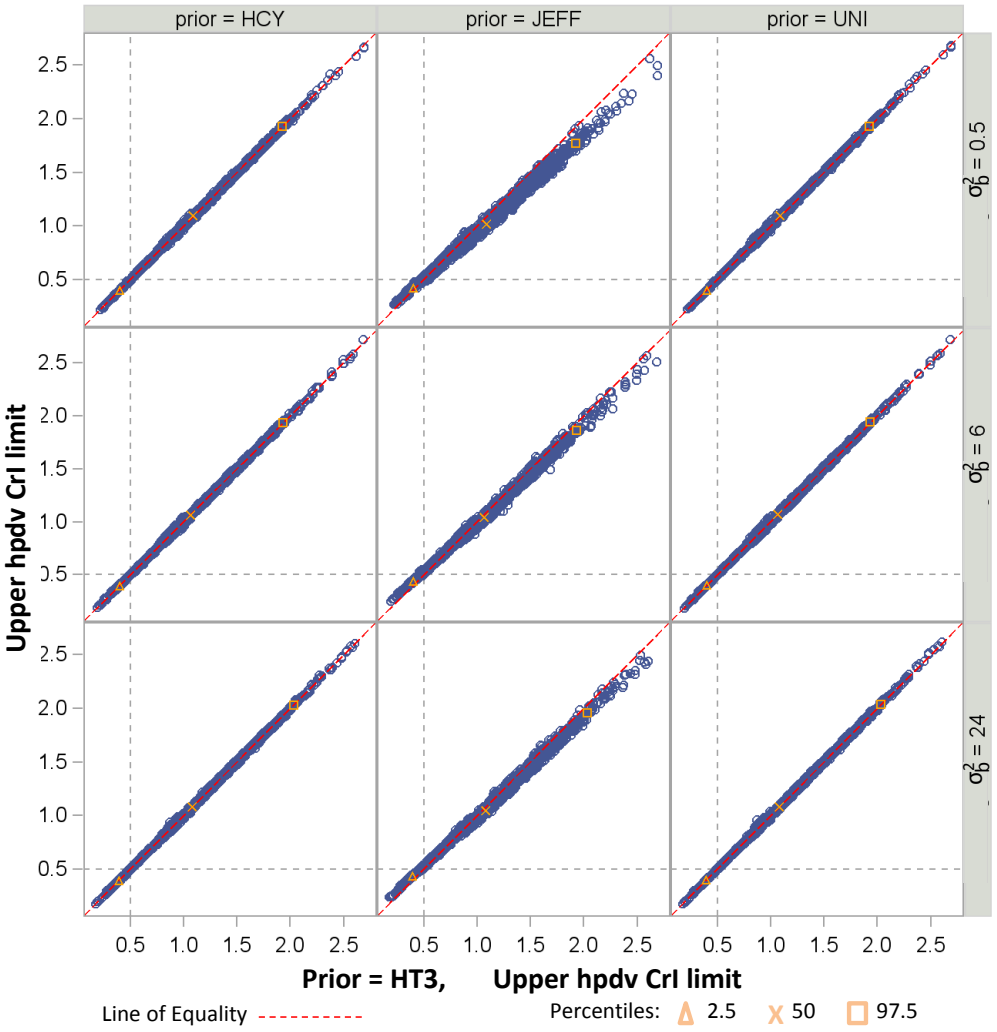
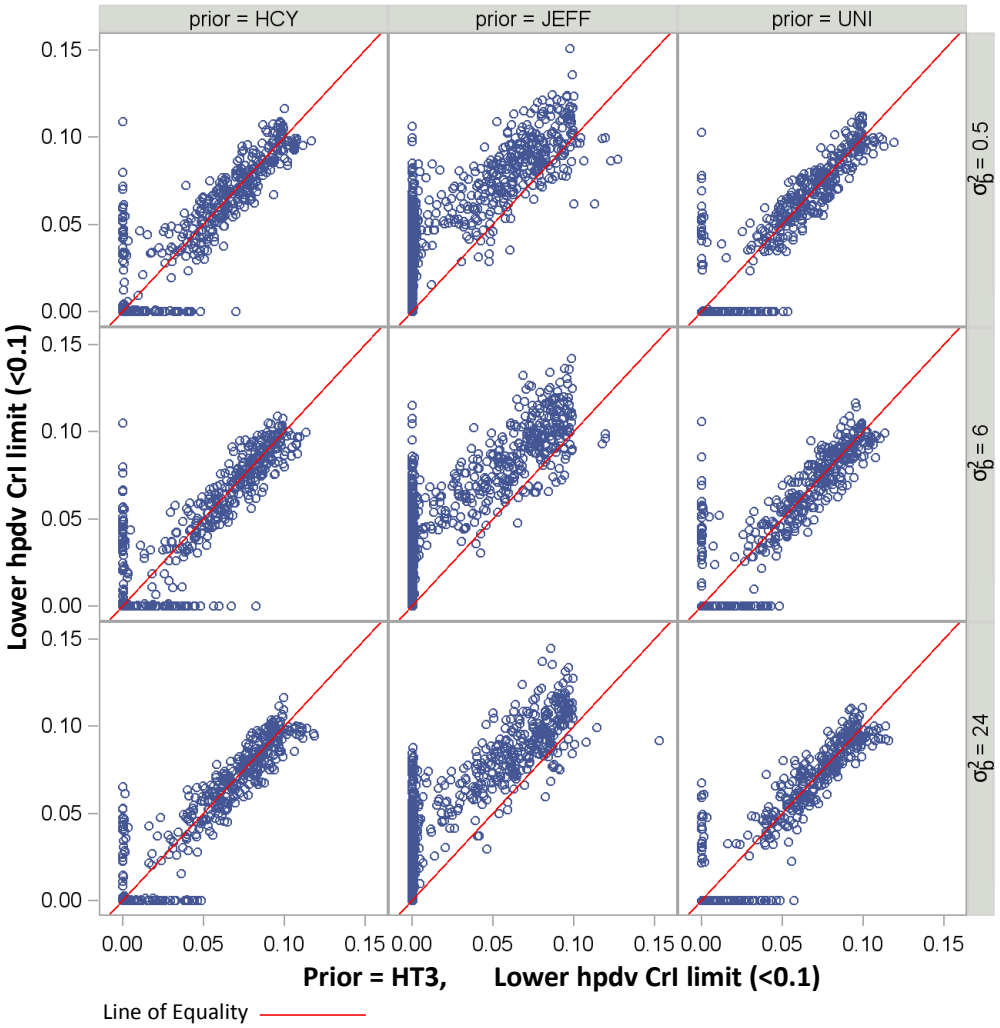


FIGURE 8.10: **Figure 8.8 with values restricted to < 0.1 for either prior in a sub plot**

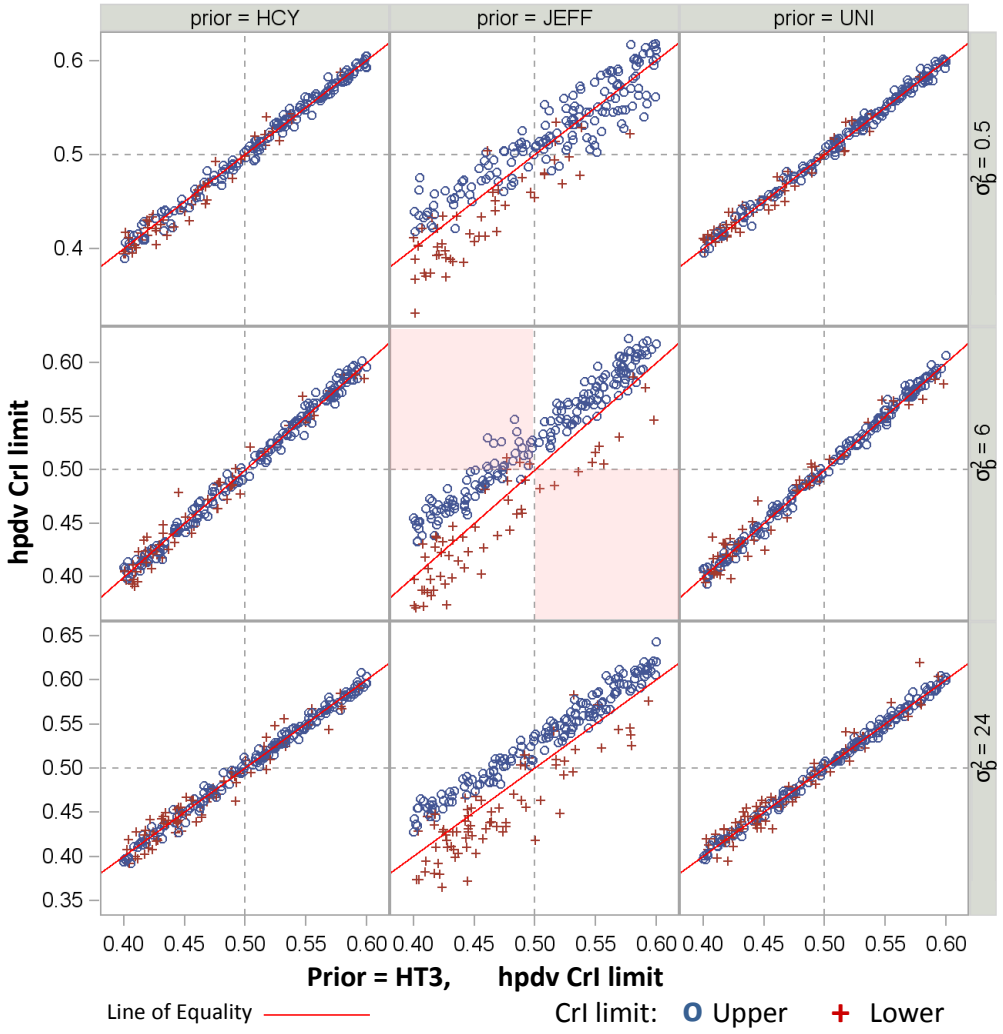


It is seen that there are some small systematic differences between the CrI limits for the JEFF prior compared with the HT3 prior:

- For the lower CrI limit (Figure 8.8) when the value is small e.g. less than 0.1 the lower limit for JEFF prior is often higher than that for the HT3 prior. However, for the larger values e.g. close to the true value (0.5) the lower limit is often lower than that for the HT3 prior.
- For the upper CrI limit (Figure 8.9) when the value is small e.g. close to the true value (0.5) the upper limit for JEFF prior is often higher than that for the HT3 prior, especially for $\sigma_b^2 = 6$ and 24. However, for the larger values the upper limit is often lower than that for the HT3 prior.

The median CrIs for the JEFF prior are shorter than those of the HT3 prior because the median CrI limits lie in the range of values where the JEFF prior has higher lower limits, and lower upper limits, than the HT3 prior. However, in the range of values around 0.5, which are critical to the estimation of the coverage of the interval, the CrIs for the JEFF prior are longer than those of the HT3 prior and hence more often include the true value. An expanded view is shown in Figure 8.11 where the CrI limits within 0.1 of the true value (0.5) for the HT3 prior are compared with those of the HCY, JEFF and UNI priors. To illustrate the difference between the JEFF prior and the HT3 prior, the upper CrI limits are discussed for σ_b^2=6. The upper left square (shaded) shows many upper CrI limits (blue circles) which are above 0.5 for the JEFF prior and below 0.5 for the HT3 prior but there are none in the lower right square (shaded) where they would be above 0.5 for the HT3 prior and below 0.5 for the JEFF prior. The CrIs for the JEFF prior being shorter than the HT3 prior for higher values suggests that i) there may not be the same difference between the JEFF prior and HT3 prior for CrIs with smaller nominal coverage, ii) though the JEFF prior gives better coverage for σ_k^2=0.5 it may not do for higher true values. For i) this was investigated for target coverages of 80% and above (those likely to be most useful) and it was found that the estimated coverages for the JEFF prior were still slightly higher than those for the HT3 prior. Intervals with smaller target coverages were not explored but given the differences between the median CrI limits it is expected that for a small enough target coverage the coverage of the HT3 prior would exceed that of the JEFF prior. For ii) it was seen for the design with 6 batches, 2 kegs/batch and 16 portions/keg (see Figure 8.4) where the JEFF prior (and IIN and IG2 priors which were similar) had good coverage for σ_k^2=0.5 but not for σ_k^2=6 or 24.

FIGURE 8.11: **Limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those between 0.4 and 0.6 for the HT3 prior for σ_k^2 for scenario k6p16_0.5.**



8.3.2 Other Situations

It was seen in Section 8.3.1 that, for the situation explored, because the relative differences in the CrIs between priors differed according to the magnitude of the CrI limits, the median CrI limits did not necessarily help explain the differences in coverage between priors. Hence some investigations were performed for the following situations where poor coverage also occurred:

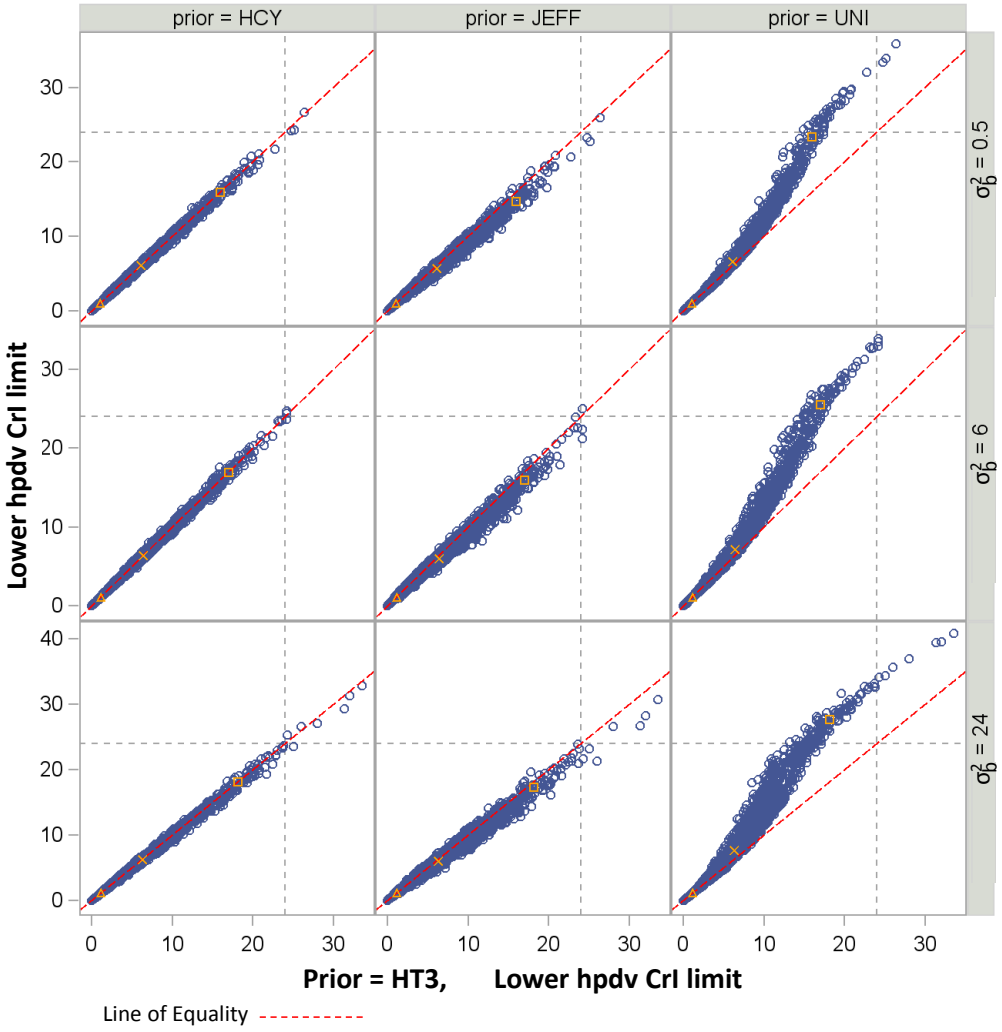
- Keg variance: The IIN, IG2 and JEFF priors have poor coverage for σ_k^2 for scenarios k2p16_6 and k2p16_24 for σ_b^2=0.5 and 6, and scenario k2p16_24 for σ_b^2=24. The UNI prior had poor coverage for scenario k2p16_24 for all σ_b^2 values. The HCY prior had poor coverage for scenario k2p16_24 for σ_b^2=0.5. Scenario k2p16_24 is explored in Figures 8.12 - 8.14.
- Batch variance: The UNI prior has poor coverage for σ_b^2 for scenarios k6p16_0.5, k6p2_0.5 and k2p16_0.5 for σ_b^2=24. Scenario k6p16_0.5 is explored in Figures 8.15 - 8.16.
- Total variance: The UNI prior has poor coverage for σ_{tot}^2 for scenarios k6p16_0.5 and k2p16_0.5 for σ_b^2=24. Scenario k6p16_0.5 is explored in Figures 8.17 - 8.18.

For the JEFF prior for the keg variance and scenario k2p16_24 it was seen in Figure 7.13 that the poor coverage was due to the true value for σ_k^2 being too often above the CrI. Figure 8.13 shows that the JEFF prior gives smaller upper limits than the HT3 prior not only for upper limit values close to the true value of 24, but also for the range of values for the upper limits. Figure 8.14 shows an expanded view around the true value of σ_k^2=24. An example of the difference is that when the HT3 prior give an upper CrI limit of 25 the JEFF prior gives a value around 15. The difference is seen across the range of values, unlike the HT3 prior for the scenario in Section 8.3.1 which had poor coverage due to its behaviour for values around the true value.

For the HCY prior it is seen in Figure 8.14 that for scenario k2p16_24 the upper limits for the keg variance are slightly lower than those of the HT3 prior accounting for the slight difference in coverage (0.939 for HCY compared to 0.944 for HT3).

For the UNI prior for the keg variance for scenario k2p16_24, the lower limit is similar to HT3 for the smaller values but for values above the median it is almost always higher, and for values close to the true value of 24 considerably higher. In the region of the true value of 24 it is seen that the upper limit is just slightly higher than the HT3 prior, but for values above the median it is almost always lower and flattens to an asymptote - a result of the prior having no support above 75. For the batch variance (Figures 8.15 and 8.16) and total variance (Figures 8.17 and 8.18) for scenario k6p16_0.5 and σ_b^2=24 a similar pattern is seen.

FIGURE 8.12: **Lower limit of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ_k^2 for scenario k2p16_24.**



Thus, whilst not all situations of low coverage were explored, the behaviour of the CrIs causing the poor coverage was usually seen across the range of CrI limit values seen suggesting that CrIs with lower probability would also exhibit low coverage.

FIGURE 8.13: **Upper limit of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ_k^2 for scenario k2p16_24.**

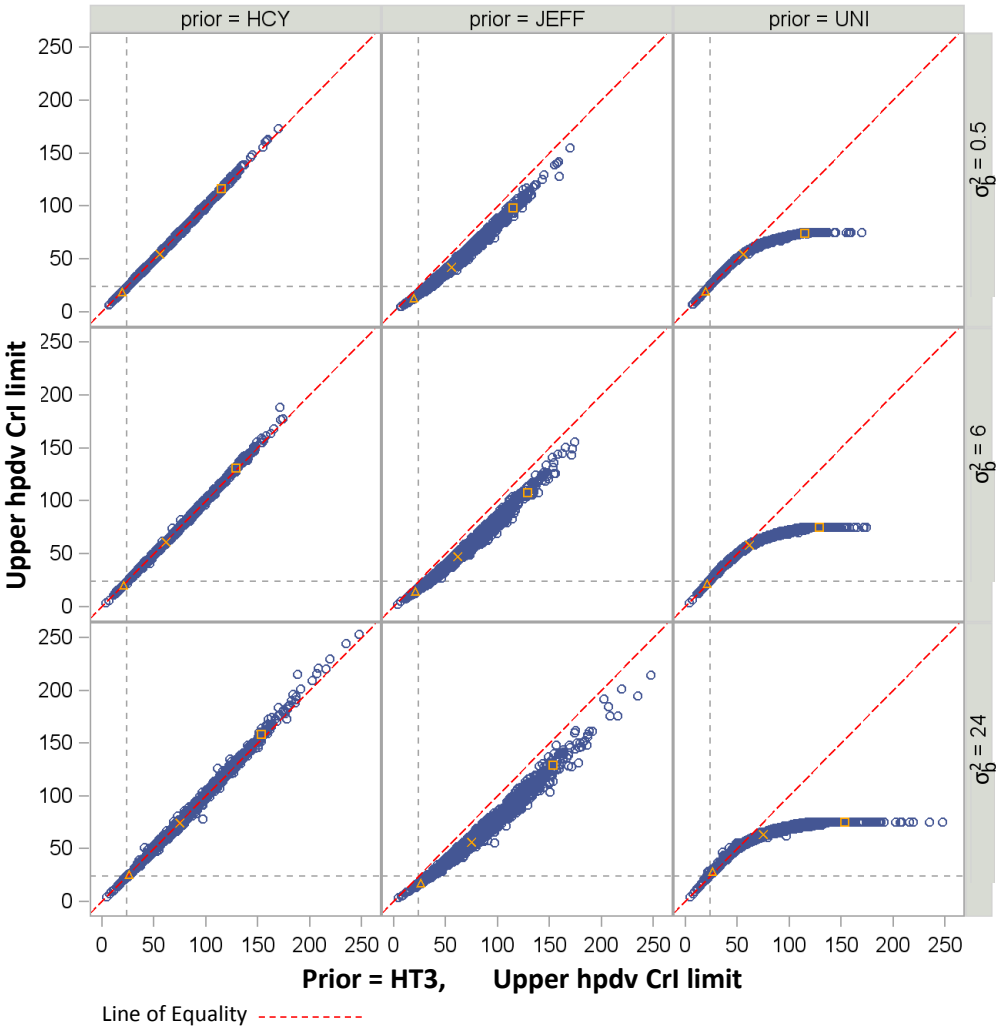


FIGURE 8.14: **Limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those between 14 and 34 for the HT3 prior for σ_k^2 for scenario k2p16_24.**

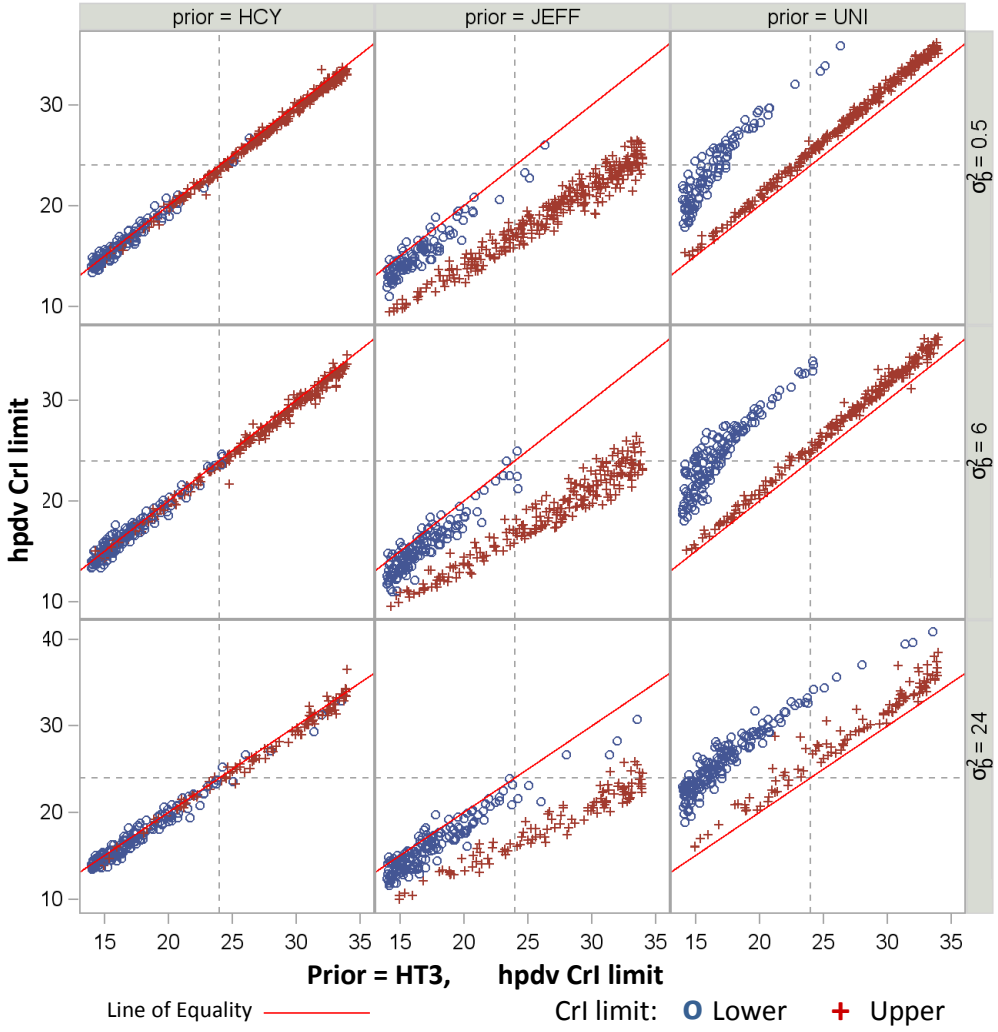


FIGURE 8.15: **Lower and upper limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ_b^2 for scenario k6p16_0.5 and σ_b^2=24.**

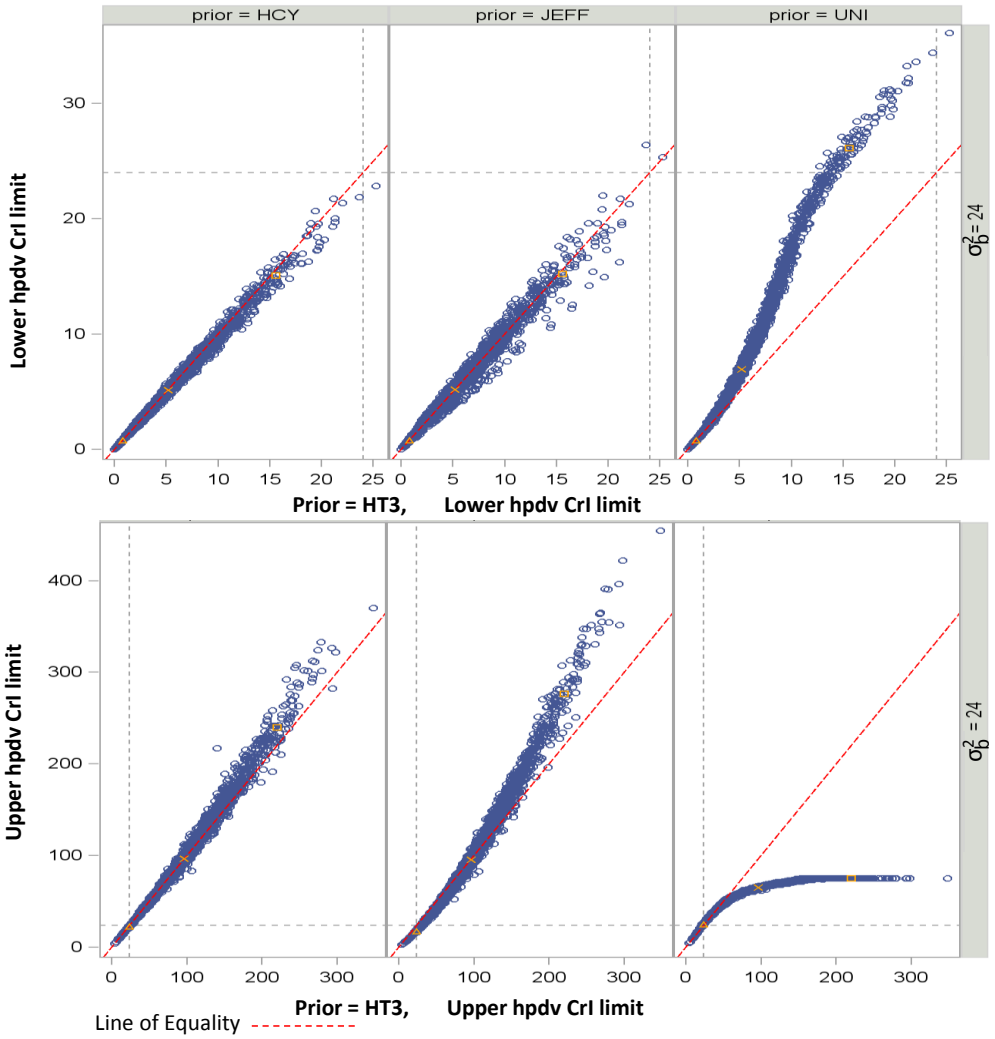


FIGURE 8.16: **Limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those between 14 and 34 for the HT3 prior for σ_b^2 for scenario k6p16_0.5.**

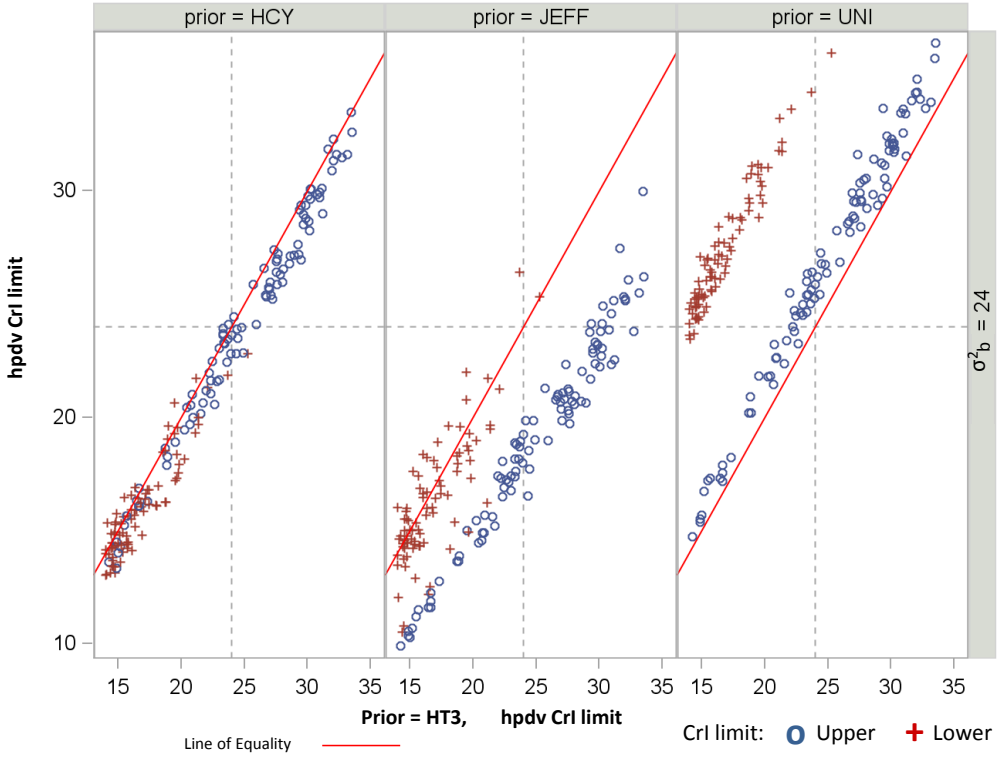


FIGURE 8.17: **Lower and upper limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those for HT3 prior for σ^2_{tot} for scenario k6p16_0.5 for σ^2_b=24.**

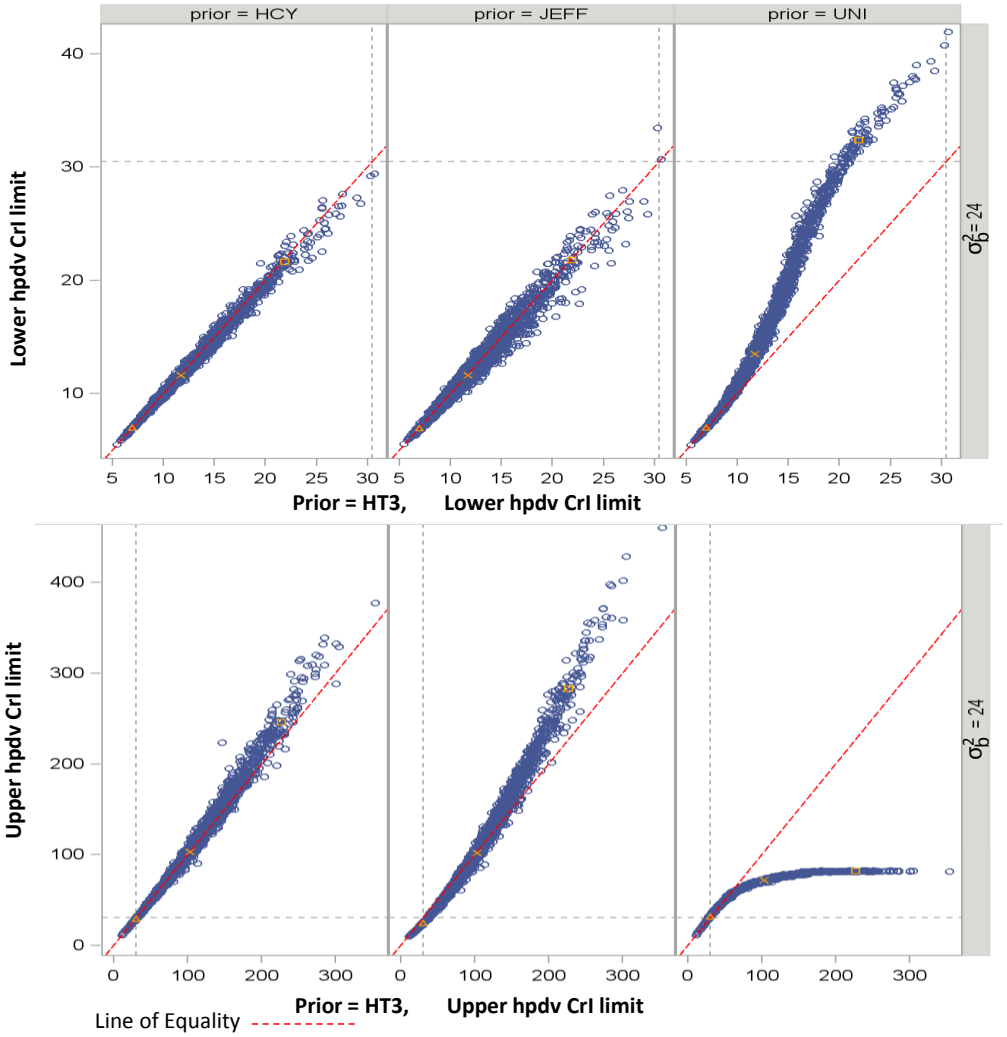
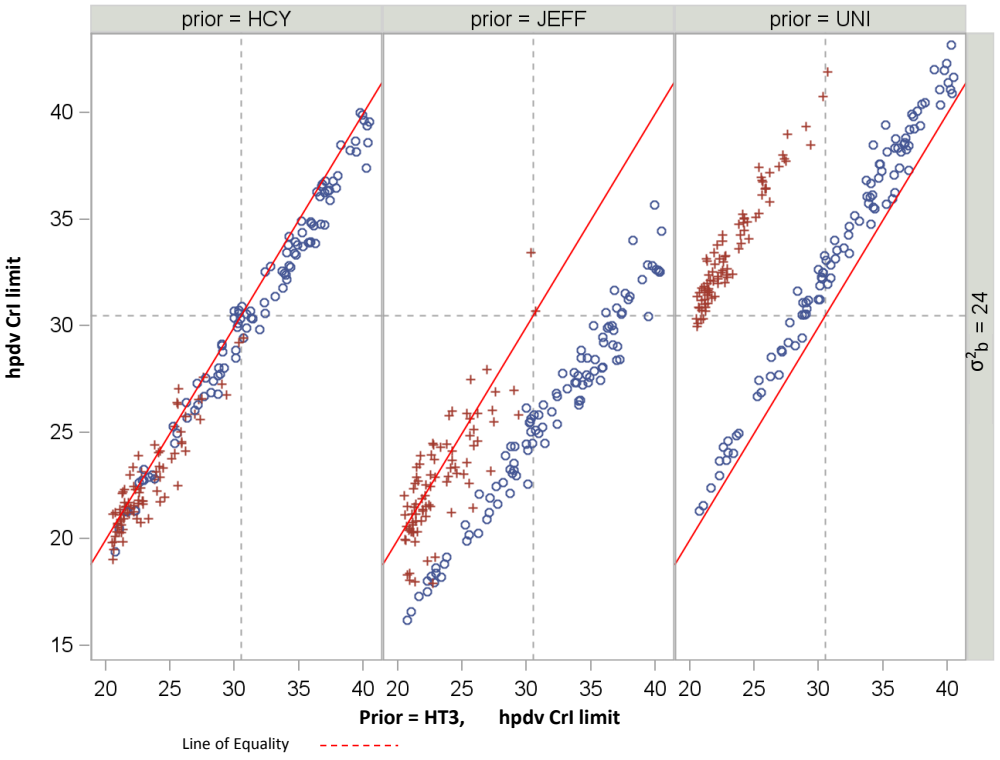


FIGURE 8.18: **Limits of hpdv CrI for HCY, JEFF and UNI priors plotted against those between 20.5 and 40.5 for the HT3 prior for σ^2_{tot} for scenario k6p16_0.5 for σ^2_b=24.**



8.4 Further Investigation of Low Coverage for Scenario k6p16_0.5 for Keg Variance

Given that the HCY and HT3 priors achieved "good" coverage (>0.94) in all situations except for scenario k6p16_0.5 for the keg variance, and no other prior did better overall, the results for this situation are investigated further. In Section 8.4.1 simulation and MCMC aspects are explored to check whether the coverage is truly poor rather than being an artefact of the simulations or MCMC procedure applied. It will be seen that it is truly "poor" (<0.94). In Section 8.4.2 the sensitivity of the coverage to the HT3 prior parameter and other influences is examined.

8.4.1 Simulation and MCMC Aspects

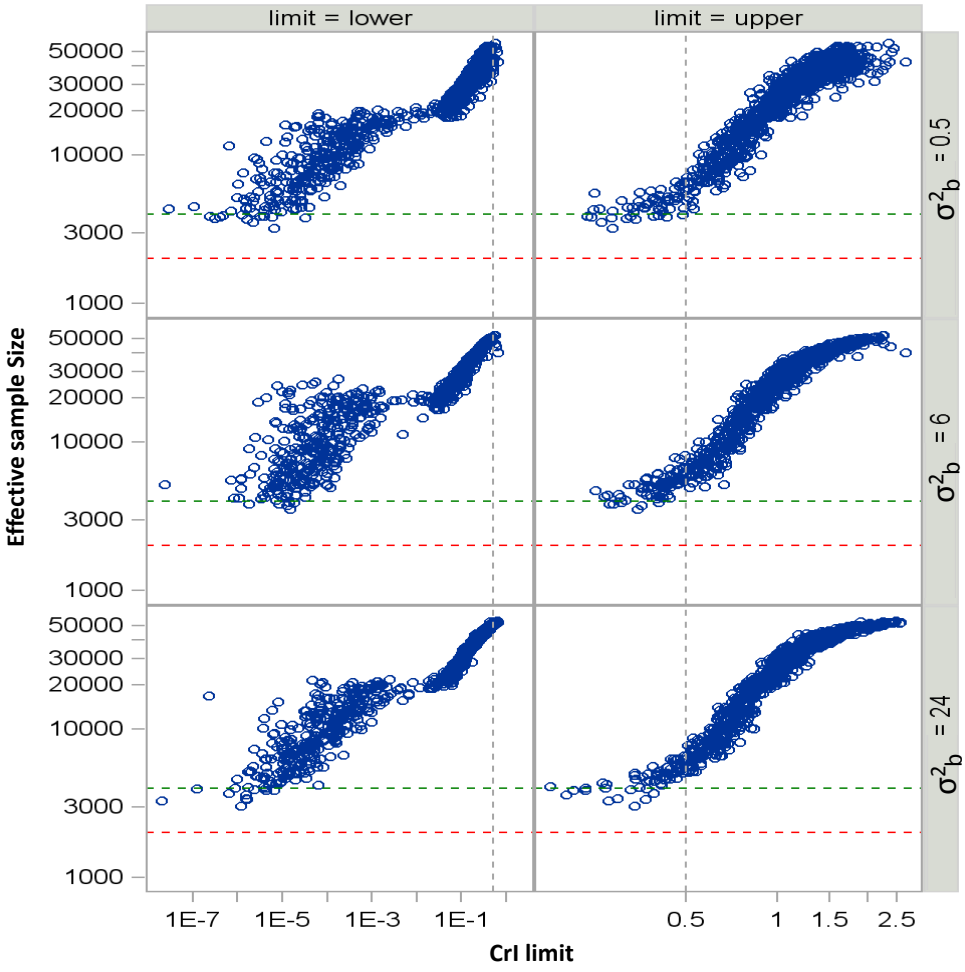
In this section aspects of the simulations performed and the MCMC procedure applied are examined.

Liu et al. (2015) showed for some simple situations that empirical hpd CrIs are biased towards shorter intervals and thus typically have a lower coverage than nominal. The larger the sample size, the smaller the bias will be. Thus the effective sample size is investigated below. In addition, if the effective sample size is too small, the "poor" coverage could be due to random chance. Though this was considered unlikely, since "poor" coverage is seen for all three σ_b^2 values, each of which has its own set of simulated datasets and associated Bayesian analyses.

The effective sample size (ESS), as discussed in Section 6.3.3, was investigated for this situation. The HCY and HT3 priors had effective sample sizes greater than 2000 for all datasets for the scenario k6p16_0.5 for keg variance, with a few datasets with ESS between 2000 and 4000 and the remainder with ESS greater than 4000. The effective sample sizes are plotted against the lower and upper CrI limits in Figure 8.19 (on log scales for clarity). For the lower CrI limits the smaller effective sample sizes are associated with very small lower CrI limits (<<0.5) and thus the estimate of coverage should not be affected. For the upper CrI limits the smaller effective sample sizes are associated with the small upper CrI limits some of which are in the region of the true value. Thus, though the desired effective sample size had been met, it is possible that the estimate of coverage could be affected and it was considered worth eliminating this as a possibility.

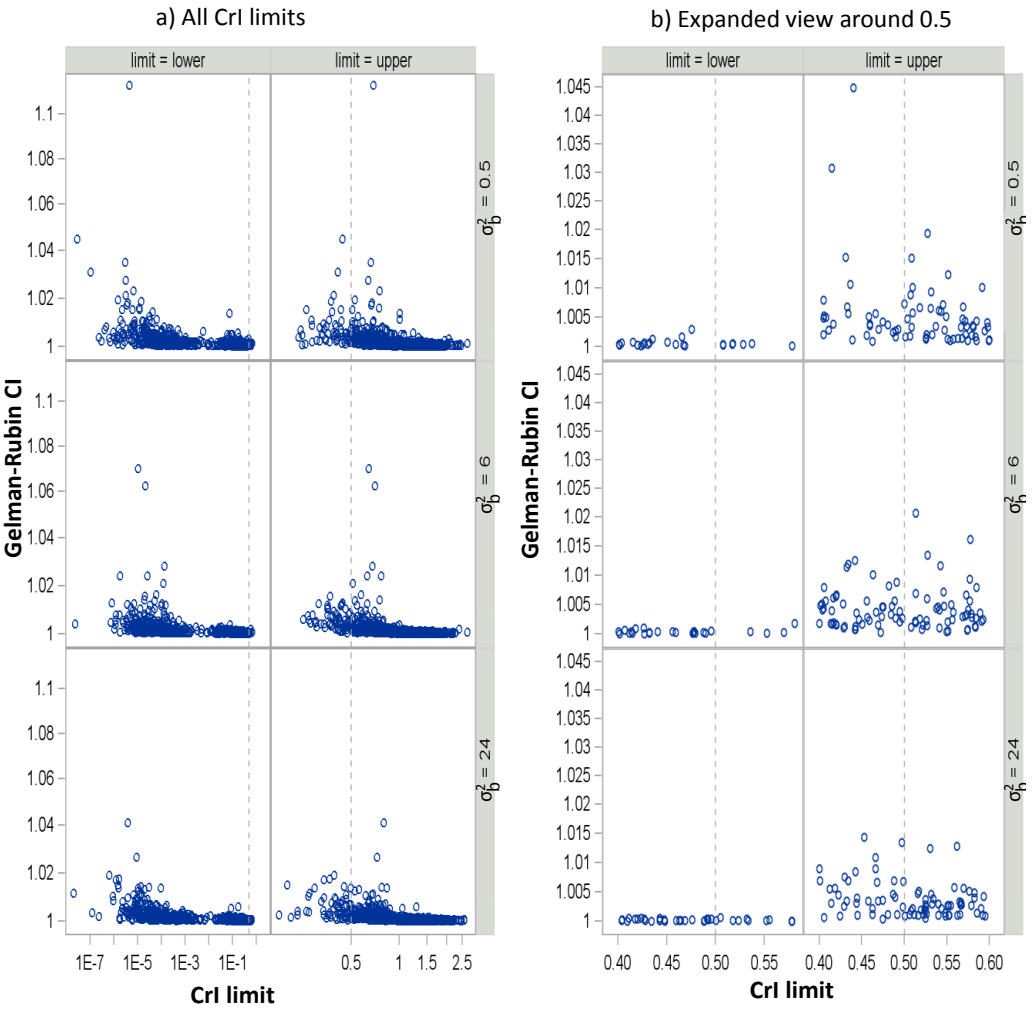
In addition to ESS the Gelman-Rubin CI (G-R CI) convergence statistic described in Section 6.3.2 is examined. In Figure 8.20, which plots the G-R CI statistic for the variance metric against the CrI limit, there are just three datasets with G-R CI greater than 1.05, and all but one are less than 1.08. For the lower CrI limit there are none close to the true value with higher G-R CI values. Though for the upper CrI limit the higher

FIGURE 8.19: **ESS of lower and upper hpdv CrI limits for HT3 prior for σ_k^2 for scenario k6p16_0.5.**



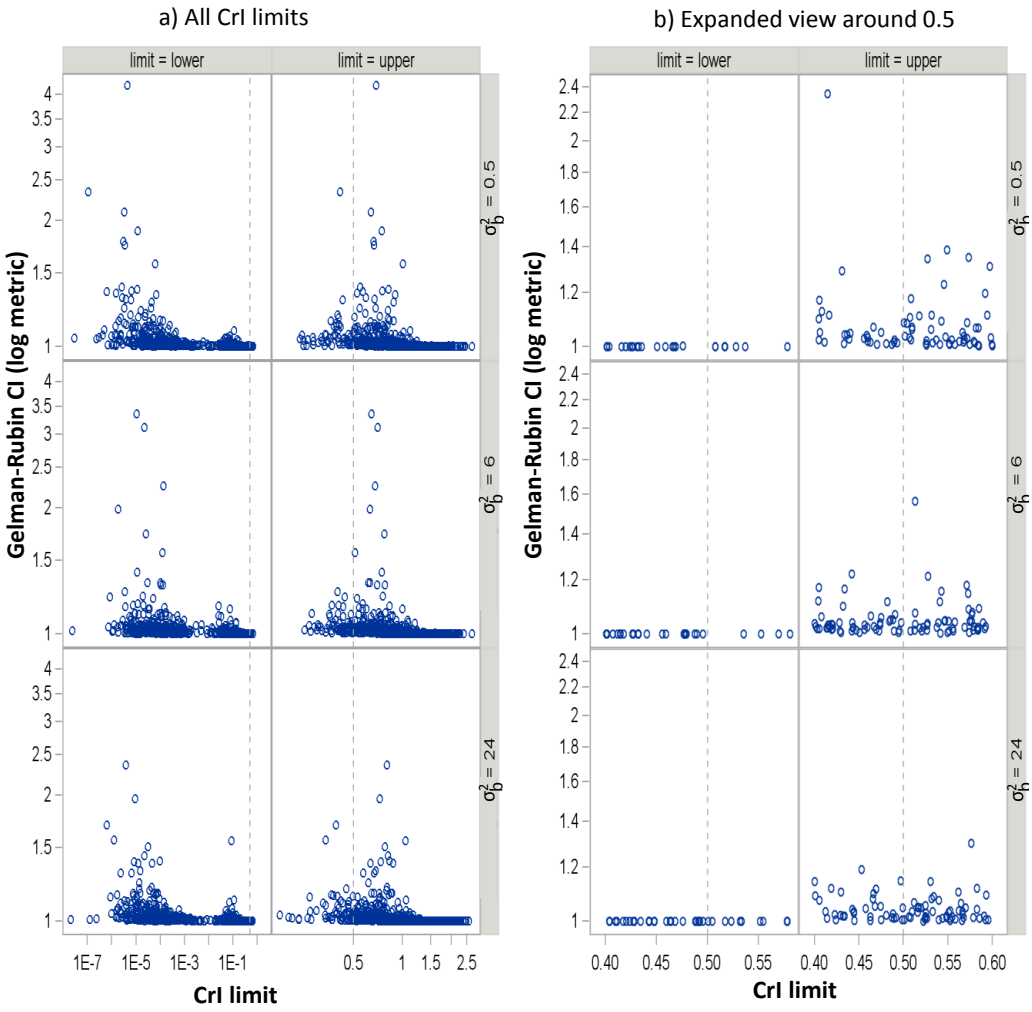
G-R CI statistic values are closer to the true value, there are so few it seems unlikely that the coverage would be affected. Thus a lack of convergence seemed an unlikely cause of the lower coverage for the variance estimate. As discussed in Section 6.3.2.3 the Gelman-Rubin statistics on the log metric provided an indication of when chains spent time at values very near zero (stickiness). Figure 8.21 plots the G-R CI statistic for the log metric against the CrI limit. Whilst there are a few datasets with a G-R CI statistic greater than 1.1 this is not particularly associated with small upper CrI limits and so the poor coverage seems unlikely to be associated with stickiness. In Figure 8.22 the CrI limits for the JEFF prior are plotted against the HT3 prior with points with higher G-R CI on the log metric identified. It is seen that the datasets where the G-R CI is higher do not appear for the lower limit, and are scattered through those for the upper limit suggesting that the estimate of coverage has not been affected.

FIGURE 8.20: **G-R CI versus lower and upper hpdv CrI limits for HT3 prior for σ_k^2 for scenario k6p16_0.5.**



In Figure 8.10 it was seen that when the lower CrI limit is at or very close to zero for one of the UNI, HCY or HT3 priors it does not necessarily have a value at or very close to zero on another. This does not affect whether the true value is above the CrI lower limit as the values are much smaller than the true value of 0.5. However, it is possible that the very low values were a result of the chain getting temporarily stuck at low values for one of the priors as described in Section 6.3.2.3. In this case the possibility that this also affected the upper CrI limit was considered, even though the G-R CI statistic didn't indicate a general problem. Lambert et al. (2005) identified a problem with "stickiness" in their simulation study which occurred with the design with the largest number of levels (30, with 5 and 10 also investigated) for the higher level random effect and smallest between-unit deviation investigated (0.001, with 0.3 and 0.8 also investigated). Their largest design might be considered comparable, 30 units compared

FIGURE 8.21: **G-R CI on log metric versus lower and upper hpdv CrI limits for HT3 prior for σ_k^2 for scenario k6p16_0.5.**



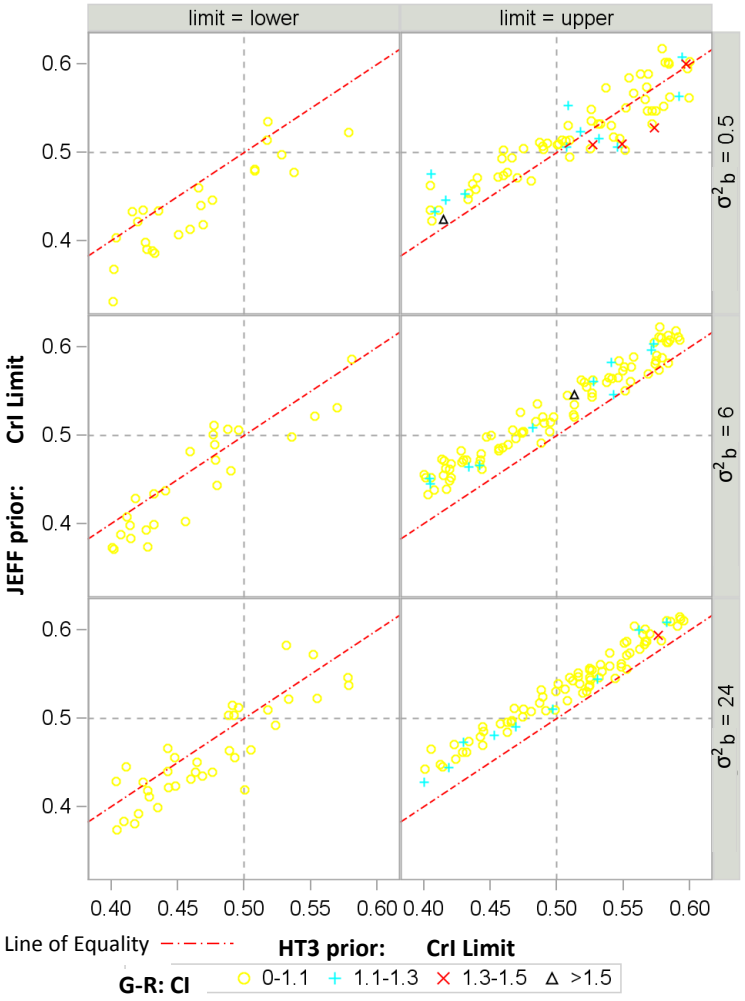
to my 36 kegs (6 batches with 6 kegs per batch). However, their between-unit standard deviation of 0.001 is extremely small compared to my keg SD of 0.71 (variance=0.5).

In Figure 8.23 the upper CrI limits are plotted against the lower limits and datasets with a difference between the lower limits for HCY and HT3 priors greater than 0.1 are identified (those within 0.1 are identified as HCT $\sim$ HT3). It is seen that the datasets where the lower limits differed by more than 0.1 did not occur where the upper CrI limits were around the true value of 0.5 and thus would not have affected the estimate of coverage.

Though none of the above seemed very likely causes of the low coverage, some further simulations were run to provide additional confidence or exploration of aspects seen.

- To see whether the smaller effective sample size associated with the smaller upper CrI limits had an effect on coverage, simulations were run with a larger number

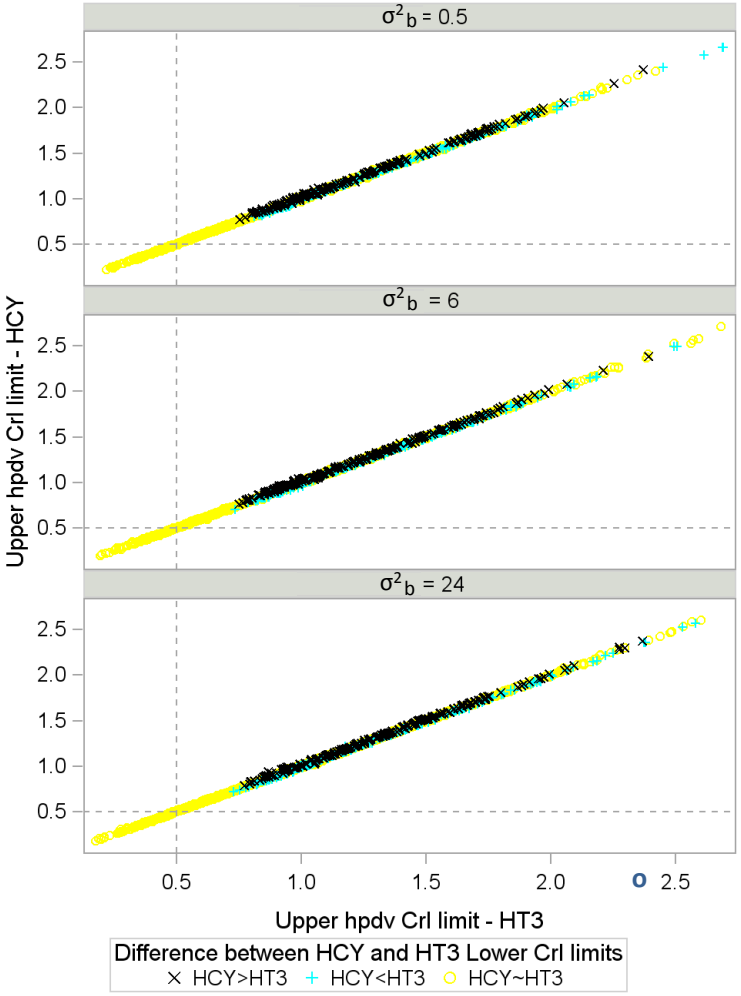
FIGURE 8.22: **CrI limits for JEFF prior plotted against those for HT3 prior with points with higher G-R CI on the log metric identified**



of posterior samples for $\sigma_b^2 = 6$. Five chains were run with a burn-in of 40,000 samples as for the original analysis, but for each chain the burn-in was followed with a further 1,800,000 posterior samples (160,000 were collected previously). Applying thinning of 1 in 10 resulted in a total of 900,000 samples post burn-in (compared to 80,000 previously). The analyses were run for 258 datasets where the CrI limits were close to 0.5 and thus an inaccuracy in the CrI limit estimation could make a difference to the estimate of coverage. The coverage of the simulations with larger number of posterior samples was 0.928 compared to 0.929 for the original simulations showing that the poor coverage is not due to the posterior sample size being too small. The results are shown in Figure 8.24 where it is seen that the CrI limits are reasonably similar and there is no observable bias.

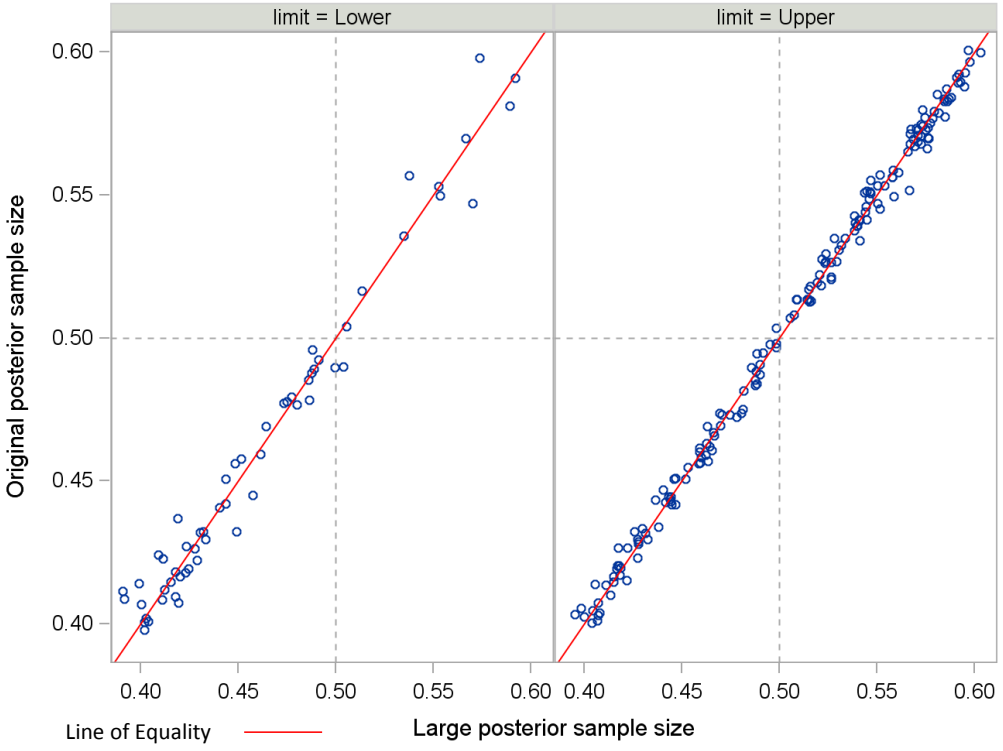
- To check whether "stickiness" accounted for the differences sometimes observed between the lower CrI limit for the HCY, UNI and HT3 priors, dataset 139 was

FIGURE 8.23: **Upper hpdv CrI limits for HCY plotted against HT3 with points with differences between the lower CrI limits identified**



re-run 60 times for the HCY, HT3 and UNI priors for $\sigma_b^2 = 0.5$. Dataset 139 was chosen as, whilst not practically important compared to a true value of 0.5, a difference was observed between the lower CrI limits for the HCY and HT3 priors (0.044 for the HCY prior and 0.00052 for the HT3 prior). Figure 8.25 shows the results where the G-R CI on the log metric and the upper CrI limit for HCY, HT3 and UNI priors are plotted against the lower CrI limit for σ_k^2. It was seen in Section 6.3.2.3 that a high G-R CI on the log metric could indicate stickiness. It is seen that the CrI limits from the repeated analyses fall into two groups - most have higher lower and upper CrI limits ("typical group") but a few have very small lower limits with a corresponding slightly lower upper CrI limit for σ_k^2 ("low" group) suggesting that the difference in results between the priors is due to an aspect of the simulations / MCMC procedure rather than a true difference. In some cases the analyses in the "low" group corresponded to high G-R CIs on the log

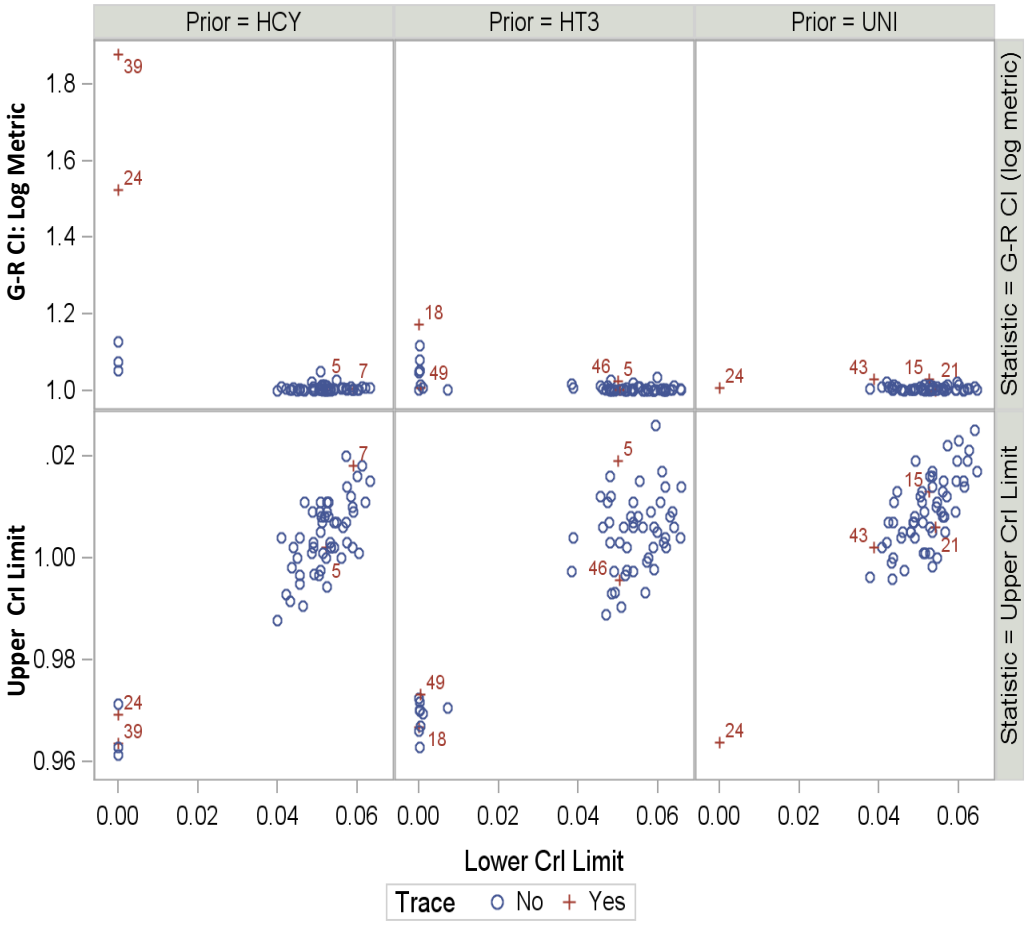
FIGURE 8.24: **Comparison of CrI limits for HT3 prior with original (160,000) posterior samples for σ_k^2 versus 900,000 samples around true value of 0.5**



metric being higher but not for all. The posterior samples for some example runs (identified in Figure 8.25 as those points with a trace) are plotted in Figure 8.26. These were chosen to include at least one from each group for each prior. It is seen that for the HCY prior, runs 24 and 39 had high G-R CI statistic on the log metric and for each of these runs one of the five chains of posterior samples spent a bit of time at very low values. Though for the HT3 prior runs 18 and 49 and for the UNI prior run 24 had very small lower limits, visually the chains did not appear that different from those with the higher lower CrI limits. Thus, whilst a high G-R CI statistic on the log metric indicated a potential problem with the chain spending some time at very low values which also resulted in smaller upper CrI limits, the absence of this did not necessarily indicate that there was no problem.

- Though examination of Figure 8.23 showed that possible stickiness which was identified through differences between priors for a dataset would not have affected the estimate of coverage (as this occurred for upper CrI limits above the true value), further simulations were performed to confirm that there was not an issue with upper CrI limits in the region of the true value. Datasets having an upper CrI limit for σ_k^2 less than 0.6 for any of HT3, HCY and UNI priors when $\sigma_b^2 = 6$ were re-analysed four times using the HT3 prior. For each of the re-run datasets

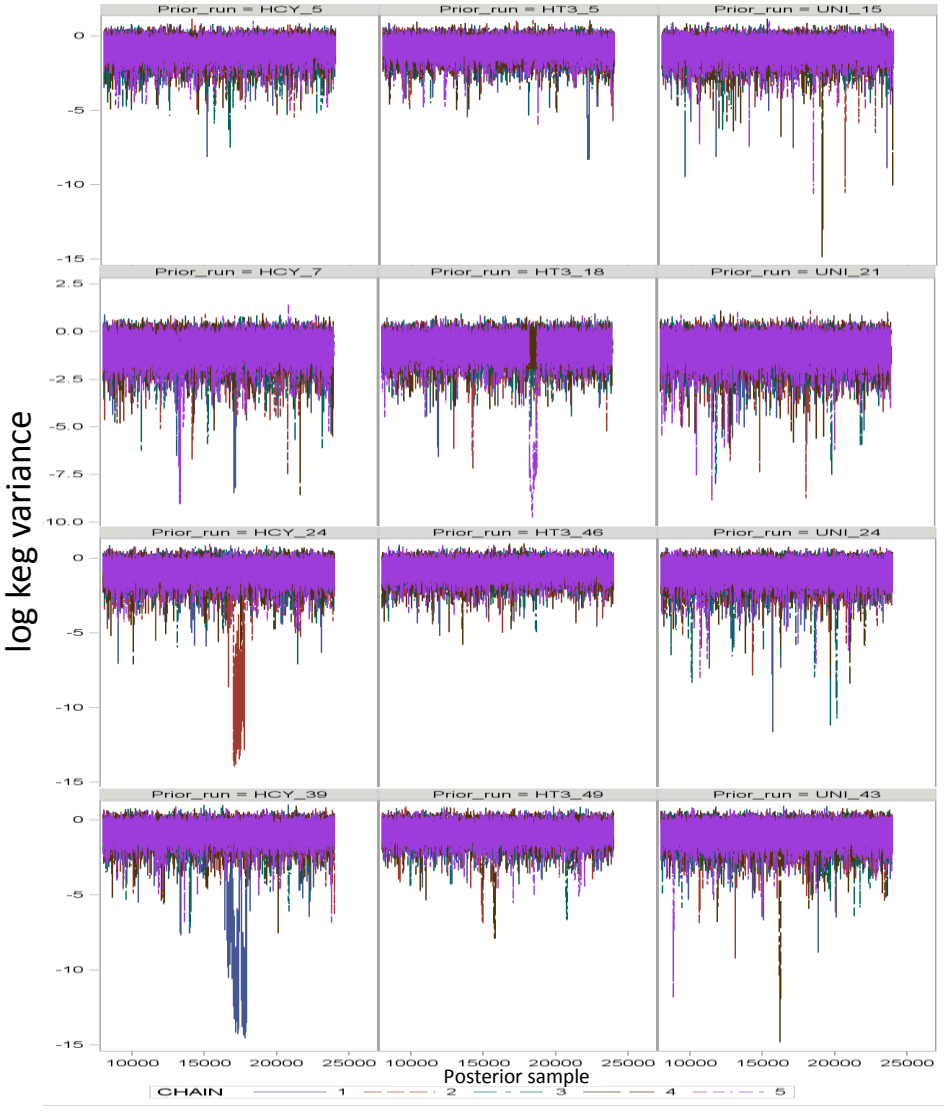
FIGURE 8.25: **Gelman-Rubin convergence CI statistic on log metric and upper CrI Limit plotted against lower CrI limit for σ_k^2 by prior**



the maximum, median and minimum of the five CrI upper limits (four re-runs and original) was calculated and used to estimate coverage. The estimates of coverage were 0.930, 0.929 and 0.928 respectively compared to 0.929 which was originally obtained, confirming that the low coverage is not due to a bias in the upper CrI limit estimate due to stickiness occurring for some analyses.

- A further 2000 datasets were analysed, giving 4000 in total. The two sets of 2000 datasets had coverages of 0.929 and 0.935 giving a coverage of 0.932 overall. Thus, though slightly higher for the second set of 2000 datasets, the coverage was confirmed to be poor.

FIGURE 8.26: **Posterior samples for example analyses of dataset 139**



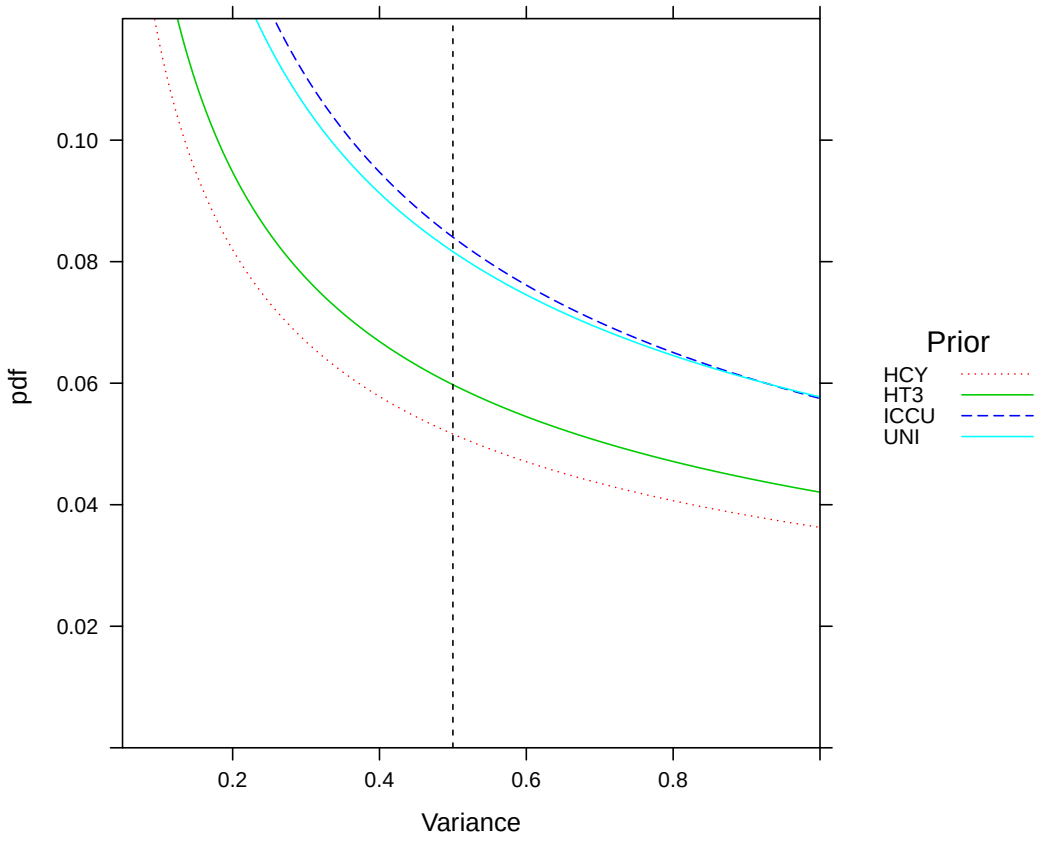
8.4.2 Prior Specification and Other Influences

In Section 8.4.1, no aspect of the simulations or the MCMC procedure applied could be found to explain the poor coverage and thus it seems likely that the poor coverage is real. In this section firstly the effect of the shape of the prior on the coverage is examined by changing the HT3 scale parameter. Secondly the coverage for higher σ_k^2 values is considered. Finally the coverage for a larger design is assessed.

The shape of the priors is firstly considered. One possibility is that the poor coverage is due to the behaviour of the priors near zero on the variance metric i.e. they have a steep slope as the variance decreases. Figure 8.27 shows the HCY, UNI, HT3 and ICCU priors on the variance metric for smaller values (this is an expanded version of Figure 5.1). It shows that around a variance component of 0.5 the pdf has a non-negligible slope which is slightly bigger for UNI and ICCU. The HT3 prior is a half-t distribution with 3 degrees of freedom and a scale parameter = $5\sqrt{3}$. The effect of changing the scale parameter on the pdf for the half-t distribution is examined in Figure 8.28. The original prior with a scale parameter = $5\sqrt{3}$ (HT3) is denoted by HT3_3. HT3_1, HT3_6 and HT3_24 represent scale parameters of $5\sqrt{k}$ where k=1, 6 and 24 respectively. Increasing the scale parameter reduces the slope at σ_h^2=0.5. The simulations were repeated with differing scale parameters. The coverages obtained were 0.929, 0.929, 0.928 and 0.929 for priors HT3_1, HT3_3, HT3_6 and HT3_24 respectively. Thus it is seen that the coverage is not affected by the scale parameter within the range investigated. This is also seen in Figure 8.29 which shows the CrI limits around the true value of 0.5 for the HT3 prior with various scale parameters plotted against those for the JEFF prior.

Next σ_k^2 is considered. The design with 6 batches, 6 kegs per batch and 16 portions per keg was only fully investigated for a true σ_k^2=0.5. Given that coverage greater than 0.94 is not reached for this scenario, it is of concern whether poorer coverage might be obtained for other true values of σ_k^2. It was noted in Section 8.3.1 that, though for true σ_k^2=0.5 the HT3 prior gives lower coverage than the JEFF prior, it is likely that as the true σ_k^2 value increases the HT3 prior will give higher coverage than the JEFF prior for this design. However, this does not give assurance for the HT3 prior since it was seen in Table 7.3 that for the original batch sampling design the JEFF prior has coverage much less than 0.94 for higher values of σ_k^2. In Figure 7.13 it was seen that for the original batch sampling design the HT3 prior had lower coverage for σ_k^2=24 compared to σ_k^2=6 and 0.5, though still above 0.94. Thus simulations were performed for a true value of σ_k^2=24 for the design k6p16. The coverages were 0.952, 0.957 and 0.945 for σ_b^2 = 0.5, 6 and 24 respectively. Thus, though for σ_k^2=0.5 coverages of 0.939, 0.929 and 0.931 were observed for σ_b^2 = 0.5, 6 and 24 respectively (which are less than 0.94 required for "good" coverage), it is expected that for σ_k^2 values in the range of interest (0.5 to 24) coverage is unlikely to drop much, if at all, below that seen for σ_k^2=0.5. In contrast the JEFF prior

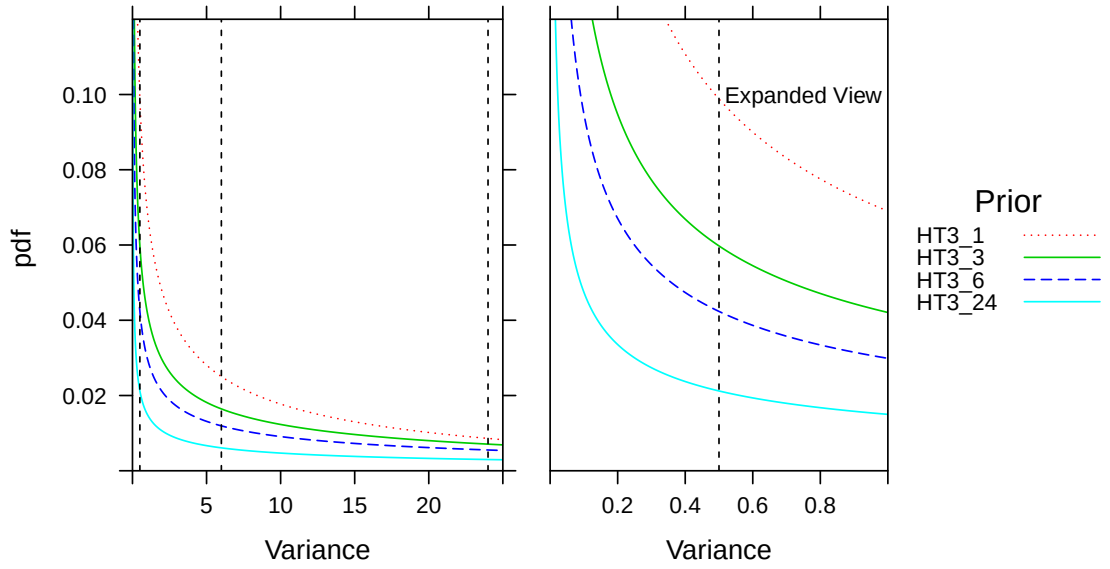
FIGURE 8.27: **PDF for HCY, HT3, ICCU and UNI priors on σ_h^2 metric around value of 0.5**



had similar coverages when $\sigma_b^2 = 6$ or 24 (0.946 and 0.944 for σ_k^2=0.5 compared to 0.940 and 0.948 for σ_k^2=24), but when $\sigma_b^2 = 0.5$ the coverage dropped from 0.944 to 0.919.

Finally analyses were performed to ascertain the sensitivity of the coverage value to the design size. Firstly, a scenario with 6 batches, 12 kegs per batch and 16 portions per keg for $\sigma_k^2 = 0.5$ (denoted k12p16_0.5) and $\sigma_b^2 = 6$ was used. Given the considerable amount of time the simulations would take using Winbugs and the close relationship seen between the CrI limits for the JEFF and HT3 priors for the design k6p16_0.5 the analyses were firstly performed using the JEFF prior. Those datasets with lower or upper CrI limits around the true value of $\sigma_k^2 = 0.5$ (between 0.45 and 0.55) were then run in Winbugs using the HT3 prior. Figure 8.30 compares the CrI limits for the HT3 prior and JEFF prior for the lower and upper limits. A regression line is fitted together with prediction intervals. This illustrates that for datasets with CrI limits outside the range 0.45 and 0.55 for the JEFF prior it is unlikely that the HT3 prior would give a different result. The coverage for the JEFF prior for σ_k^2 was 0.956 and for the HT3 prior 0.951. A similar approach was taken for designs with four (k4p16_0.5) or eight kegs

FIGURE 8.28: **PDF for half-t distribution priors with 3df and varying scale parameters on variance metric around σ_h^2=0.5**



(k8p16_0.5) per batch where the coverage was found to be 0.942 and 0.936 respectively. Thus, though the coverage for scenario k6p16_0.5 for the HT3 prior was only 0.929, it was seen for the smaller designs investigated and for the large design k12p16_0.5 the coverage is "good", and for scenario k8p16_0.5 it was 0.936, suggesting that it is unlikely that the coverage for the HT3 prior will drop much below the desired value for other designs.

FIGURE 8.29: **Comparison of CrI limits for changed scale parameter compared to original HT3 prior for σ_k^2 around true value of 0.5**

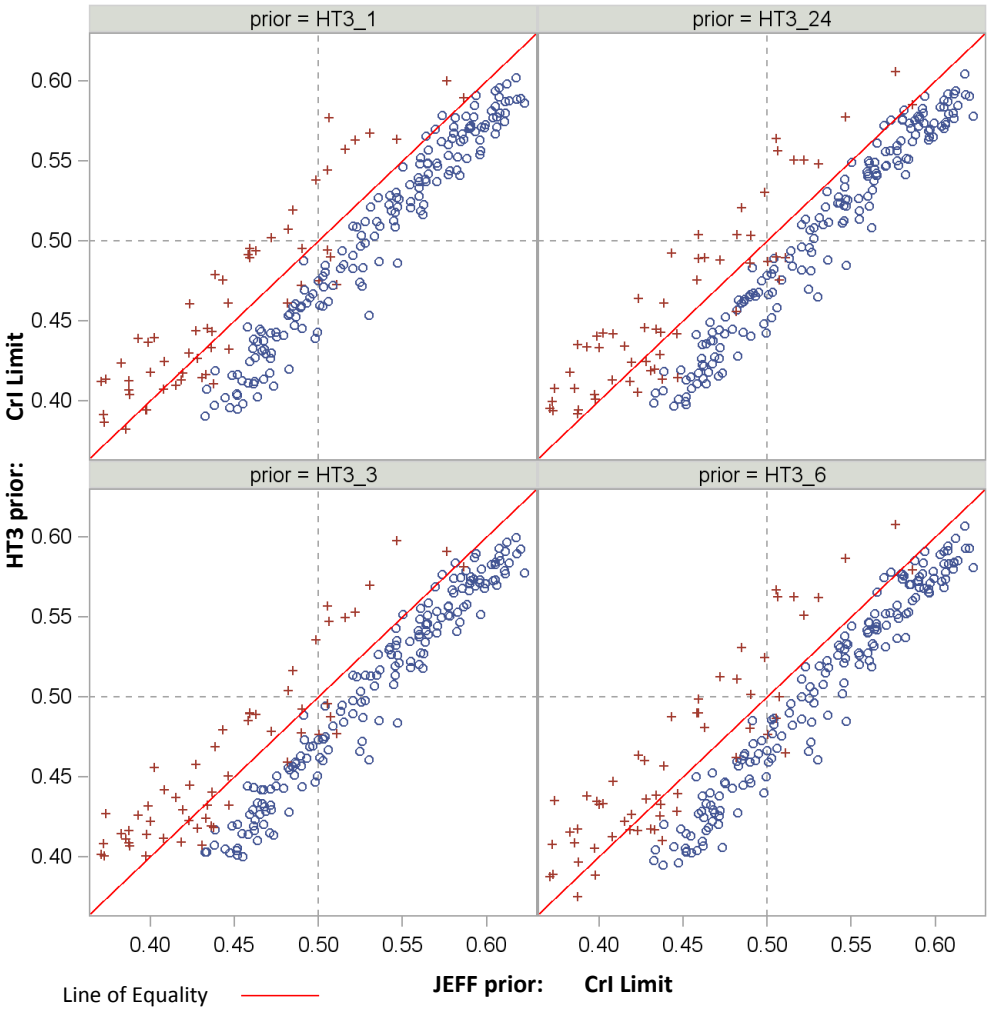
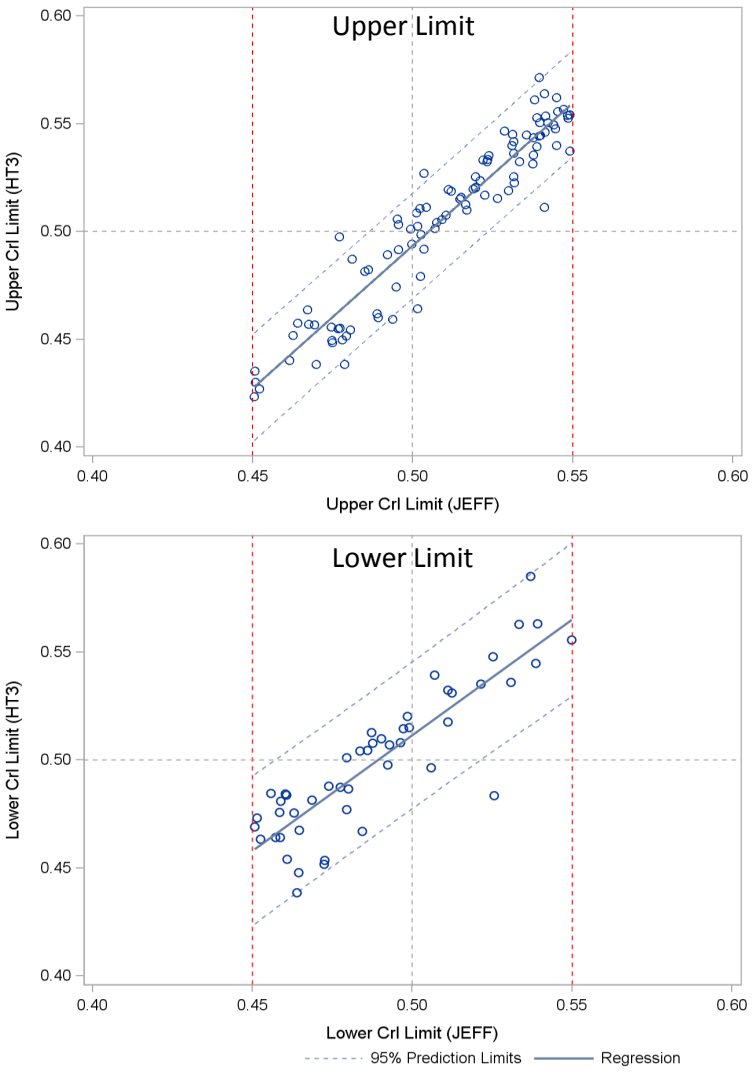


FIGURE 8.30: **Comparison of CrI limits for HT3 prior and JEFF prior for σ_k^2 around true value of 0.5**



8.5 Comparison of Posterior Medians with REML Estimates

The posterior medians are compared with the REML estimates for the HT3 prior. Figures 8.31 - 8.34 compare these for the total variance, total SD, batch, keg and portion variance components. The plots use the icc scale developed in Section 8.1.

It is seen that the posterior medians are typically higher than the REML estimates and the difference is greater for the smaller designs. For both posterior medians and REML estimates the mean over the datasets is higher than the median as expected given the skewed distribution of the variance estimates. For the posterior medians the mean and median are generally similar to or higher than the true value for all designs and all variances. For the REML estimates the median over the datasets is similar to or lower than the true value. However, the mean is lower than, similar to or higher than the true value dependent on the variance being estimated, the scenario and the value of σ_b^2. For σ_p^2 it is lower whilst for σ_k^2, σ_b^2 and σ_{tot}^2 it is similar or higher.

FIGURE 8.31: **Posterior median against REML estimate for σ^2_{tot} for various priors and scenarios**

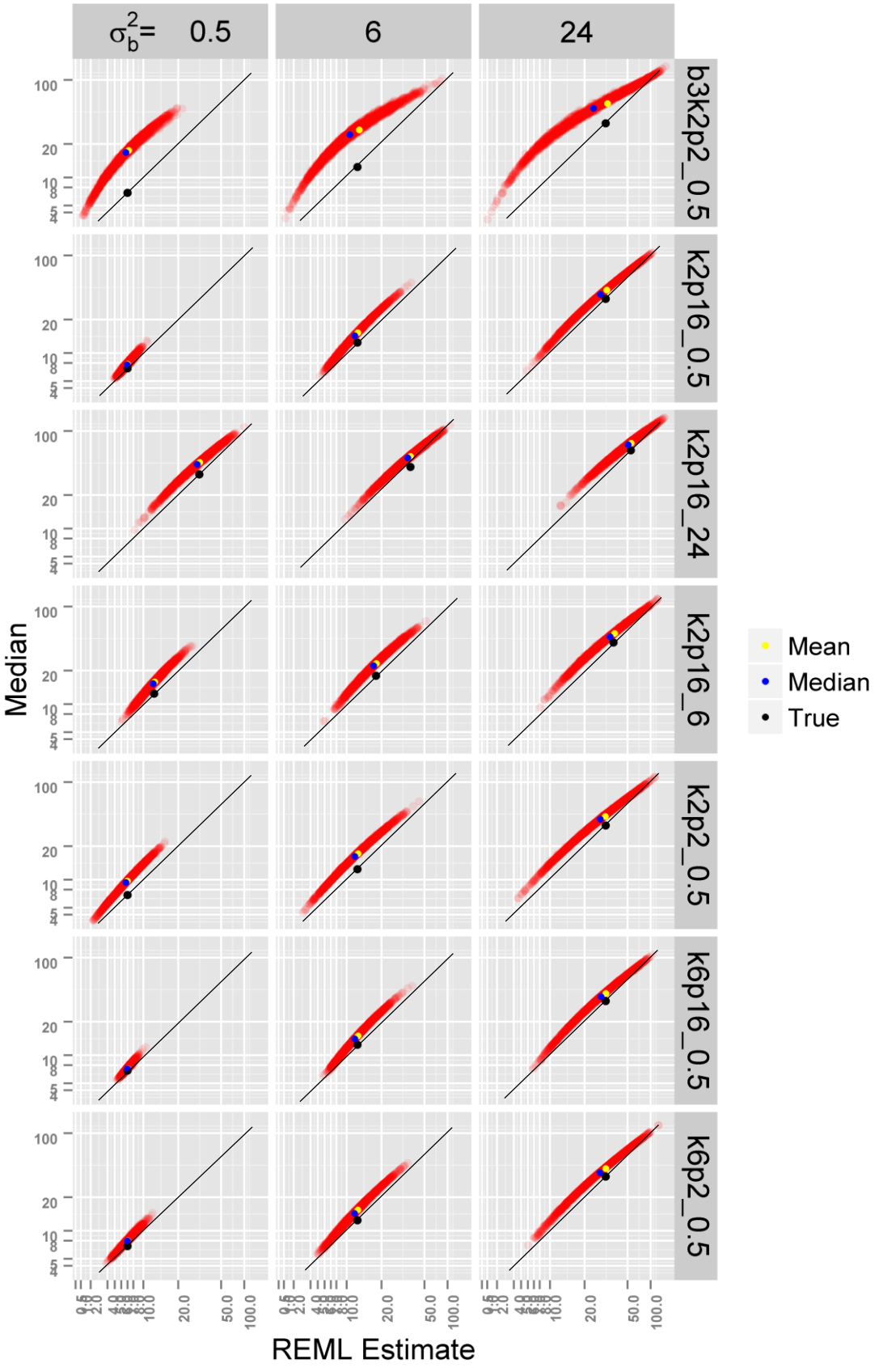


FIGURE 8.32: **Posterior median against REML estimate for σ_b^2 for various priors and scenarios**

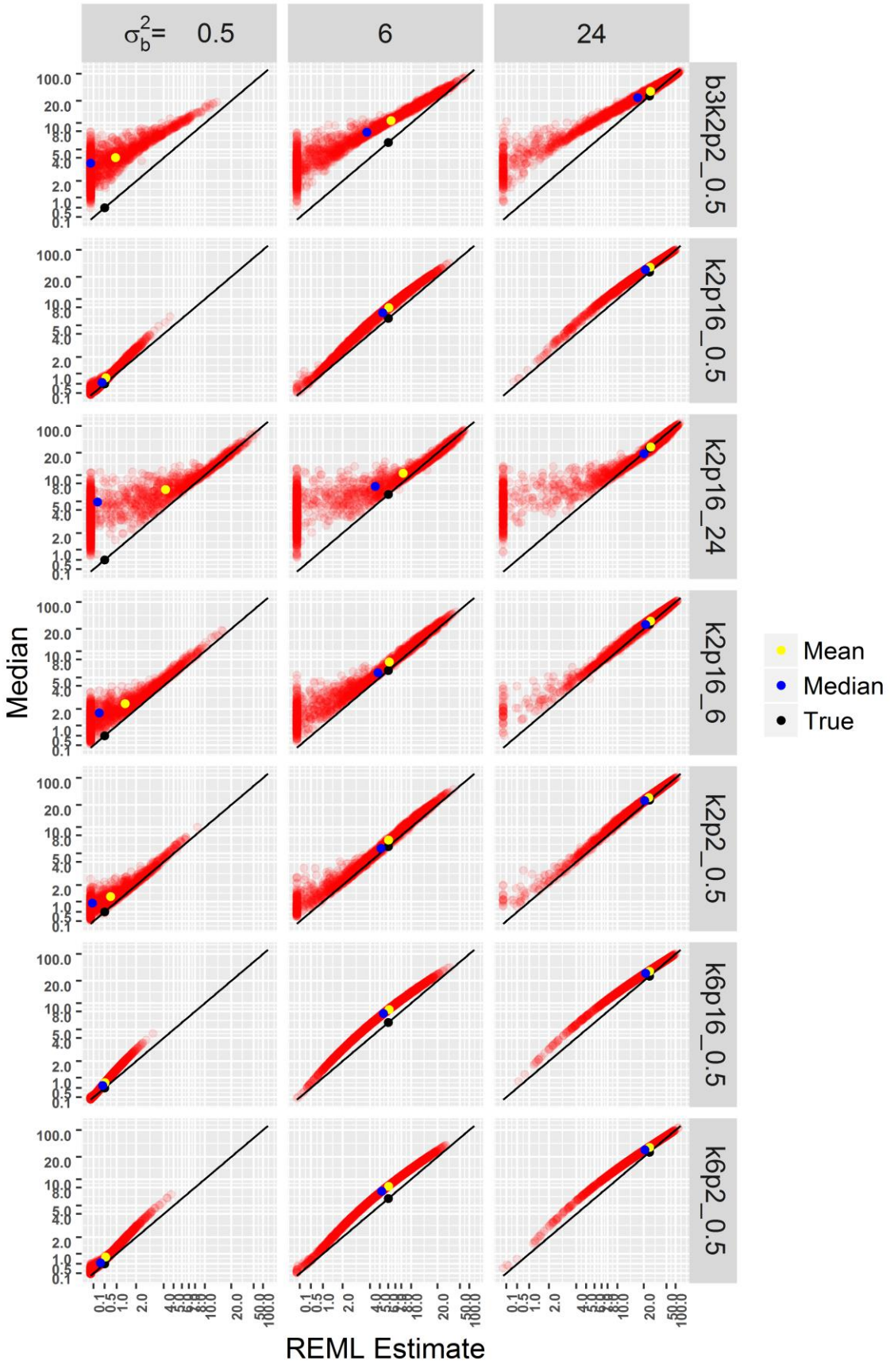


FIGURE 8.33: **Posterior median against REML estimate for σ_k^2 for various priors and scenarios**

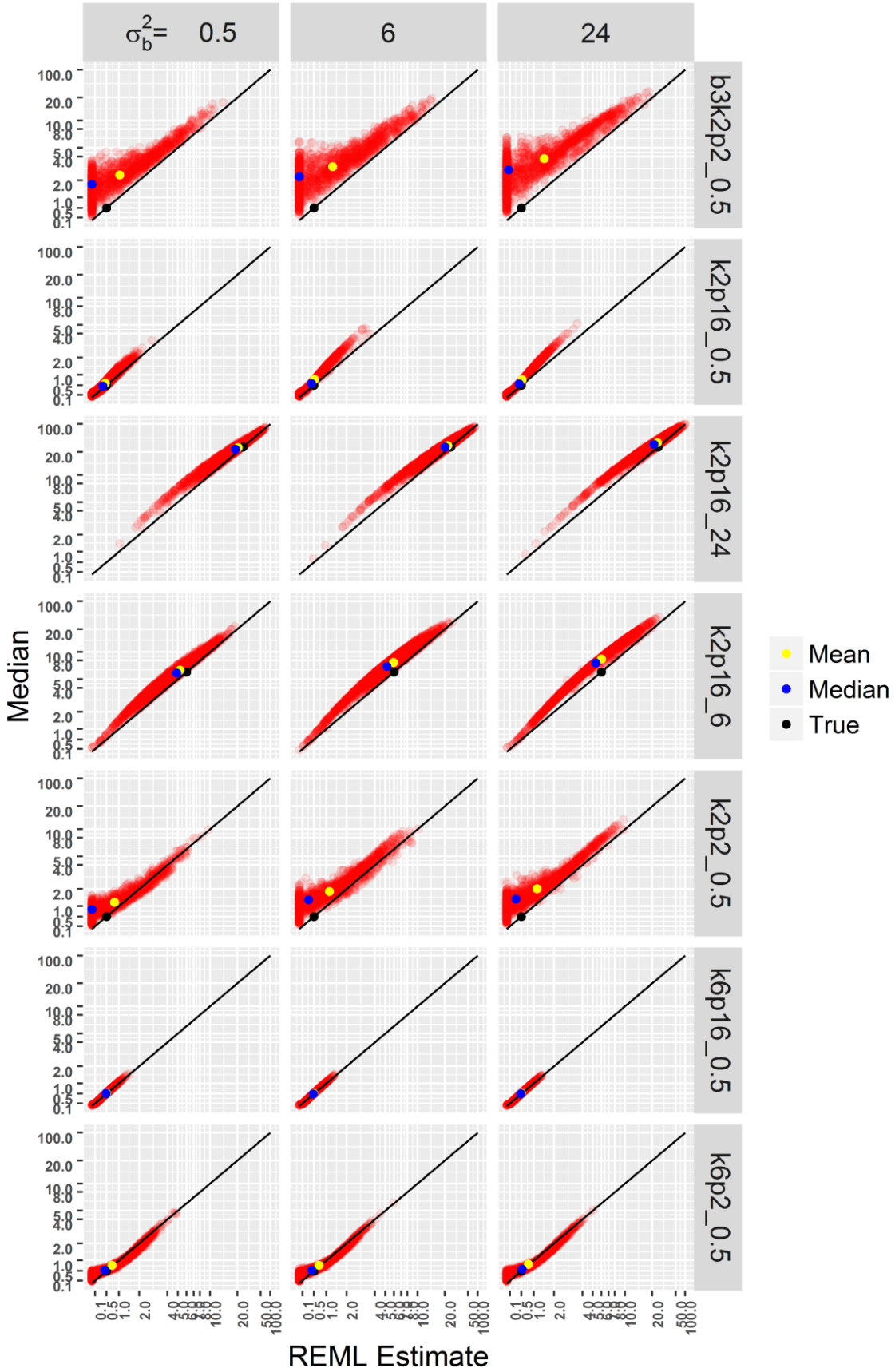
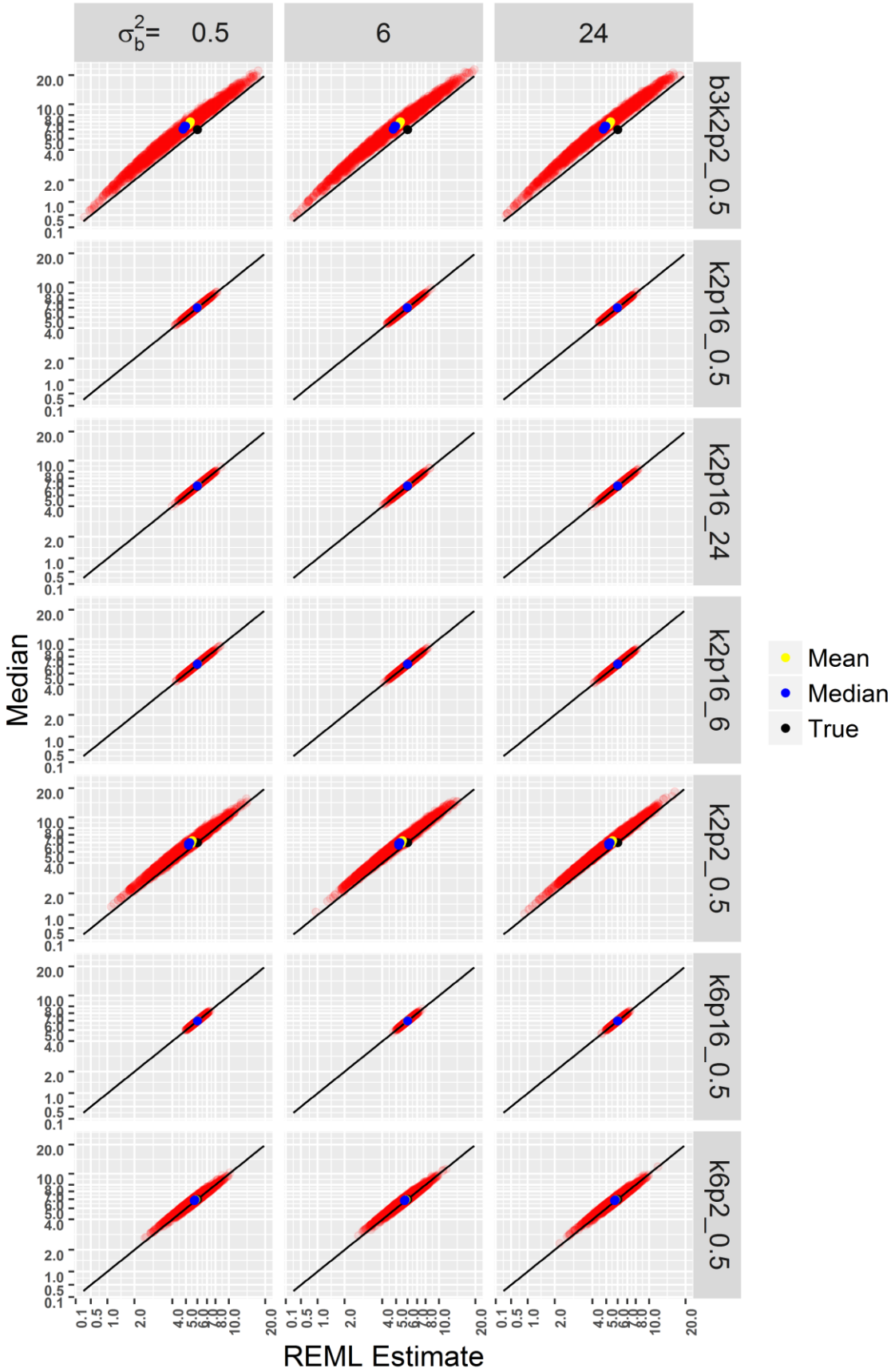


FIGURE 8.34: **Posterior median against REML estimate for σ_p^2 for various priors and scenarios**



8.6 Summary

This chapter examined the credible interval limits and posterior medians produced for the individual variance components and the total variance from the Bayesian analyses performed, in the light of the coverages obtained in Chapter 7. It also investigated the low coverage for scenario k6p16_0.5 seen for HCY, HT3 and UNI priors.

Firstly a novel way of plotting the credible intervals was developed in Section 8.2 - the icc (intraclass correlation) scale. This provides a better visualisation for variance components (for which the posterior samples have a skewed distribution) than either a linear or log scale.

Examination of the medians of the credible interval limits and the posterior median in Section 8.2 showed that the HCY and HT3 priors are very similar (as indeed their coverage was). The UNI prior is more similar to HCY and HT3 priors than the other priors but there are some differences for batch, keg and total variance where the CrIs were sometimes higher and/or shorter (on the icc scale). This occurred for σ_b^2 CrIs for some scenarios when σ_b^2=24 and also for σ_k^2 CrIs when σ_k^2=24. This also manifested itself in the total variance for these situations. In some cases the coverage was still "good" but in others being too high or too short was the reason for the poor coverage. For example, in Figure 8.17 it was seen that for σ_{tot}^2 for the scenario k6p16_0.5 and when $\sigma_b^2 = 24$, the lower CrI limit is higher than the HCY, HT3 and JEFF priors for the majority of the datasets and this results in the "poor" coverage. The effect of the UNI prior having no support for higher variances was also seen in Figure 8.17 for the upper CrI limits. The IG2 and JEFF priors are very similar and also similar to the IIN prior for the batch and keg variance components. The IIN prior CrIs differed for the portion variance which is not surprising as it was designed to be informative for that variance. For total variance the IIN prior was also similar to the IG2 and JEFF priors except for the scenarios with the two smallest designs and $\sigma_b^2 = 0.5$ or 6, where the higher median CrI limits and posterior median probably result from the informative prior for the portion variance. The JEFF, IG2 and IIN priors had poor coverage for σ_k^2 when σ_k^2=6 or 24 for most values of σ_b^2, where the upper CrI limits were too frequently below the true value. The considerably smaller upper limits were seen in Figure 8.14 for scenario k2p16_24 where it was seen that when the HT3 prior gives an upper CrI limit of 25 the JEFF prior gives a value around 15. The FLAT prior gives higher (often considerably higher) medians of the upper CrI limit and posterior median in all situations for the keg, batch and total variance and SD than the other priors, and almost all situations for the portion variance. This aligns with Gelman (2006) who stated that a uniform prior on σ_h^2 has a miscalibration towards higher values which is worse than a uniform prior on σ_h. The width of the FLAT CrI on the icc scale was often considerably larger than other priors which also achieved "good" coverage.

In Figure 8.4 it was seen for scenario k6p16_0.5 that, despite the coverage of the hpdv CrIs for UNI, HCY and HT3 priors being less than 0.94 and the coverage of the hpdv CrIs for IIN, IG2 and JEFF priors being greater than 0.94 for σ_k^2, the median CrIs seemed wider for the former priors compared to those for the latter. The CrI limits for individual datasets for scenario k6p16_0.5 were examined in Section 8.3.1 to investigate this, at first, counter-intuitive observation. The reason was as follows. The median CrIs for the JEFF prior are shorter than those of the HT3 prior because the median CrI limits lie in a range of values where the JEFF prior has higher lower limits, and lower upper limits, than the HT3 prior. However, in the range of values around 0.5, which are critical to the estimation of the coverage of the interval, the CrIs for the JEFF prior are longer than those of the HT3 prior and hence more often include the true value. It was confirmed that the JEFF prior had slightly higher coverage than the HT3 prior for nominal coverages between 80 and 99% (those likely to be most useful) despite the difference between the CrIs for the JEFF and HT3 priors depending on the value. Intervals with smaller target coverages were not explored but given the differences between the median CrI limits it is expected that for a small enough target coverage the coverage of the HT3 prior would exceed that of the JEFF prior. Also, though the JEFF prior gives better coverage than the HT3 and HCY priors when σ_k^2=0.5, it may not do for higher true values. For σ_k^2=24 and when σ_b^2=0.5 the coverage dropped to 0.919. However, for the HT3 prior for scenario k6p16_24 "good" coverage was achieved. Additionally, for the original batch sampling design with 6 batches, 2 kegs/batch and 16 portions/keg where the JEFF prior had good coverage for σ_k^2=0.5 it was seen that it did not for σ_k^2=6 or 24.

Given that the HCY and HT3 priors achieved "good" coverage (>0.94) in all situations except for scenario k6p16_0.5 for the keg variance, and no other prior did better overall, the results for this situation were explored in Section 8.4.1 to check whether the coverage was truly poor rather than an artefact of the simulations or MCMC procedure applied. Nothing was identified that was thought to have affected the coverage estimates. The desired effective sample size of 2000 had been achieved with almost all analyses of datasets achieving greater than 4000 for σ_k^2 for the HT3 prior. In addition for σ_b^2=6 the analyses were repeated with greater than ten times the original number of posterior samples and similar coverage was obtained. The Gelman-Rubin CI statistic also indicated good convergence with only three datasets with G-R CI greater than 1.05 and just one greater than 1.08. The Gelman-Rubin CI statistic on the log metric was examined as this provided an indication of stickiness. There were some datasets with higher values than desirable but with very few where the CrI limits were around the true value it was thought unlikely to have affected the coverage estimate. It was observed that when the lower CrI limit was at or close to zero for one of the UNI, HCY or HT3 priors it did not necessarily have a value at or close to zero on another. It was thought that this might be due to stickiness. Though a smaller estimate of the lower CrI limit wouldn't affect the coverage (as it was very small compared to the true value), if this

caused a smaller upper CrI limit, that might affect the coverage estimate. However, none occurred where the upper CrI limit was around the true value of 0.5. In addition, datasets which had upper CrI limits less than 0.6 (i.e. includes those datasets which, if stickiness had caused a problem, might have an upper limit above the true value of 0.5 on re-analysis) were re-analysed four times. It was seen that even taking the minimum upper CrI limit obtained over all 5 analyses for each dataset gave similar coverage. Finally a further 2000 datasets were analysed giving 4000 in total and similar, though slightly higher, coverage was obtained. Thus it was confirmed that the coverage was poor.

The sensitivity of the coverage to the shape of the HT3 prior and other influences was examined in Section 8.4.2 and found to be largely insensitive. The original HT3 prior had a scale parameter $= 5\sqrt{k}$ where k=3. Changing k to 1 ,6 or 24 for $\sigma_b^2 = 6$ gave coverages in the range 0.928 - 0.929 compared to the original coverage of 0.929. The effect of σ_k^2 on the coverage was considered. For σ_k^2=24 the coverages were 0.952, 0.957 and 0.945 for $\sigma_b^2 = 0.5$, 6 and 24 respectively, compared to 0.939, 0.929 and 0.931 for $\sigma_k^2 = 0.5$. Finally the effect of increasing the design size was assessed. For designs with 4, 8 and 12 kegs per batch (and 6 batches, 16 portions per keg, $\sigma_k^2 = 0.5$ and $\sigma_b^2 = 6$) the coverages for the HT3 prior were 0.942, 0.936 and 0.951 respectively compared to 0.929 for 6 kegs per batch. Thus, though the coverage for the HT3 prior for scenario k6p16_0.5 is only in the range 0.929 - 0.939 and has not achieved "good" coverage, it is concluded that the coverage is unlikely to drop much below the desired value for other situations.

In Section 8.4, though not the main topic of the thesis, the posterior medians were compared to the REML estimates. The posterior medians were typically higher than the REML estimates and the difference is greater for the smaller designs. For both posterior medians and REML estimates the mean over the datasets is higher than the median as might be expected given the skewed distribution of the variance estimates. For the posterior medians the mean and median are generally similar to or higher than the true value for all designs and all variances, with the difference being greater for small designs. For the REML estimates the median over the datasets is similar to or lower than the true value. However, the mean is lower, similar to or higher than the true value dependent on the situation.

Chapter 9

Application of Bayesian Analysis with Half-t Prior to Examples

Contents

It was seen in Chapters 7 and 8 that the HCY and HT3 priors gave good coverage in almost all situations. None of the other priors investigated gave better coverage. In addition the HT3 prior required fewer burn-in samples compared to the HCY prior as seen in Section 6.3.2.2. In this chapter some example datasets are analysed using a Bayesian approach, primarily using the HT3 prior. The examples provide an insight into the value of applying a Bayesian approach in the application area of process and analytical variation studies which is helped by having a recommended prior for routine use. Unless otherwise stated, the Bayesian analysis using the HT3 prior was performed in WinBUGS. Five chains of posterior samples were obtained, each having a burn-in of 360,000 samples (though this is probably unnecessarily large) followed by 160,000 posterior samples. These were then thinned and only 1 in 10 was retained.

9.1 Application to Batch Sampling Example

It was seen in Section 7.2.2, Table 7.3, that for scenario k2p16_6 and the JEFF prior, the coverage of the CrIs for the keg variance is less than 0.9. An example dataset is analysed using the JEFF and HT3 priors to compare the results. The design was described in Section 4.1. The analysis using the HT3 prior was performed in WinBUGS, whilst the analysis using the JEFF prior was performed in PROC MIXED in SAS software (where

20,000 posterior samples were requested). In addition, the results from two frequentist approaches are presented. A REML analysis was performed in PROC MIXED and CIs based on the chi-squared distribution described in Chapter 2, Equation (2.2), are presented (denoted Chi). Generalised confidence limits were obtained from an analysis performed in PROC VARCOMP in SAS software (denoted GCL).

The results of the analyses are provided in Table 9.1. The posterior median (denoted estimate) and 95% highest posterior density interval, the effective sample size (ESS) and the Gelman-Rubin convergence CI statistic (for analyses performed in WinBUGS) are provided for the variance components for the Bayesian analyses. The ESS (effective sample size) and Gelman-Rubin CI statistics are considered acceptable. The estimate and 95% CI are provided for the variance components for the frequentist analyses (note CIs are not provided for total variance/SD for the chi-squared method as these are not produced by PROC MIXED).

The CrI upper limits for the keg variance differ considerably for the HT3 and JEFF priors (5.96 for the HT3 prior versus 3.92 for the JEFF prior) and this difference is also reflected in the upper CrI limits for the total variance and SD. Without the simulations and evaluation performed in this thesis (the coverage for the keg variance for the HT3 prior was estimated as 0.957 whilst that for the JEFF prior was estimated as 0.873 for 95% nominal intervals), a user would not know which results to use. The generalised upper confidence limit for the keg variance lies between those of the HT3 and JEFF priors. The CrIs for the batch variance are fairly similar for the HT3 and JEFF priors. The CI limits for the batch and total variances for the GCL method are higher than those for the Bayesian analyses, whilst the upper CI limits for the batch and keg variance components for the chi-squared method are very high compared to the other methods. The CrIs/CIs for the portion variance are fairly similar for all four methods.

9.2 Application to Examples 3-6

A frequentist REML analysis performed on four example analytical datasets was described in Section 2.1. The results were shown in Table 2.2. The four datasets are now analysed using the HT3 prior in WinBUGS and the results are provided in Table 9.2. The bottom-level variance was anticipated to be around 6, as was the case in the batch sampling study. Thus the HT3 prior placed on all the higher-level variance components uses the same value for the prior parameter. The same uniform prior was placed on the bottom-level variance.

The Gelman-Rubin CI statistic on the linear metric (the metric of the variance or SD being estimated) for the HPLC variance component for dataset 5 is a little high (1.202) and might warrant a longer burn-in for a final analysis. Otherwise the ESS (effective sample size) and Gelman-Rubin CI statistics are considered acceptable. In contrast to

TABLE 9.1: **Results for batch sampling design with example data from scenario k2p16_6 analysed using Bayesian and frequentist approaches**

		Bayesian		Frequentist	
		HT3	**JEFF**	**Chi**	**GCL**
Total variance	Estimate	10.34	9.63	8.72	8.72
	CrI/CI	(6.20,21.56)	(6.28,18.48)	-	(7.18,24.43)
	ESS	20040	20000	-	-
	G-R CI (linear)	1.0016	-	-	-
	G-R CI (log)	1.0001	-	-	-
Total SD	Estimate	3.22	3.10	2.95	2.95
	CrI/CI	(2.54,4.68)	(2.53,4.32)	-	(2.68,4.94)
	ESS	19920	20000	-	-
	G-R CI (linear)	1.0004	-	-	-
	G-R CI (log)	1.0001	-	-	-
Batch variance	Estimate	1.96	2.05	1.62	1.62
	CrI/CI	(0.00,12.55)	(0.00,10.14)	(0.44,58.87)	(0.86,16.59)
	ESS	17439	20000	-	-
	G-R CI (linear)	1.0019	-	-	-
	G-R CI (log)	1.0057	-	-	-
Keg variance	Estimate	1.78	1.30	1.23	1.23
	CrI/CI	(0.06,5.96)	(0.11,3.92)	(0.42,12.37)	(0.19,5.03)
	ESS	25449	20000	-	-
	G-R CI (linear)	1.0007	-	-	-
	G-R CI (log)	1.0006	-	-	-
Portion variance	Estimate	5.93	5.89	5.87	5.87
	CrI/CI	(4.78,7.28)	(4.75, 7.17)	(4.82,7.30)	(4.76,7.43)
	ESS	79580	20000	-	-
	G-R CI (linear)	1.0002	-	-	-
	G-R CI (log)	1.0002	-	-	-

the bounded REML analysis, intervals are now obtained for all variance components. They are also easily obtained for the total variance and total SD. The site variance component for dataset 3 had an upper confidence limit from the bounded REML analysis of 33864 whilst for the unbounded analysis the upper limit was 86.9. The Bayesian analysis has an upper CrI limit of 366 which is more believable than 33864, though still very large. However, as the variance component is being estimated from a notional one degree of freedom this is not surprising. For the unbounded REML analysis for dataset 3 the analyst variance component has an upper CI limit of 0.85. This seems small given that for an ANOVA type analysis there would only be two dfs to estimate this variance. In Section 2.2.4, Table 2.4 it was seen that the coverage for the CIs from the frequentist analysis is low. In contrast the Bayesian analysis provides a CrI of (0.00,40.95) which is more plausible.

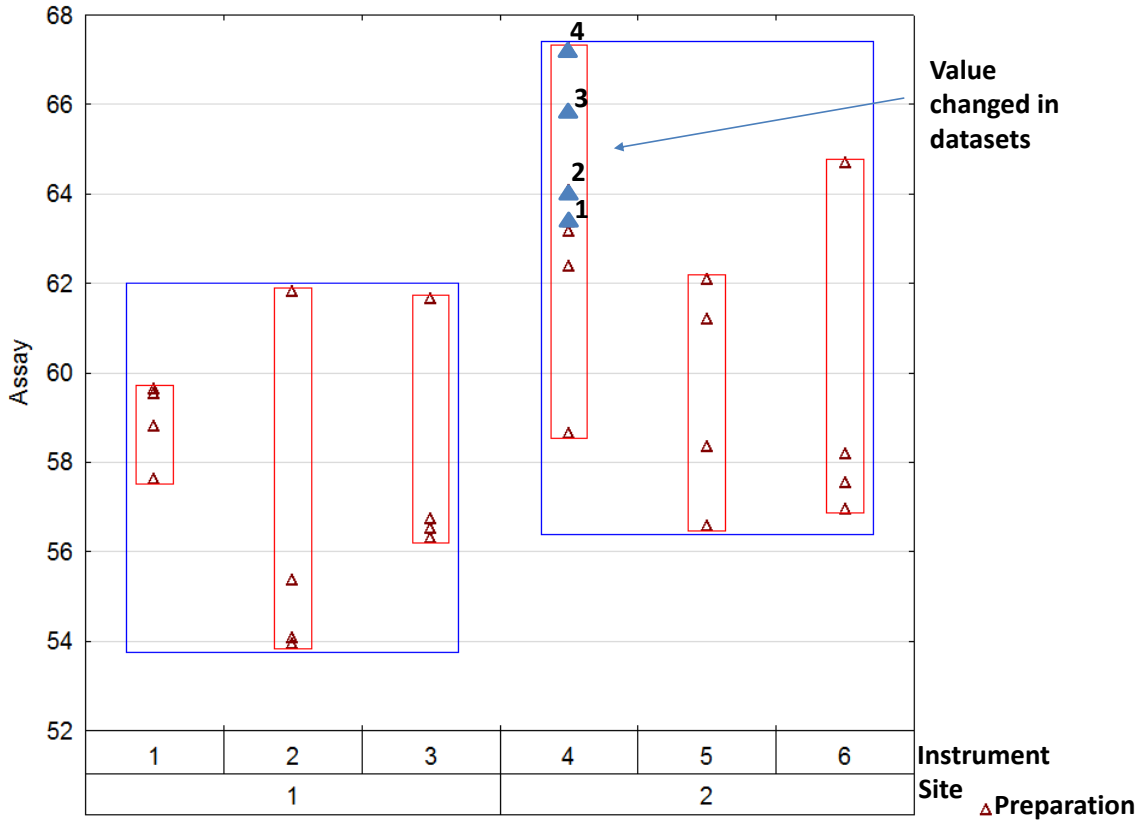
TABLE 9.2: **Results for Bayesian analysis of Examples 3-6 using HT3 prior**

Dataset		Variance	Median	CrI limit		ESS	G-R CI	
				Lower	Upper		linear	log
Balanced	**3**	Site	29.71	0.00	366.4	4321	1.030	1.006
		Analyst(Site)	1.54	0.00	40.95	12889	1.004	1.006
		Instr(Site)	2.24	0.00	48.93	14228	1.002	1.008
		Column(Site)	2.47	0.00	50.66	14977	1.004	1.015
		Residual	5.73	1.97	13.61	75318	1.000	1.000
		Total variance	67.47	6.89	430.1	4298	1.028	1.005
		Total SD	8.21	2.88	20.83	4689	1.015	1.005
	5	Site	13.83	0.00	252	11802	1.012	1.001
		Analyst(Site)	8.96	0.00	88.63	19745	1.005	1.006
		Instr(Site)	1.82	0.00	42.47	12066	1.202	1.005
		Column(Site)	3.19	0.00	54.54	17777	1.022	1.105
		Residual	11.52	3.65	27.57	62896	1.000	1.000
		Total variance	67.78	8.13	374.7	9262	1.011	1.001
		Total SD	8.23	3.23	19.56	8627	1.002	1.001
Unbalanced	**4**	Site	26.32	0.00	307.60	10337	1.005	1.002
		Analyst(Site)	5.10	0.00	68.60	13043	1.016	1.002
		Instr(Site)	1.28	0.00	37.57	11434	1.056	1.012
		Column(Site)	1.25	0.00	37.10	10201	1.003	1.002
		Residual	3.37	0.74	11.16	61531	1.001	1.000
		Total variance	62.63	4.81	378.60	9549	1.005	1.003
		Total SD	7.91	2.53	19.59	7947	1.003	1.003
Unbalanced	**6**	Site	15.04	0.00	257.80	11865	1.009	1.001
		Analyst(Site)	9.15	0.00	92.07	26257	1.001	1.001
		Instr(Site)	2.23	0.00	47.73	19554	1.021	1.004
		Column(Site)	3.98	0.00	64.12	19930	1.006	1.002
		Residual	11.55	2.85	35.18	68014	1.000	1.000
		Total variance	75.12	6.88	385.80	12281	1.007	1.000
		Total SD	8.67	3.23	19.89	12682	1.001	1.000

9.3 Analytical Equivalence Example

The final example is a simple equivalence study constructed to demonstrate two issues that occur with the REML analysis when variance components are small, and illustrate the advantage of taking a Bayesian approach. The study has a three-level nested design similar to the batch sampling study, but in this case the highest-level effect (site) will be treated as a fixed effect. There are three factors: site, instrument and preparation. It is wished to compare the means for the two sites and demonstrate that the means of the sites are equivalent (sufficiently similar). The data is shown in Figure 9.1. There are four datasets which are identical except for one set of conditions where the value is changed slightly for the four datasets. These points are identified by blue triangles.

FIGURE 9.1: **Equivalence study example**



An analysis using the REML approach is performed in PROC MIXED and a Bayesian analysis using the HT3 prior is performed in WinBUGS. The example was constructed so that bottom-level variance was anticipated to be around 6. Thus the same HT3 prior is placed on the instrument variance component as was used for the higher-level variance components in the batch sampling study. The same uniform prior as was previously used was placed on the bottom-level variance. An extremely diffuse normal prior, $\mathcal{N}(0, \sigma^2 = 10^{10})$ is placed on each of the site means. Estimates and intervals for the variance components (confidence intervals for the REML analysis and credible intervals for the Bayesian analysis) are given in Table 9.3. The posterior samples from applying a Bayesian approach make it easy to find intervals for functions of the variance components such as the total variance reported in the table. The probability that the instrument variance is greater than the preparation variance can easily be ascertained. An indicator variable is used to identify when the posterior sample of the instrument variance component is larger than the preparation variance. Taking means over this variable for all the posterior samples gives an estimate of the probability that the instrument variance component is larger than the preparation variance. The frequentist method based on the chi-squared distribution (Equation (2.2)) cannot provide a confidence interval when a variance component is zero or very small as is seen for datasets 1 and 2. It is also seen

in Table 9.3 that despite dataset 4 having a larger estimate for the instrument variance component than dataset 3, it has a smaller upper confidence limit which is likely to be an artefact of the approximation to the degrees of freedom being inappropriate.

TABLE 9.3: **Instrument, preparation and total variance with accompanying 95% intervals**

Anal.*		**Instrument var.**			**Residual var.**			**Total var.**			**Prob.+**
		Est.	Limit		Est.	Limit		Est.	Limit		
			L	U		L	U		L	U	
R	1	0.00	-	-	7.59	4.54	15.21	7.59	-	-	-
	2	0.06	-	-	7.77	4.44	16.99	7.83	-	-	-
	3	0.35	0.03	1.07E+44	8.29	4.73	18.13	8.63	-	-	-
	4	0.57	0.05	4.42E+20	8.89	5.08	19.44	9.46	-	-	-
B	1	1.08	0.00	13.32	7.98	3.86	14.67	9.96	4.08	23.94	0.103
	2	1.18	0.00	15.00	8.18	3.99	15.13	10.33	4.18	25.56	0.113
	3	1.50	0.00	17.17	8.85	4.20	16.26	11.38	4.48	28.38	0.124
	4	1.72	0.00	18.94	9.58	4.56	17.62	12.46	4.90	30.80	0.132

* Analysis: R=REML, B=Bayesian for datasets 1-4

\+ Probability that instrument variance is larger than the preparation variance

The mean difference between the sites together with 95% intervals (confidence intervals for the REML analysis and credible intervals for the Bayesian analysis) are provided in Table 9.4. In addition, the p-value for testing the null hypothesis that the difference is zero (REML) or probability that the difference is greater than zero (Bayesian) is provided. The datasets have been chosen to illustrate what happens as a variance component goes to zero. It is seen in the REML analysis that there is a discontinuity between datasets 1 and 2 where the dfs increase from 4 to 22 and bigger changes in the p-value and the upper confidence limit occur than seen between the other datasets. It is seen that the Bayesian analysis does not have this issue. The upper limits for the instrument variance are also much more believable than those for the REML analysis (where provided).

In summary: The Bayesian approach provides a methodology for interval estimation which does not suffer from the issues seen in this chapter for the frequentist approach when applied to small studies with small variance components, and is very flexible in being able to cope with a variety of designs and a variety of outcomes of interest.

TABLE 9.4: **Mean difference between sites with accompanying 95% intervals**

Analysis	**Dataset**	**Mean Site Difference**				
		Estimate	Limit		P-value	df
			lower	upper		
REML	1	-2.59	-4.93	-0.26	0.031	22
	2	-2.65	-5.86	0.56	0.083	4
	3	-2.80	-6.32	0.73	0.092	4
	4	-2.92	-6.70	0.87	0.100	4
Bayesian	1	-2.59	-6.34	1.19	0.066	-
	2	-2.65	-6.66	1.16	0.068	-
	3	-2.79	-6.95	1.39	0.070	-
	4	-2.92	-7.18	1.53	0.070	-

Chapter 10

Conclusions

"The more you know, the more you know you don't know."
- Aristotle[1]

Contents

10.1 Summary and Original Contributions

Principal conclusion: Through extensive evaluation and simulations the thesis demonstrates the half-t is the prior of choice for estimating uncertainty of variance estimators in routine analysis of analytical and process variance components studies. The coverage of 95% credible intervals for variance components and total variance or SD is 93% (approximately) or above, across a range of realistic scenarios.

Industrial studies are often performed to estimate sources of variation e.g. to assess, improve and quantify analytical (measurement) or process capability. These are particularly important in the pharmaceutical industry where there is a need to demonstrate to regulatory authorities that processes are under control and meet the required quality. These variability studies and their issues were introduced in Chapters 1 and 2. This thesis has investigated methods for industry to use routinely to quantify the uncertainty of variance estimates, especially for small studies. Understanding the uncertainty is important so that efforts to improve capability are directed most effectively, and so that

[1]This saying is typically attributed to Aristotle - possibly derived from his work **Metaphysics**

appropriate assurance of the measurement or process capability is ascertained. Complications of the industrial studies include: their small size, in particular the small number of levels of random factors incorporated into the design; the number of sources of variation included; the sources of variation are often small, perhaps negligible; designs may include nested and/or crossed factors; the outcomes of interest are not just variance components but other functions including sums and ratios. Though frequentist methods exist for providing confidence intervals for variance components in some situations, as discussed in Chapter 3, a Bayesian approach was thought most likely to provide an effective methodology for routine use which accommodates the complexities that these small studies exhibit. It is also noted that, though not a subject of this thesis, a Bayesian approach also provides a framework for incorporating historical information through an informative prior.

The research has sought to find prior(s) that can be used routinely and also are likely to satisfy regulatory authorities that the choice of prior has not unduly reduced the estimates of uncertainty to the extent of damaging the coverage. Six families of priors (described in Chapter 5) were chosen to be investigated from consideration of the literature, the software typically in use in the pharmaceutical industry and my idea that often an analytical method or process will have an intrinsic level of variance. The latter provides a benchmark for either the choice of prior itself (one prior investigated made use of this by being based on the intra-class correlation coefficient) or the parameter for a prior (for example, the scale parameters for the half-t and half-Cauchy priors for the higher-level variances were based on the existing knowledge of the intrinsic variance). Vague or mildly informative priors were assessed against the objective that the 95% credible intervals (CrIs) should ideally be consistent with a true coverage of at least 0.95 (subject to simulation error) for variances in the range of interest. The range of interest was chosen to be $\frac{1}{12}$ to 4 times the intrinsic variance. In practice an estimated coverage greater than 0.94 was sought to allow for the simulation error. In addition the estimation of credible intervals for total variance and SD was considered of primary importance (as the total variance/SD are typically reported to the regulatory authorities when a method is registered), with those for the variance components being secondary. An extensive simulation study has been performed covering a variety of scenarios based on a two-way variance component study (two higher-level nested random effects in addition to the bottom-level effect) as described in Chapter 4. The Gelman-Rubin convergence statistics and effective sample size were examined to check for any lack of convergence and that an adequate number of posterior samples were taken. Literature on assessing priors for estimating uncertainty of variance component estimators is limited and concentrates on one-way variance component (two-level) studies, as McNeish and Stapleton (2016a) describe. None have been seen that perform extensive simulation evaluations for priors for the two-way variance component study. Thus the work in this thesis is novel, in evaluating the two-way variance components model and in the range of priors

evaluated. McNeish and Stapleton (2016b) state that "To date, although analytical arguments for half-t and half-Cauchy have been made (e.g. Polson and Scott (2012), the performance (both absolute and relative to other priors) of these recommendations has not been systematically assessed."

From the extensive evaluation performed, a prior from the half-t family (originally proposed by Gelman (2006)) with 3 degrees of freedom on the higher-level variance component SD (σ_h) is recommended. This achieved estimated coverage ≥ 0.929 for the total variance and SD in all the situations evaluated, and coverage greater than 0.94 for all three variance components in almost all situations (unfortunately no prior was found with coverage ≥ 0.94 for all situations). The half-t scale parameter is recommended to be set to $5 \times \sqrt{k}$ where k is the best guess of what the variance component might be, though in investigations coverage of CrIs was found to be insensitive to varying the value of k. An alternative is to set it to the intrinsic variance for a process study, or for an analytical precision study, so that the variance components for the random effects that form part of the intermediate precision sum to the repeatability variance. The simulations incorporated a proper uniform on the residual variance where the upper limit was chosen to be $5 \times \sqrt{k}$. However, as Gelman (2006) noted there is often sufficient information at this level that any reasonably vague prior may work satisfactorily. The half-t prior achieved coverage greater than 0.94 for 95% highest posterior density intervals based on σ^2 or σ for total variance/SD for all scenarios investigated. Coverage greater than 0.94 was also obtained for the top-level and residual variances. For the mid-level variance component, coverage greater than 0.94 was achieved in all but one of the scenarios investigated, where coverage was in the range 0.929-0.939. A sensitivity analysis was undertaken for that scenario and for the investigations performed the coverage did not drop further. Given that, the primary goal for coverage was achieved (total variance and SD), the secondary goals were almost achieved (variance components), the sensitivity analysis did not identify a further drop in coverage, and in the absence of a prior performing better, the half-t prior is considered acceptable to be recommended for routine use. It is noted that though the desired level of coverage was not achieved it was better than that seen in the literature for many priors (for example Browne and Draper (2006) coverage was around 89% and 91% for the higher-level variance component for the smallest size study). Lambert (2006)'s response to Browne and Draper (2006) commented "It is clear is that for any Bayesian hierarchical model involving a small number of units, the role of the prior distribution for the hierarchical variance parameters is crucial and that there is unlikely to be an 'off-the-shelf' vague prior distribution suitable for all scenarios". This thesis has not attempted to provide a recommended prior for "all scenarios", but within the boundaries set for the investigation, it is proposed that the half-t prior can be routinely used.

As well as the recommendation of the half-t prior for routine use supported by extensive simulations other findings and original contributions were made as follows:

- The half-t prior based on 3 degrees of freedom (HT3) is recommended over the half-Cauchy (HCY - which is a half-t with 1 degree of freedom) since, though the credible intervals and coverages achieved are very similar, the half-t with 3 degrees of freedom had better results for the Gelman-Rubin convergence statistic in general and required fewer burn-in samples.
- The proper uniform prior on σ (UNI) often had shorter CrIs than half-t priors. However, coverage was less than 0.94 (in range 0.921-0.939) for the total variance and the highest-level variance component for two scenarios when σ_b^2=24, with the CrI often being above the true value. For other scenarios the coverage was closer to 0.95 than that for the HCY and HT3 priors. Gelman (2006) recommended starting with a non-informative uniform prior density on σ_h and expected that the uniform prior would "generally work well unless the number of groups is low (below 5, say)" (his paper compared group sizes of 3 and 8 for σ_h, with the half-t working well for 8 but thought unnecessary). However, in this study when the number of groups at the highest-level random effect was 6 the half-t was preferred to maintain coverage, thus suggesting that it would be preferable to use a half-t prior routinely for small studies. In addition, use of a half-t prior instead of a proper uniform prior avoids the hard cut-off of the latter (seen in the results in Section 8.3.2), a precaution that seems sensible in case your prior belief is substantially inconsistent with the data.
- Jeffreys' prior (JEFF) has not been recommended by the literature because it relies on the design and can lead to an improper posterior (though not for the designs investigated here). However, the pharmaceutical industry has been using it as it is a convenient prior in PROC MIXED in SAS/STAT® software (SAS Institute Inc.: SAS/STAT, 2011)[1]. The research has shown that the coverage can fall below 90% in some situations for the mid-level variance component and in some situations the credible intervals appear practically different from those obtained for the recommended half-t prior. It was also seen that the coverage of the credible intervals for one variance component could be strongly dependent on the true value of another variance component. This dependence was less for the half-t and uniform priors which are set independently on the variance components.
- Vague inverse gamma priors placed on linear combinations of variance components gave fairly comparable results to Jeffreys' prior (note that IG(0.001,0.001) is an approximation to Jeffreys') and are not recommended for similar reasons.
- The documentation for SAS PROC MIXED suggests use of inverse gamma priors on linear combinations of variance components for informative priors. The investigations for this thesis found that it was not easy to ensure that the prior is not over-informative if interest is in the variance components. Those investigated

[1]SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

were found to be too informative to achieve the desired coverage and none were satisfactory.

- The FLAT prior available in SAS PROC MIXED (improper uniform on variance components) is not recommended as it has an improper posterior when the highest-level random effect has only 3 levels, coverage was less than 0.9 for total SD for one scenario and where the desired coverage was achieved, the upper credible limits were often particularly high compared to other priors evaluated.
- It is recommended that highest posterior density CrIs for variance components based on σ^2 (hpdv interval) or σ (hpds interval) metrics are used in preference to percentile CrIs as the latter generally gave poorer coverage. Note it has been noticed in the literature that percentile intervals are often used - perhaps to align the credible interval type used for both fixed and random effects, or to be able to align a one-sided interval to the two sided interval, or perhaps because the limits can be transformed to apply on either the variance or SD metric. Generally hpdv intervals provided higher coverage than hpds intervals. It was found for the uniform prior that if good coverage was obtained for the variance on the σ^2 metric, this did not necessarily correspond to good coverage for the SD on the σ metric. Thus even though the UNI prior is placed on the σ metric, and the σ might be of interest, better coverage can be gained by obtaining the hpd CrI for the σ^2 estimate and then taking the square root to obtain the CrI for σ. Highest posterior density CrIs using logged values did not offer advantages.
- In PROC MIXED sampling from the posterior stops when either the proposal distribution has insufficient probability of positive values or the proposal distribution is sufficiently different from the posterior distribution. A novel method was implemented to modify the parameters of the proposal distribution subject to the constraints of the software to overcome this issue.
- A bug in PROC MCMC SAS version 9.3 TS1M2 (see problem note SAS Institute Inc.: PN48940 (2013)) was identified as causing erroneous results for the two-way variance component study though correct results were obtained for one-way variance component study.
- A scale based on the intraclass correlation coefficient is proposed for visualising credible intervals for variance components so that relevant features can be seen, unlike the linear scale for many examples.
- For simulation studies involving vague priors and small studies, a computing-resource effective strategy for calibration studies for the coverage of credible intervals is to use PROC MIXED with Jeffreys' prior in SAS and then follow up with datasets critical to determining coverage with priors requiring the more computer-intensive MCMC Gibbs algorithm implemented in WinBUGS.

Some of the advantages of applying a Bayesian approach and the recommendation that

the half-t prior is used routinely were illustrated in Chapter 9 in the analysis of example datasets e.g. it was seen that the Bayesian approach avoids a discontinuity in fixed effect testing as would typically happen with a frequentist approach.

Possible Limitations

- **Design of variance component study evaluated**
  The thesis has only investigated balanced two-way variance component study designs (nested design). In practice designs will often have more than two higher-level variance components, may have crossed factors and may be unbalanced. Whilst the simulations only covered studies with three variance components, it is expected that the results for CrIs for variance components would extend to nested studies with more variance components. For example, suppose there were four random effects A, B, C and D where A is the highest and D is the bottom. To estimate credible intervals for the variance components for A and B, for each level of factor C the results over factor D could be summed, converting it to a two-way variance component design. Similarly, to estimate the credible interval for the variance component for D, a new factor AB formed from the generalised factor of A and B again converts the problem to a two-way variance component design. However, for the total variance it is not clear what would happen to the coverage if the design had many more variance components. Unbalanced designs were not investigated but Browne and Draper (2006) (Table 6) showed that for Bayesian analyses unbalanced designs gave longer interval lengths than the balanced designs and in all but one case this led to increased coverage, though it is acknowledged that this is for a one-way design. Only a nested design has been evaluated. The methodology can be implemented for crossed factors but coverage has not been explored. For balanced crossed designs it seems reasonable that coverage for a variance component is more likely to be affected by factors nested within the random effect or factors nesting it, than those with which it is crossed. However, this may not be the case for unbalanced designs.
- **Range of scenarios evaluated**
  It is impossible to evaluate every combination of number of levels for each random effect and all values the variance components might take within the ranges for which good coverage was required. Consider the middle value of three variance component values explored. When the proportion of CrIs where the lower limit is higher than the true value, or the upper limit is lower than the true value, is highest at this value it may be that a higher proportion occurs at another value within the range of interest, in which case the estimate of minimum coverage across the range would be an overestimate. Thus there is an assumption that any overestimate is small.

- **Sensitivity to parameter choice for the half-t prior**
  The half-t prior with one and three degrees of freedom gave similar results. The effect of increasing the degrees of freedom has not been investigated generally (when dfs are infinite this corresponds to the half-normal distribution). However, as the half-t prior for one situation has coverage less than 0.94 with the CrI interval being too frequently lower than the true value, increasing the dfs which results in a less vague distribution seems likely to make the problem worse. For one situation the half-t scale parameter was varied from 5 to 24 and the coverage differed by at most one in the third decimal place indicating the robustness of the results to the choice of scale parameter.
- **Sensitivity to parameter choice for the other priors investigated**
  The half-t was chosen over the uniform prior since, in all but one case, it achieved coverage greater than 0.94 whereas the uniform prior had more situations where coverage of 0.94 was not achieved. In most cases the latter was because the CrI was too frequently above the true value. It might be thought that it would be desirable to reduce the upper limit of the proper uniform prior to avoid this, but as there were also situations where the CrI was too frequently below the true value, this seems unlikely to result in the uniform prior being preferred to the half-t.
- **Evaluation of fixed effects**
  In the model evaluated the only fixed effect was the overall mean. However, it is anticipated that the coverage of the half-t prior will be similar when other fixed effects are included in the model provided the design is sufficient to estimate these and the variance components.

In addition to the potential limitations discussed above it is noted that the focus of the thesis is restricted to the uncertainty of estimators for estimation of variance components and sum of variance components; random effects with normal distributions; and models without random covariates (where priors on covariance matrices would need consideration).

Some comments on the findings compared to those in the literature

Browne and Draper (2006) investigated the use of an inverse gamma prior or a proper uniform prior on the higher-level variance component as well as likelihood based-methods for a two-level (one-way) variance component model and found that both approaches had difficulties achieving nominal coverage in small samples (around 91%). In this thesis it is seen that the half-t prior (and indeed the uniform prior) on the SD attained higher coverage than this; it is noted that Browne and Draper suggested that uniform priors on the standard deviation might have better repeated-sampling characteristics than those they had chosen.

In Chapter 3 it was reported that Ionan et al. (2014) found in favour of frequentist methods for less than 8 levels of a higher-level random effect. However, they had used inverse gamma priors on the variances or improper uniform priors on the variance component standard deviations in their Bayesian analyses. These are not the priors evaluated in this thesis and as Gelman (2006) stated the inverse gamma prior is not recommended and the uniform prior has a disagreeable miscalibration towards positive values.

My results on the coverage of CrIs using the half-t prior are broadly aligned with the results obtained by Quiroz and Baumgartner (2017) who compared Bayesian and frequentist approaches for one-way variance components model and found that the half-Cauchy prior maintained coverage for small studies. In their study however, they also found the Jeffreys' reference prior maintained coverage whilst, in the extensive study performed here for the two-way model, this was not found to be the case for all situations. As others have found, they found the inverse gamma prior on the variance component did not maintain coverage for small studies.

Gelman (2006) suggested using a scale parameter set to a value that is high but not off the scale. He used a value of 25 when the plausible range of the SD was less than 50. I have used around 9 when the SD for the higher-level variance component was expected to be less than 5. However, the coverage results were found to be insensitive to the scale parameter when this was explored for one situation.

The results of the simulations in this thesis support Polson and Scott (2012)'s deduction that "The half-Cauchy ... performs well near the origin, but does not lead to drastic compromises in other parts of the parameter space", albeit for a different object of interest.

10.2 Further work

Connected to Main Focus of the Thesis

The following work is suggested to further support the arguments made in the discussion of the limitations of the work and the half-t being the prior of choice for analytical and process variation studies.

- Further evaluation of design sizes for which the CrIs for keg variance for the half-t prior does not reach the nominal coverage, for example, sensitivity of the coverage to changing the number of portions per keg or the number of batches.
- Evaluate more values for the variance components, for example, to obtain the coverage for the keg variance for the design with 6 batches, 6 kegs per batch and 16 portions per keg for true keg variances between the values already investigated (0.5 and 24).
- Evaluate the half-t prior for unbalanced designs and for crossed designs.

- Assess the coverage of the CrIs for total variance for designs with many variance components for the half-t prior.
- Assess coverage for CrIs for other quantities of interest derived from variation studies e.g. ratios of variances, tolerance intervals, predicted % falling outside specifications.

Widening the Scope

Widening the scope from that of the thesis but of potential relevance to the analysis of variation studies the following aspects could benefit from further investigation:

- **Estimates of variances**
  The purpose of the thesis is to provide estimates of the uncertainty of the estimators for the variance components and total variance. I question the appropriateness of providing point estimates of variance components as: in most cases the posterior distribution should be used to provide uncertainty for other quantities of interest which are based on the variance components; and the large uncertainty associated with the estimators for the small studies suggests that little focus should be placed on the estimate itself. However, in practice estimates will be demanded as a summary statistic e.g. for comparison with the results from other studies, products, analytical methods. From the results seen in Section 8.5 and the results seen in Browne and Draper (2006) the median is to be preferred to the mean. Browne and Draper (2006) also investigated the mode and found that for the inverse gamma prior the median was to be preferred based on relative bias but for the uniform prior the mode was to be preferred. Under certain conditions there is equivalence between REML estimates and the mode (Tsai and Hsiao, 2008). Some have suggested that priors are better chosen to avoid zero estimates and thus the mode should be avoided unless a prior density which is zero at zero is used (Chung et al., 2013). In contrast Simpson et al. (2017) advocating penalizing complexity states that "a prior forces overfitting if the density of the prior is zero at the base model". Thus investigation of the preference for the summary measure to be the median or mode could be useful. Baldwin and Fellingham (2013) discuss that the optimality of an estimator is related to the loss function, with the mean being optimal for a squared error loss function, the median being optimal for an absolute error loss function and the mode being optimal under the 0-1 loss function, which they say is most appropriate for discrete parameter spaces. So this would imply the median is to be preferred to the mode. However, the squared error and absolute error loss functions will typically be placed on the linear metric which I argue is not the best metric when estimators have skewed distributions and thus other loss functions and estimators may be more appropriate.
- **Fixed effects**
  The focus of the thesis is the estimation of the variance components. However, fixed

effects and their associated intervals are often also of interest e.g. for equivalence assessment. In this thesis variance components are restricted to be non-negative. It is noted that Stroup and Littell (2002) recommended that unbounded REML variance component estimates should be used when testing fixed effects to avoid bias (it was seen in Chapter 2 that unbounded REML is often a problem for the unbalanced datasets investigated). Investigation of bias and coverage of intervals for fixed effects in a variance component analysis with the half-t prior would be of interest, though given results such as those of Li and Stern (1997) and McNeish and Stapleton (2016b), it is expected that bias will be reasonably small and the intervals will generally have higher coverage than nominal.

- **Attitude towards coverage**

  The criteria for evaluating the priors in this thesis is that the coverage of the CrIs should be nominal or greater in order that they are correctly or conservatively calibrated. In a regulatory environment this will give assurance to regulators that satisfactory control of processes or precision of analytical methods is obtained. This perspective is in agreement with that by Quiroz and Baumgartner (2017) and Stegmueller (2013), the latter stating "While confidence interval non-coverage is undesirable in itself, it is the direction of this bias which is cause for concern. Without exception, ML produces confidence intervals that are too short, so that hypothesis tests are anti-conservative". However, others take a different perspective, Gustafson et al. (2006) proposed a prior aiming to prevent over estimation of the higher-level variance component, Simpson et al. (2017) propose penalising complexity. An evaluation of the choice of prior if the aim had been for a prior close to nominal coverage would be useful (an appropriate loss function would need to be decided upon as discussed in Section 7.1).

- **Other areas for exploration**

  The half-t prior did not quite achieve the nominal coverage in all cases - would a slightly more flexible prior achieve this? For example, Perez et al. (2017) propose a wider family - the Scaled Beta2 distribution. Gelman (2006) suggested placing a prior on the scale parameter used in the half-Cauchy or half-t prior (though proposed for when there are many variance components).

  Some variation studies in industry have random covariates either of interest in themselves or as nuisance variables. In this case priors are required on the covariance matrix. In the past an inverse Wishart has typically been used. However this suffers from the problems seen with an inverse gamma prior on the variance components. Alternative proposals have been made in the literature e.g. Huang and Wand (2013), Chung et al. (2015) and Demirhan and Kalaylioglu (2015). Evaluation and recommendations for the priors to be placed on the covariance matrix in routine analysis of small variation studies is desirable.

  The simulations performed in this thesis have been very time consuming, especially those performed using the MCMC algorithms in WinBUGS. In order to

make it feasible to perform such simulation studies more frequently it is desirable that the computing efficiency is improved. There are many efforts in the literature to improve the efficiency of applying the Bayesian approach, both in improving the MCMC efficiency and alternatives such as using integrated nested Laplace approximations (Rue et al., 2009). However, given that the independence chain algorithm used in SAS was seen to be much less time consuming, it is worth considering whether it together with a truncated inverse gamma distribution proposal distribution, can be used more widely.

Application area

Information on analytical or process variability is used in industry to assess whether the analytical and process variation meets requirements. The focus of the thesis has been on the estimation of uncertainty of variance estimators in analytical and process variability studies. There is external influence to consider the uncertainty, for example the USP stimuli article (Barnett et al., 2016), and a move towards Bayesian methodology (Lira, 2016). Having the ability to, with one approach, obtain estimates of this uncertainty for all types of such studies, in conjunction with the changing external environment is expected to:

1. Enable the uncertainty to be taken into account when making subsequent decisions e.g. deciding whether more data is desirable or whether further process or analytical improvement should be undertaken.
2. Encourage the use of larger studies where total variance and/or variance component estimation is important and the typical designs used are too small. For gage R&R studies Vardeman and VanValkenburg (1999) promoted the importance of providing intervals to accompany point estimates for a two-way gage R&R study and demonstrated extremely wide confidence intervals for a fairly typical design. It is expected that this will lead on to a prospective interest in the size of design required. For example, Minitab 17 (2018) provides useful information on the effect of sample size on variance component estimates in two-way gage R&R studies. However, the two-way gage R&R study is a very specific design and most complex designs are likely to require extensive simulations which will be time consuming and practically infeasible.
3. Identify that there are some situations where the designs that would be required to adequately estimate the total variance and variance components are very large when just based on the data and a vague prior such as the half-t, and not commensurate with the risk relative to the impact of the variance being slightly larger than specified. Even when testing a fixed effect the design required may be too large. This will focus attention on alternative approaches such as using more informative priors. Though in a different application area and for fixed effects, van de Schoot

et al. (2015) stated "Only when Bayesian analysis, in conjunction with informative priors, was used power increased to acceptable levels". If there is minimal chance of any really big values Stan (2017) say a half-normal might be considered. Gustafson et al. (2006) propose prior distributions for variance components which deliberately give weight to small values. Neither seem appropriate in the regulatory environment of the pharmaceutical industry where we need to give assurance that the analytical or process variation is sufficiently small, without justification for these priors. Simpson et al. (2017) propose informative priors based on penalising complexity. They define a model component to be a flexible extension of a base model and state it is important to consider the joint effect of the priors. To avoid overfitting they propose priors which i) control the overall variance and ii) control how much each term contributes to it. This could be aligned to analytical precision where a rule of thumb is that the intermediate precision is twice that of the repeatability. They say that a prior forces overfitting if the density of the prior is zero at the base model. A prior avoids overfitting if its density is decreasing with a maximum at the base model. The half-t prior satisfies this property on both the SD and variance metric.

4. Focus interest on incorporating historical information for the type of analytical method or type of process into the prior and/or statistical analysis, especially when fixed effects are of prime importance e.g. equivalence studies. A possible approach might be, in a situation where variability is thought to be similar to that seen previously, to first assess whether the random effects are consistent with historical information. If they are not, then more data would need to be collected. If they are consistent, then estimate variability on the basis of a prior based on historical information combined with the data itself. The issue would then be how to weight the historical information relative to that from the study itself.

Appendices

Appendix A

Example Datasets

Table A.1 contains the data for a typical ruggedness study, Example 1, plotted in Figure 1.4.

TABLE A.1: **Data for a typical ruggedness study- Example 1**

Site	Analyst	Instrument	Column	Assay
Site A	1	1	1	99.680
Site A	1	1	1	99.627
Site A	1	1	2	99.270
Site A	1	1	2	99.424
Site A	1	2	1	100.391
Site A	1	2	1	100.170
Site A	2	1	2	99.299
Site A	2	1	2	99.222
Site A	2	2	1	100.970
Site A	2	2	1	100.614
Site A	2	2	2	99.729
Site A	2	2	2	99.319
Site B	3	3	3	100.426
Site B	3	3	3	101.425
Site B	3	3	4	98.993
Site B	3	3	4	99.872
Site B	3	4	3	101.015
Site B	3	4	3	100.624
Site B	4	3	4	100.844
Site B	4	3	4	100.306
Site B	4	4	3	100.126
Site B	4	4	3	100.781
Site B	4	4	4	101.085
Site B	4	4	4	101.716

Table A.2 contains the data used in Examples 3 to 6 as described in Section 2.1.1.

TABLE A.2: **Data for Examples 3 - 6**

Site	**Analyst**	**Instrument**	**Column**	**Exclusion Indicator**	**Examples 3, 4**	**Examples 5, 6**
1	1	1	1	0	26.1482	15.2735
1	1	1	2	0	25.3342	18.6503
1	1	2	1	0	24.7832	14.7861
1	1	2	2	1	23.1596	21.9481
1	2	1	1	1	30.4698	19.3294
1	2	1	2	0	21.7685	19.6912
1	2	2	1	0	21.9764	19.5200
1	2	2	2	0	25.0900	21.2964
2	3	3	3	0	17.2705	23.4274
2	3	3	4	0	16.3489	25.7537
2	3	4	3	0	17.7218	20.3414
2	3	4	4	1	17.9507	19.4016
2	4	3	3	1	17.8107	13.5417
2	4	3	4	0	21.2281	13.8848
2	4	4	3	0	17.2407	22.3028
2	4	4	4	0	18.8356	17.9183

Table A.3 contains an example dataset for the batch sampling design which was plotted in Figure 4.1 and analysed in Section 9.1.

TABLE A.3: **Example dataset for batch sampling design**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
1	1	1	96.128	2	1	1	96.735
1	1	2	96.391	2	1	2	104.212
1	1	3	95.522	2	1	3	100.734
1	1	4	95.442	2	1	4	99.835
1	1	5	97.531	2	1	5	96.971
1	1	6	95.118	2	1	6	97.774
1	1	7	100.166	2	1	7	100.350
1	1	8	93.999	2	1	8	105.848
1	1	9	96.630	2	1	9	98.061
1	1	10	96.998	2	1	10	101.814
1	1	11	96.762	2	1	11	96.472
1	1	12	96.133	2	1	12	96.820
1	1	13	97.653	2	1	13	99.813
1	1	14	96.209	2	1	14	104.267
1	1	15	99.611	2	1	15	101.664
1	1	16	96.674	2	1	16	98.915
1	2	1	94.583	2	2	1	97.664
1	2	2	99.215	2	2	2	94.666
1	2	3	96.727	2	2	3	94.079
1	2	4	95.931	2	2	4	97.721
1	2	5	95.854	2	2	5	97.506
1	2	6	91.412	2	2	6	97.788
1	2	7	97.492	2	2	7	92.827
1	2	8	94.366	2	2	8	99.497
1	2	9	99.338	2	2	9	97.745
1	2	10	97.168	2	2	10	95.283
1	2	11	93.113	2	2	11	95.874
1	2	12	96.313	2	2	12	97.625
1	2	13	99.937	2	2	13	98.675
1	2	14	89.958	2	2	14	97.088
1	2	15	97.206	2	2	15	97.958
1	2	16	96.079	2	2	16	98.259

TABLE A.3: **Example dataset for batch sampling design**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
3	1	1	97.765	4	1	1	94.808
3	1	2	101.309	4	1	2	97.472
3	1	3	100.024	4	1	3	99.454
3	1	4	100.145	4	1	4	100.169
3	1	5	98.596	4	1	5	98.829
3	1	6	99.957	4	1	6	95.731
3	1	7	102.610	4	1	7	96.343
3	1	8	101.749	4	1	8	99.483
3	1	9	100.864	4	1	9	96.491
3	1	10	105.515	4	1	10	95.934
3	1	11	98.955	4	1	11	97.808
3	1	12	97.932	4	1	12	104.002
3	1	13	94.938	4	1	13	100.607
3	1	14	102.195	4	1	14	96.805
3	1	15	98.888	4	1	15	101.255
3	1	16	99.535	4	1	16	99.593
3	2	1	95.465	4	2	1	98.918
3	2	2	102.783	4	2	2	101.058
3	2	3	100.485	4	2	3	105.729
3	2	4	104.958	4	2	4	100.578
3	2	5	97.906	4	2	5	98.577
3	2	6	102.145	4	2	6	103.446
3	2	7	99.252	4	2	7	100.989
3	2	8	100.741	4	2	8	97.733
3	2	9	100.566	4	2	9	98.811
3	2	10	100.682	4	2	10	98.206
3	2	11	101.310	4	2	11	100.695
3	2	12	101.036	4	2	12	100.534
3	2	13	102.977	4	2	13	99.196
3	2	14	98.799	4	2	14	96.905
3	2	15	102.150	4	2	15	100.415
3	2	16	100.604	4	2	16	103.678

TABLE A.3: **Example dataset for batch sampling design**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
5	1	1	93.141	6	1	1	103.095
5	1	2	105.353	6	1	2	103.139
5	1	3	101.179	6	1	3	99.546
5	1	4	99.713	6	1	4	101.321
5	1	5	95.227	6	1	5	97.614
5	1	6	104.557	6	1	6	104.384
5	1	7	98.202	6	1	7	101.872
5	1	8	98.390	6	1	8	102.585
5	1	9	101.997	6	1	9	100.506
5	1	10	102.349	6	1	10	101.595
5	1	11	98.980	6	1	11	97.533
5	1	12	99.735	6	1	12	98.251
5	1	13	97.541	6	1	13	103.142
5	1	14	101.985	6	1	14	101.983
5	1	15	100.130	6	1	15	96.150
5	1	16	100.824	6	1	16	101.591
5	2	1	100.954	6	2	1	96.311
5	2	2	100.898	6	2	2	101.184
5	2	3	97.199	6	2	3	102.615
5	2	4	96.614	6	2	4	100.928
5	2	5	99.994	6	2	5	102.380
5	2	6	98.033	6	2	6	102.828
5	2	7	98.805	6	2	7	96.676
5	2	8	95.364	6	2	8	96.098
5	2	9	95.707	6	2	9	101.447
5	2	10	98.687	6	2	10	98.290
5	2	11	97.537	6	2	11	101.117
5	2	12	93.717	6	2	12	99.848
5	2	13	97.272	6	2	13	103.494
5	2	14	100.655	6	2	14	103.443
5	2	15	97.334	6	2	15	98.709
5	2	16	97.653	6	2	16	98.004

Appendix B

REML

Let Y be an n-dimensional random vector which is multivariate normal $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma})$ then the probability density function is:

$$f_Y(\mathbf{y}) = \frac{1}{(2\boldsymbol{\pi})^{n/2}|\boldsymbol{\sigma}|^{1/2}} \exp\Big(-\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\Big),$$

where $|\boldsymbol{\sigma}|$ is the determinant of $\boldsymbol{\sigma}$.

For the log likelihood we obtain:

$$-2l(\boldsymbol{\phi}, \boldsymbol{\mu}; \mathbf{y}) = n\log(2\boldsymbol{\pi}) + \log|\boldsymbol{\sigma}(\boldsymbol{\phi})| + (\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$$

For the mixed model in section 1.3 we had:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}),$$

where $\mathbf{V} = Var[\mathbf{Y}] = \mathbf{ZGZ}' + \mathbf{R}$ and p is the rank of $\mathbf{X}$ (i.e we have p fixed effects).

SAS Institute Inc.: The MIXED Procedure (2013) describe that REML estimation is equivalent to likelihood estimation for transformed data as follows. Instead of the log likelihood of $\mathbf{Y}$, the log likelihood of $\mathbf{KY}$ is considered where the matrix $\mathbf{K}$ is chosen so that $E[\mathbf{KY}] = 0$. This gives:

$$-2l_R(\boldsymbol{\phi}; \mathbf{KY}) = \log|\mathbf{KV}(\boldsymbol{\phi})\mathbf{K}'| + \mathbf{y}'\mathbf{K}'\mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{Ky} + c_R$$

Choosing $\mathbf{K}$ to be n-p independent rows from the matrix $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gives

$$-2l_R(\boldsymbol{\phi}; \mathbf{y}) = \log|\mathbf{V}(\boldsymbol{\phi})| + \log|\mathbf{X}'\mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{X}| + \Big(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\phi})\Big)' \mathbf{V}(\boldsymbol{\phi})^{-1} \Big(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\phi})\Big) + c_R,$$

where $c_R = (n-p)\log(2\pi)$; and p is the rank of $\mathbf{X}$.

REML is generally preferred over maximum likelihood for several reasons as described by McCulloch et al. (2008). The REML solutions have the attractive property that, for balanced data, they have the same values as the ANOVA estimators and these are minimum variance unbiased under normality. The REML estimators take into account the degrees of freedom for the fixed effects in the model and are invariant to the value of $\boldsymbol{\beta}$. McCulloch et al. (2008) also suggests that the REML estimators do not seem to be as sensitive to outliers in the data.

Appendix C

Transforming Parameters

Contents

When a prior distribution is defined on one scale the following equations are useful to construct the distribution on the alternative scale of interest.

C.1 Transforming a Single Parameter

If a prior on σ is $f(\sigma)$ then the prior for σ^2 is $\frac{1}{2\sigma}f(\sigma)$
If a prior on σ^2 is $f(\sigma^2)$ then the prior for σ is $2\sigma f(\sigma^2)$ as follows:
Let $Y = g(X)$. Then if function g is monotonic the resulting density function is:

$$f_Y(y) = \left|\frac{1}{g'(g^{-1}(y))}\right| f_X(g^{-1}(y)). \tag{C.1}$$

Here g^{-1} denotes the inverse function and g' denotes the derivative.

If $X = \sigma$ and has prior $f(\sigma)$ then for $Y = \sigma^2 = X^2 = g(X)$ then $g'(g^{-1}(y)) = g'(y^{\frac{1}{2}}) = \frac{1}{2}y^{-1/2} = \frac{1}{(2y^{\frac{1}{2}})}$, giving a prior on $Y = \sigma^2$ of $\frac{1}{2y^{\frac{1}{2}}}f(\sqrt{y}) = \frac{1}{2(\sigma^2)^{\frac{1}{2}}}f(\sqrt{\sigma^2}) = \frac{1}{2\sigma}f(\sigma)$.

If $X = \sigma^2$ and has prior $f(\sigma^2)$ then for $Z = \sigma = \sqrt{X} = g(X)$ then $g'(g^{-1}(z)) = g'(z^2) = 2z$, giving a prior on $Z = \sigma$ of $2zf(z^2) = 2\sigma f(\sigma^2)$.

C.2 Transforming Multiple Parameters

Assume $(x_1, x_2, ..., x_n)$ has pdf $f(x_1, x_2, ..., x_n)$. Let $y_1 = h_1(x_1, x_2, ..., x_n)$ $y_2 = h_2(x_1, x_2, ..., x_n)$ .... $y_n = h_1(x_1, x_2, ..., x_n)$

Then $(y_1, y_2, ..., y_n)$ has pdf

$$g(y_1, y_2, ..., y_n) = f(x_1, x_2, ..., x_n)J(x_1, x_2, ..., x_n|y_1, y_2, ..., y_n) \tag{C.2}$$

where J is the Jacobian $\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$ or $\begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix}^{-1}$

Appendix D

Jeffreys' Prior

Contents

For the mixed maximum likelihood model, and using the notation given in Section 1.3, McCulloch et al. (2008) (P176) gives the Information matrix as:

$$I(\boldsymbol{\beta},\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{X^T V^{-1} X} & 0 \\ 0 & \mathbf{\frac{1}{2}[tr[Z_i^T V^{-1} Z_j (Z_i^T V^{-1} Z_j)^T]]_{0\leq i,j\leq r}} \end{pmatrix}, \tag{D.1}$$

where r is the number of variance components.

D.1 Simple Normal Distribution

For $Y_m \propto N(\mu,\sigma^2), m = 1,...n$ then $\mathbf{V} = \sigma^2\mathbf{I_n}$ where $\mathbf{I_n}$ is an $\mathbf{n}$x$\mathbf{n}$ identity matrix. So $\mathbf{V^{-1}} = \frac{1}{\sigma^2}\mathbf{I_n}$, $r = 0$, $\mathbf{Z_0} = (1,1,1,...,1)_n$, and thus,

$\mathbf{\frac{1}{2}tr[Z_i^T V^{-1} Z_j (Z_i^T V^{-1} Z_j)^T]_{0\leq i,j\leq r}} = \mathbf{\frac{1}{2}tr[Z_0^T V^{-1} Z_0 (Z_0^T V^{-1} Z_0)^T]} = \mathbf{\frac{1}{2}tr[\frac{n}{\sigma^2}(\frac{n}{\sigma^2})^T]} = \frac{1}{2}[\frac{n}{\sigma^2}]^2$.

So Jeffreys' prior is given by $\pi(\sigma^2) \propto det(\mathbf{I}(\sigma^2))^{1/2} \propto \frac{1}{\sigma^2}$.

D.2 One-Way Nested Variance Components Model

D.2.1 Information Matrix and Jeffreys' Prior

Now consider a one-way variance components model with measurements taken from P portions from each of K kegs from one batch with two variance components: between keg variance component σ^2_{keg} and between portions variance $\sigma^2_{portion}$.

The Fisher Information matrix is given in Berger and Bernardo (1992) and substituting $\sigma^2_k{=}\tau^2$, $\sigma^2_p{=}\sigma^2$, $P{=}n$, $K{=}p$ (to use my notation) gives:

$$\mathbf{H_1}(\mu, \sigma^2_k, \sigma^2_p) = \begin{pmatrix} \frac{KP}{P\sigma^2_k+\sigma^2_p} & 0 & 0 \\ 0 & \frac{KP^2}{2(P\sigma^2_k+\sigma^2_p)^2} & \frac{KP}{2(P\sigma^2_k+\sigma^2_p)^2} \\ 0 & \frac{KP}{2(P\sigma^2_k+\sigma^2_p)^2} & \left[\frac{K(P-1)}{2\sigma^4_p} + \frac{K}{2(P\sigma^2_k+\sigma^2_p)^2}\right] \end{pmatrix}.$$

Assuming Jeffreys' prior for σ^2_k and σ^2_p ignores the part of the information matrix relating to μ then $\pi(\sigma^2_k, \sigma^2_p)$ is given by $\sqrt{|\mathbf{H_1}(\sigma^2_k, \sigma^2_p)|}$.

$$\begin{aligned} |\mathbf{H_1}(\sigma^2_k, \sigma^2_p)| &= \frac{KP^2}{2(P\sigma^2_k+\sigma^2_p)^2}\left[\frac{K(P-1)}{2\sigma^4_p} + \frac{K}{2(P\sigma^2_k+\sigma^2_p)^2} - \frac{K}{2(P\sigma^2_k+\sigma^2_p)^2}\right] \\ &= \frac{KP^2}{2(P\sigma^2_k+\sigma^2_p)^2}\frac{K(P-1)}{2\sigma^4_p} \\ &= \frac{K^2P^2(P-1)}{4\sigma^4_p(P\sigma^2_k+\sigma^2_p)^2}, \end{aligned},$$

and thus Jeffreys' prior is

$$\pi(\sigma^2_k, \sigma^2_p) \propto \frac{1}{\sigma^2_p(P\sigma^2_k+\sigma^2_p)}. \tag{D.2}$$

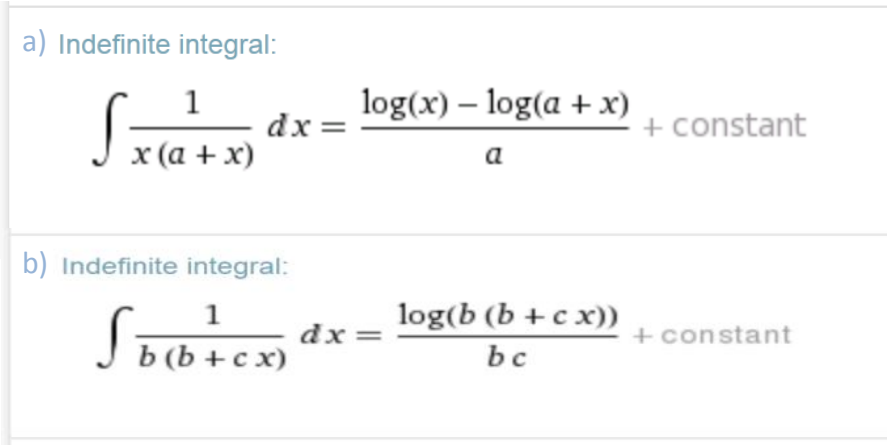
D.2.2 Density, Marginal and Conditional Distributions

For comparison with other priors it was desired to find the density function for the Jeffreys' prior and/or look at the marginal distribution for σ^2_{keg}. The density involves integrating Jeffreys' prior function over both σ^2_{keg} and $\sigma^2_{portion}$ and the marginal distribution involves integration over $\sigma^2_{portion}$.

Firstly I tried integrating over $\sigma^2_{portion}$. Figure D.1a) shows the solution to integrating over $\sigma^2_{portion}$ to obtain the marginal distribution for σ^2_{keg}, obtained from Wolfram|Alpha (2015), accessed 11$^{\text{th}}$ January 2015, where $x = \sigma^2_p$ and $a = P\sigma^2_k$.

However, we need to integrate from $x(=\sigma^2_p) = 0$ to ∞ and it is seen that the solution is undefined at $x = 0$ as follows. Taking the solution in Figure D.1 and substituting for x

FIGURE D.1: **Integration results from Wolfram for one-way model**



and a gives:

$$\begin{aligned}
\int_0^\infty \frac{1}{\sigma_p^2(\sigma_p^2 + P\sigma_k^2)} \mathrm{d}\sigma_p^2 &= \left[\log(\sigma_p^2) - \log(P\sigma_k^2 + \sigma_p^2)\right]_0^\infty \\
&= \left[\log\left(\frac{\sigma_p^2}{(P\sigma_k^2 + \sigma_p^2)}\right)\right]_0^\infty \qquad = \left[\log\left(\frac{1}{\left(\frac{P\sigma_k^2}{\sigma_p^2} + 1\right)}\right)\right]_0^\infty \\
&= \log\left(\frac{1}{\left(\frac{P\sigma_k^2}{\infty} + 1\right)}\right) - \log\left(\frac{0}{(P\sigma_k^2 + 0)}\right) \\
&= \log\left(\frac{1}{(0+1)}\right) - \log\left(\frac{0}{P\sigma_k^2}\right) \\
&= 0 - \log(0).
\end{aligned} \tag{D.3}$$

Thus the marginal distribution for σ^2_{keg} is undefined.

Secondly I try looking at the conditional Jeffreys' prior distribution of σ^2_{keg} given $\sigma^2_{portion}$. Using Figure D.1b) and letting x=σ_k^2, b=σ_p^2 and c=P the conditional Jeffreys' prior for $L \leq \sigma_k^2 \leq U$ is as follows:

$$
\begin{aligned}
f(\sigma_k^2|\sigma_p^2) &= \frac{\pi(\sigma_k^2,\sigma_p^2)}{f(\sigma_p^2)} \\
&= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)}}{\int_L^U \frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)}\mathrm{d}\sigma_k^2} \\
&= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)}}{\left[\frac{\log(\sigma_p^4+P\sigma_p^2\sigma_k^2)}{P\sigma_p^2}\right]_L^U} \\
&= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)}}{(\frac{log(\sigma_p^4+P\sigma_p^2U)}{P\sigma_p^2} - \frac{\log(\sigma_p^4+P\sigma_p^2L)}{P\sigma_p^2})}.
\end{aligned}
\tag{D.4}
$$

The denominator is undefined at $U = \infty$ and thus the conditional distribution is undefined.

Similarly the prior function is improper as a normalising constant obtained from integrating over σ_{keg}^2 and $\sigma_{portion}^2$ will be undefined as seen from Equations (D.3) and (D.4).

Thirdly I try examining the distribution of $\frac{\sigma_{keg}^2}{\sigma_{portion}^2}$. Letting $y_1 = \sigma_p^2$ and $y_2 = \frac{\sigma_k^2}{\sigma_p^2}$ and using equations (C.2) and (D.2) we have:

$$
\begin{aligned}
h\left(\sigma_p^2, \frac{\sigma_k^2}{\sigma_p^2}\right) = h(y_1,y_2) &= g(\sigma_p^2,\sigma_k^2)\left|\begin{matrix}\frac{\partial y_1}{\partial\sigma_p^2} & \frac{\partial y_1}{\partial\sigma_k^2}\\ \frac{\partial y_2}{\partial\sigma_p^2} & \frac{\partial y_2}{\partial\sigma_k^2}\end{matrix}\right|^{-1} \\
&= \frac{1}{\sigma_p^2(P\sigma_k^2+\sigma_p^2)}\left|\begin{matrix}1 & 0\\ \frac{-\sigma_k^2}{\sigma_p^4} & \frac{1}{\sigma_p^2}\end{matrix}\right|^{-1} \\
&= \frac{\sigma_p^2}{\sigma_p^2(P\sigma_k^2+\sigma_p^2)} \\
&= \frac{1}{\sigma_p^2\left(P\frac{\sigma_k^2}{\sigma_p^2}+1\right)}.
\end{aligned}
$$

It is shown that this is also undefined for both the marginal distribution and the distribution conditional on $\sigma_{portion}^2$ as follows.

For the marginal distribution:

$$
\begin{aligned}
h\left(\frac{\sigma_k^2}{\sigma_p^2}\right) = h(y_2) &= \int_0^\infty \frac{1}{\sigma_p^2\,(1+Py_2)}\,\mathrm{d}\sigma_p^2 \\
&= \left[\frac{\log(\sigma_p^2)}{(1+Py_2)}\right]_0^\infty,
\end{aligned}
$$

which is undefined at $\sigma_p^2 = \infty$ and $\sigma_p^2 = 0$.

For the conditional distribution:

$$\begin{aligned} h\left(\frac{\sigma_k^2}{\sigma_p^2}|\sigma_p^2\right) = h(y_2|\sigma_p^2) &= \frac{\frac{1}{\sigma_p^2(1+Py_2)}}{\int_0^\infty \frac{1}{\sigma_p^2\left(1+Py_2\right)}\,\mathrm{d}y_2} \\ &= \frac{\frac{1}{\sigma_p^2(1+Py_2)}}{\frac{1}{\sigma_p^2}\left[\frac{\log(1+Py_2)}{P}\right]_0^\infty} \\ &= \frac{\frac{1}{\sigma_p^2(1+Py_2)}}{\frac{1}{\sigma_p^2}\frac{(\log(1+P\infty)-\log(1))}{P}} \\ &= \frac{\frac{1}{\sigma_p^2(1+Py_2)}}{\frac{1}{\sigma_p^2}\frac{(\log(1+P\infty))}{P}}, \end{aligned}$$

which becomes 0 when the denominator is evaluated at $y_2 = \infty$.

All the above attempts to find proper probability distributions based on Jeffreys' prior for comparison with other priors failed. Thus the joint (improper) distribution was used to visualise Jeffreys' prior in Section 5.2.5.2 for the one-way variance components model but any comparisons with other priors should be treated with caution.

D.3 Two-Way Nested Variance Components Model

D.3.1 Information Matrix and Jeffreys' Prior

Searle (1970) derives the Information matrix for a maximum likelihood analysis (P514) of unbalanced and balanced two-way nested variance components designs.

The model for the measurement y_{ijk} made on α-level i (effect α_i), β-level j (effect β_{ij}) and random error term ϵ (effect ϵ_{ijk}) is:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk},$$

where μ is the overall mean,
$\alpha_i \sim N(0, \sigma_\alpha^2)$ for $i = 1, ...a$, $\beta_{ij} \sim N(0, \sigma_\beta^2)$ for $j = 1, ...c_i$,
$\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ for $k = 1, ...n_{ij}$, and all random variables are assumed to be independent.

Then the information matrix ($\mathbf{T}$) is given by:

$$\begin{pmatrix} t_{\alpha,\alpha} & t_{\alpha,\beta} & t_{\alpha,\epsilon} \\ t_{\alpha,\beta} & t_{\beta,\beta} & t_{\beta,\epsilon} \\ t_{\alpha,\epsilon} & t_{\beta,\epsilon} & t_{\epsilon,\epsilon} \end{pmatrix},$$

where the elements of $\mathbf{T}$ are defined as follows:

$$\begin{aligned}
m_{i,j} &= n_{i,j}\sigma_\beta^2 + \sigma_\epsilon^2 \\
A_{ipq} &= \sum_{\mathrm{j}=1}^{c_i} (n_{ij}^p / m_{ij}^q) \text{ for integers } p \text{ and } q, \\
q_i &= 1 + \sigma_\alpha^2 A_{i11}, \\
t_{\alpha,\alpha} &= \sum_{\mathrm{i}=1}^{a} (A_{i11}^2 / q_i^2), \\
t_{\alpha,\beta} &= \sum_{\mathrm{i}=1}^{a} (A_{i22} / q_i^2), \\
t_{\alpha,\epsilon} &= \sum_{\mathrm{i}=1}^{a} (A_{i12} / q_i^2), \\
t_{\beta,\beta} &= \sum_{\mathrm{i}=1}^{a} (A_{i22} - 2\sigma_\alpha^2 A_{i33}/q_i + \sigma_\alpha^4 A_{i22}^2 / q_i^2), \\
t_{\beta,\epsilon} &= \sum_{\mathrm{i}=1}^{a} (A_{i12} - 2\sigma_\alpha^2 A_{i23}/q_i + \sigma_\alpha^4 A_{i12} A_{i22} / q_i^2), \text{ and} \\
t_{\epsilon,\epsilon} &= \sum_{\mathrm{i}=1}^{a} (A_{i02} - 2\sigma_\alpha^2 A_{i13}/q_i + \sigma_\alpha^4 A_{i12}^2 / q_i^2) + (n_{..} - c_.)/\sigma_\epsilon^4.
\end{aligned}$$

For balanced data where $n_{ij} = n$ for all i and j and $c_i = c$ for all i, Searle (1970) shows that $m_{i,j}$, A_{ipq} and q_i simplify as follows (though Searle abbreviates $\sigma_\alpha^2, \sigma_\beta^2$ and σ_ϵ^2 to α, β and ϵ respectively):

$$\begin{aligned}
m_{i,j} &= n\sigma_\beta^2 + \sigma_\epsilon^2, \\
A_{ipq} &= \sum_{\mathrm{j}=1}^{c} \left(\frac{n^p}{(n\sigma_\beta^2 + \sigma_\epsilon^2)^q} \right) = \frac{cn^p}{(n\sigma_\beta^2 + \sigma_\epsilon^2)^q}, \\
q_i &= 1 + \sigma_\alpha^2 A_{i11} = 1 + \sigma_\alpha^2 \left(\frac{cn}{(n\sigma_\beta^2 + \sigma_\epsilon^2)} \right) = \frac{n\sigma_\beta^2 + \sigma_\epsilon^2 + \sigma_\alpha^2 cn}{n\sigma_\beta^2 + \sigma_\epsilon^2} = \frac{cn\sigma_\alpha^2 + n\sigma_\beta^2 + \sigma_\epsilon^2}{n\sigma_\beta^2 + \sigma_\epsilon^2},
\end{aligned}$$

and $\mathbf{T}$ simplifies to:

$$\begin{pmatrix} c^2n^2y & cn^2y & cny \\ cn^2y & n^2(x+y) & n(x+y) \\ cny & n(x+y) & x+y+z \end{pmatrix}.$$

where

$$x = \frac{a(c-1)}{(n\sigma^2_\beta + \sigma^2_\epsilon)^2},$$
$$y = \frac{a}{(cn\sigma^2_\alpha n\sigma^2_{beta} + \sigma^2_{epsilon})^2},$$
$$z = \frac{ac(n-1)}{\sigma^2_{epsilon}}.$$

For the balanced case Searle (1970) gives the determinant $|\mathbf{T}|$ as c^2n^4xyz and thus Jeffreys' prior is $= \sqrt{|\mathbf{T}|} \propto \frac{1}{(cn\sigma^2_\alpha + n\sigma^2_\beta + \sigma^2_\epsilon)(n\sigma^2_\beta + \sigma^2_\epsilon)\sigma^2_\epsilon}$

Now consider the two-way nested variance components model with measurements taken from P portions from each of K kegs from each of B batches. The three variance components are: the between batch variance component; the between keg variance component and the between portion variance. This is the design of the batch sampling study described in Section 4.1. Thus for the batch sampling study Jeffreys' prior is

$$\propto 1/(\sigma^2_p(\sigma^2_p + P\sigma^2_k)(\sigma^2_p + P\sigma^2_k + PK\sigma^2_b)). \tag{D.5}$$

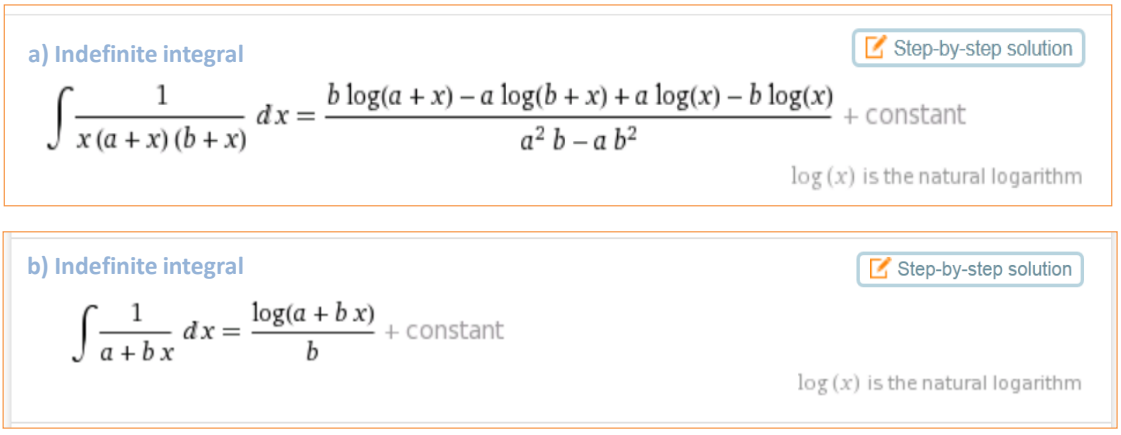
D.3.2 Density, Marginal and Conditional Distributions

For comparison with other priors it was desired to find the density function for the Jeffreys' prior and/or look at the marginal distributions for σ^2_{keg} and σ^2_{batch}. The density involves integrating Jeffreys' prior function over σ^2_{batch}, σ^2_{keg} and $\sigma^2_{portion}$ and the marginal distributions involve integration over $\sigma^2_{portion}$ and σ^2_{batch} or σ^2_{keg}.

Firstly I look at finding the marginal distribution for σ^2_{keg} and σ^2_{batch} by integrating over $\sigma^2_{portion}$. Figure D.2a) shows the solution to integrating over $\sigma^2_{portion}$ to obtain the marginal distribution for σ^2_{keg}, obtained from Wolfram|Alpha (2015), accessed 25$^{\text{th}}$ February 2015, where $x = \sigma^2_p$ and $a = P\sigma^2_k$.

However, we need to integrate from $x(= \sigma^2_p) = 0$ to ∞ and it is seen that the solution is undefined at $x = 0$ as follows. Taking the solution in Figure D.2a) and substituting for

FIGURE D.2: **Integration results from Wolfram for two-way model**



$x = \sigma_p^2$, $a = P\sigma_k^2$ and $b = P\sigma_k^2 + PK\sigma_b^2$ we have:

$$
\begin{aligned}
&\int_0^\infty \frac{1}{\sigma_p^2(\sigma_p^2 + P\sigma_k^2)(\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2)} \mathrm{d}\sigma_p^2 \\
&= \frac{1}{(P\sigma_k^2)(P\sigma_k^2 + PK\sigma_b^2)(P\sigma_k^2 - (P\sigma_k^2 + PK\sigma_b^2))} \times \\
&\quad \left[(P\sigma_k^2 + PK\sigma_b^2)\log(\sigma_p^2 + P\sigma_k^2) - (P\sigma_k^2)\log((\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2))\right. \\
&\quad \left. + (P\sigma_k^2 - (P\sigma_k^2 + PK\sigma_b^2))\log(\sigma_p^2)\right]_0^\infty \\
&= \frac{\left[(P\sigma_k^2 + PK\sigma_b^2)\log(\sigma_p^2 + P\sigma_k^2) - (P\sigma_k^2)\log((\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2)) - PK\sigma_b^2\log(\sigma_p^2)\right]_0^\infty}{(P\sigma_k^2)(P\sigma_k^2 + PK\sigma_b^2)(-PK\sigma_b^2)} \\
&= \frac{\left[\log\left(\frac{(\sigma_p^2 + P\sigma_k^2)^{(P\sigma_k^2 + PK\sigma_b^2)}}{(\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2)^{(P\sigma_k^2)}(\sigma_p^2)^{PK\sigma_b^2}}\right)\right]_0^\infty}{(P\sigma_k^2)(P\sigma_k^2 + PK\sigma_b^2)(-PK\sigma_b^2)} \\
&= \frac{\left[\log\left(\left(\frac{(\sigma_p^2 + P\sigma_k^2)}{(\sigma_p^2 + P\sigma_k^2 + PK\sigma_b^2)}\right)^{(P\sigma_k^2)}\left(\frac{(\sigma_p^2 + P\sigma_k^2)}{\sigma_p^2}\right)^{PK\sigma_b^2}\right)\right]_0^\infty}{(P\sigma_k^2)(P\sigma_k^2 + PK\sigma_b^2)(-PK\sigma_b^2)}.
\end{aligned}
\tag{D.6}
$$

When $\sigma_p^2 = \infty$, the numerator of Equation (D.6) is $\log(1) = 0$. However when $\sigma_p^2 = 0$ the numerator is $\log\left(\left(\frac{(0+P\sigma_k^2)}{(0+P\sigma_k^2+PK\sigma_b^2)}\right)^{(P\sigma_k^2)}\left(1 + \frac{P\sigma_k^2}{0}\right)^{PK\sigma_b^2}\right)$ which is undefined due to the zero denominator.

Thus instead of examining the marginal distribution, I try looking at the conditional Jeffreys' prior distribution of σ_{keg}^2 and σ_{batch}^2 given $\sigma_{portion}^2$. Using Figure D.2b) and letting x=σ_k^2, b=$\frac{\sigma_p^2}{P}$ and c=$\frac{(\sigma_p^2 + K\sigma_b^2)}{P}$ the conditional Jeffreys' prior is established as follows:

$$
\begin{aligned}
f(\sigma_k^2, \sigma_b^2|\sigma_p^2) &= \frac{\pi(\sigma_k^2, \sigma_b^2, \sigma_p^2)}{f(\sigma_p^2)} \\
&= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\int_0^\infty \int_0^\infty \frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)} \mathrm{d}\sigma_b^2 \mathrm{d}\sigma_k^2} \\
&= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\int_0^\infty \left[\frac{\log(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\right]_0^\infty \mathrm{d}\sigma_k^2}.
\end{aligned}
\tag{D.7}
$$

The integral within the denominator is undefined at $\sigma_b^2 = \infty$.

Thirdly I try examining the distribution of $\sigma^2_{keg}/\sigma^2_{portion}$ and $\sigma^2_{batch}/\sigma^2_{portion}$. Letting $y_1 = \sigma_p^2$, $y_2 = \frac{\sigma_k^2}{\sigma_p^2}$ and $y_3 = \frac{\sigma_b^2}{\sigma_p^2}$ and using Equations (C.2) and (D.5) then

$$
\begin{aligned}
h(\sigma_p^2, {}^{\sigma_k^2}/_{\sigma_p^2}, {}^{\sigma_b^2}/_{\sigma_p^2}) = h(y_1, y_2, y_3) &= g(\sigma_p^2, \sigma_k^2, \sigma_b^2) \left| \begin{matrix} \frac{\partial y_1}{\partial \sigma_p^2} & \frac{\partial y_1}{\partial \sigma_k^2} & \frac{\partial y_1}{\partial \sigma_b^2} \\ \frac{\partial y_2}{\partial \sigma_p^2} & \frac{\partial y_2}{\partial \sigma_k^2} & \frac{\partial y_2}{\partial \sigma_b^2} \\ \frac{\partial y_3}{\partial \sigma_p^2} & \frac{\partial y_3}{\partial \sigma_k^2} & \frac{\partial y_3}{\partial \sigma_b^2} \end{matrix} \right|^{-1} \\
&= g(\sigma_p^2, \sigma_k^2, \sigma_b^2) \left| \begin{matrix} 1 & 0 & 0 \\ \frac{-\sigma_k^2}{\sigma_p^4} & \frac{1}{\sigma_p^2} & 0 \\ \frac{-\sigma_b^2}{\sigma_p^4} & 0 & \frac{1}{\sigma_p^2} \end{matrix} \right|^{-1} \\
&= g(\sigma_p^2, \sigma_k^2, \sigma_b^2) \left({}^{1}/_{\sigma_p^4}\right)^{-1} \\
&= const_1 \times \sigma_p^4 \times \frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)} \\
&= const_1 \times \frac{\sigma_p^2}{(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}.
\end{aligned}
\tag{D.8}
$$

It can be shown that this is also undefined for the marginal distribution, $h\left(\frac{\sigma_k^2}{\sigma_p^2}, \frac{\sigma_b^2}{\sigma_p^2}\right)$. The distribution conditional on $\sigma^2_{portion}$, $h\left(\frac{\sigma_k^2}{\sigma_p^2}, \frac{\sigma_b^2}{\sigma_p^2}|\sigma_p^2\right)$, is also undefined since $f(\sigma_p^2)$ was seen to be undefined in Equation (D.7).

Given the above attempts failed to find a useful density function with which to visualise Jeffreys' prior the approach of restricting the function to a finite region used for the one-way variance components model will be used for the two-way model. This is derived in Section D.3.3.

D.3.3 Normalising the Prior Density Function for the Density within a Certain Region

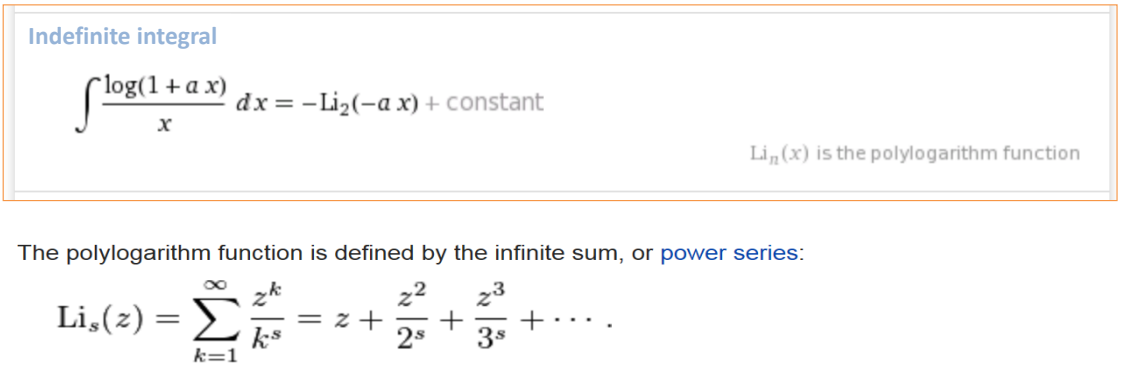
Whilst the probability associated with a range of σ_k^2 or σ_b^2 cannot be calculated since Jeffreys' prior is improper, it may be useful to try to normalise the prior density function for the density within a certain region.

From Equation (D.7) above

$$\begin{aligned} f(\sigma_k^2, \sigma_b^2|\sigma_p^2) &= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\int_0^{U_k}\left[\frac{\log(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\right]_0^{U_b} \mathrm{d}\sigma_k^2} \\ &= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\int_0^{U_k}\left(\frac{\log(\sigma_p^2+P\sigma_k^2+PKU_b)-\log(\sigma_p^2+P\sigma_k^2)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\right)\mathrm{d}\sigma_k^2} \\ &= \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\int_0^{U_k}\frac{\log\left(1+\frac{PKU_b}{\sigma_p^2+P\sigma_k^2}\right)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\mathrm{d}\sigma_k^2} \end{aligned} \tag{D.9}$$

Let $z = \frac{\sigma_p^2+P\sigma_k^2}{PKU_b}$, $a = \sigma_p^2P^2K^2U_b$, $L_z = \frac{\sigma_p^2}{PKU_b}$, $U_z = \frac{\sigma_p^2}{PKU_b} + \frac{U_k}{KU_b}$ and $Li_2(x) = \Sigma_{m=1}^{\infty}x^m/m^2$. Figure D.3 shows a result used in the solution to the integral in the denominator, obtained from Wolfram|Alpha (2015), accessed 23rd May 2015.

FIGURE D.3: **Further integration result from Wolfram for two-way model**



Then the denominator

$$\int_0^{U_k} \frac{\log\left(1+\frac{PKU_b}{\sigma_p^2+P\sigma_k^2}\right)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\mathrm{d}\sigma_k^2$$
$$= KU_b \int_{L_z}^{U_z} \frac{\log\left(1+\frac{1}{z}\right)}{az}\mathrm{d}z$$
$$= KU_b \left[\frac{Li_2(-\frac{1}{z})}{a}\right]_{L_z}^{U_z}.$$

and substituting back in,

$$\int_0^{U_k} \frac{\log\left(1+\frac{PKU_b}{\sigma_p^2+P\sigma_k^2}\right)}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)PK}\mathrm{d}\sigma_k^2$$
$$= \left[Li_2\left(\frac{-KU_b}{\left(\frac{\sigma_p^2}{P}+\sigma_b^2\right)}\right) PKU_b\sigma_p^2P^2K\right]_0^{U_k} \tag{D.10}$$
$$= \frac{\sum_{k=1}^{\infty}\frac{\left(\frac{-KU_b}{\left(\frac{\sigma_p^2}{P}+\sigma_b^2\right)}\right)^k}{k^2}}{\sigma_p^2P^2K}.$$

Thus from Equations (D.9) and (D.10),

$$f(\sigma_k^2,\sigma_b^2|\sigma_p^2) = \frac{\frac{1}{\sigma_p^2(\sigma_p^2+P\sigma_k^2)(\sigma_p^2+P\sigma_k^2+PK\sigma_b^2)}}{\frac{\sum_{k=1}^{\infty}\frac{\left(\frac{-KU_b}{\left(\frac{\sigma_p^2}{P}+\sigma_b^2\right)}\right)^k}{k^2}}{\sigma_p^2P^2K}}. \tag{D.11}$$

Appendix E

Posterior Sampling

Contents

This appendix contains more details regarding sampling from the posterior and the diagnostics used. Section E.1 has details of the Gelman-Rubin statistic used to check convergence in Section 6.3.2. Section E.2 contains an example dataset for the batch sampling design. Sections E.3 - E.6 relate to sampling using a proposal distribution based on the inverse gamma distribution performed in PROC MIXED in SAS software - see Section 6.4.

E.1 Gelman-Rubin Statistic

The documentation associated with the package CODA (Plummer et al., 2006) provides details of the calculations of the Gelman-Rubin statistic based on Gelman and Rubin (1992) and Brooks and Gelman (1998).

The Gelman-Rubin convergence statistic is based on a comparison of within-chain and between-chain variances and is given by

$$R = \sqrt{\frac{(d+3)\hat{V}}{(d+1)W}} = \sqrt{\frac{(d+3)}{(d+1)}\left(\frac{n-1}{n} + \frac{(m+1)B}{mnW}\right)}, \tag{E.1}$$

where

$$d = \frac{2 * \hat{V}^2}{Var(\hat{V})}, \quad \hat{V} = \frac{(n-1)W}{n} + \frac{(m+1)B}{mn},$$

W = the mean of the empirical variance within each chain, B/n = the empirical between-chain variance, m = number of chains, n = number of samples in each chain.

The upper 97.5% confidence limit for R is also calculated in CODA where $\frac{B}{W}$ in Equation (E.1) is approximated by an F distribution.

E.2 Data for Investigating Sampling algorithms in SAS

Table E.1 contains an example dataset for the batch sampling design which was used to investigate sampling algorithms in Section 6.4.1.3.

TABLE E.1: **Example batch sampling dataset used to investigate sampling algorithms**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
1	1	1	107.838	2	1	33	108.497
1	1	2	111.053	2	1	34	111.05
1	1	3	102.308	2	1	35	113.501
1	1	4	109.788	2	1	36	108.367
1	1	5	107.601	2	1	37	112.736
1	1	6	104.96	2	1	38	109.664
1	1	7	109.407	2	1	39	113.344
1	1	8	108.509	2	1	40	108.73
1	1	9	106.367	2	1	41	114.556
1	1	10	107.066	2	1	42	111.993
1	1	11	109.844	2	1	43	112.53
1	1	12	110.628	2	1	44	115.044
1	1	13	108.818	2	1	45	109.132
1	1	14	103.97	2	1	46	111.643
1	1	15	108.104	2	1	47	113.037
1	1	16	107.828	2	1	48	110.318
1	2	17	107.136	2	2	49	110.151
1	2	18	101.183	2	2	50	110.167
1	2	19	106.48	2	2	51	110.256
1	2	20	104.735	2	2	52	111.589
1	2	21	106.834	2	2	53	105.969
1	2	22	105.664	2	2	54	107.41
1	2	23	104.618	2	2	55	113.014
1	2	24	108.172	2	2	56	108.986
1	2	25	108.108	2	2	57	109.064
1	2	26	100.539	2	2	58	111.616
1	2	27	105.779	2	2	59	105.831
1	2	28	103.239	2	2	60	109.124
1	2	29	108.259	2	2	61	111.027
1	2	30	102.905	2	2	62	107.995
1	2	31	108.039	2	2	63	108.656
1	2	32	108.755	2	2	64	109.717

TABLE E.1: **Example batch sampling dataset used to investigate sampling algorithms**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
3	1	65	108.894	4	1	97	113.388
3	1	66	118.967	4	1	98	113.8
3	1	67	115.53	4	1	99	118.698
3	1	68	112.956	4	1	100	119.974
3	1	69	118.562	4	1	101	120.803
3	1	70	113.417	4	1	102	115.808
3	1	71	114.171	4	1	103	117.486
3	1	72	110.764	4	1	104	120.487
3	1	73	115.286	4	1	105	118.811
3	1	74	114.961	4	1	106	117.613
3	1	75	114.915	4	1	107	112.548
3	1	76	118.961	4	1	108	116.24
3	1	77	116.394	4	1	109	111.379
3	1	78	118.014	4	1	110	116.72
3	1	79	111.875	4	1	111	118.417
3	1	80	116.073	4	1	112	114.353
3	2	81	111.248	4	2	113	119.37
3	2	82	115.213	4	2	114	116.815
3	2	83	114.974	4	2	115	116.176
3	2	84	107.75	4	2	116	114.674
3	2	85	113.378	4	2	117	116.122
3	2	86	110.391	4	2	118	114.387
3	2	87	111.546	4	2	119	113.497
3	2	88	111.565	4	2	120	114.89
3	2	89	111.483	4	2	121	114.307
3	2	90	111.197	4	2	122	116.518
3	2	91	111.071	4	2	123	119.905
3	2	92	111.42	4	2	124	109.98
3	2	93	109.654	4	2	125	115.146
3	2	94	112.917	4	2	126	113.028
3	2	95	114.602	4	2	127	113.986
3	2	96	112.826	4	2	128	117.335

TABLE E.1: **Example batch sampling dataset used to investigate sampling algorithms**

Batch	Keg	Portion	Assay	Batch	Keg	Portion	Assay
5	1	129	118.458	6	1	161	122.045
5	1	130	118.319	6	1	162	122.866
5	1	131	122.62	6	1	163	122.645
5	1	132	122.162	6	1	164	122.308
5	1	133	116.85	6	1	165	126.938
5	1	134	122.296	6	1	166	123.352
5	1	135	120.239	6	1	167	124.563
5	1	136	125.802	6	1	168	120.837
5	1	137	120.363	6	1	169	122.116
5	1	138	119.419	6	1	170	128.292
5	1	139	115.507	6	1	171	123.128
5	1	140	121.534	6	1	172	120.459
5	1	141	115.06	6	1	173	125.001
5	1	142	118.528	6	1	174	124.213
5	1	143	120.743	6	1	175	122.619
5	1	144	121.205	6	1	176	124.553
5	2	145	117.901	6	2	177	124.419
5	2	146	119.052	6	2	178	124.65
5	2	147	115.2	6	2	179	117.316
5	2	148	121.39	6	2	180	119.523
5	2	149	119.104	6	2	181	118.012
5	2	150	115.757	6	2	182	122.341
5	2	151	118.274	6	2	183	119.98
5	2	152	116.647	6	2	184	121.028
5	2	153	119.862	6	2	185	114.744
5	2	154	116.53	6	2	186	119.721
5	2	155	117.819	6	2	187	121.167
5	2	156	118.385	6	2	188	119.675
5	2	157	114.407	6	2	189	119.606
5	2	158	120.159	6	2	190	118.671
5	2	159	118.632	6	2	191	123.247
5	2	160	115.464	6	2	192	123.531

E.3 Acceptance Rates in SAS - Additional Graphs

Sampling acceptance rates were investigated in Section 6.4.3.2. This section provides additional plots. Graphs showing the sampling acceptance rates for a subset of the priors previously shown in Figure 6.19 and for values of σ^2_{batch} of 0.5, 6 and 24 over the 10,000 datasets evaluated are provided. Sampling acceptance rates are shown for the various scenarios which varied the numbers of batches, numbers of kegs per batch and portions per keg and size of keg variance component. The sampling acceptance rates are summarised using boxplots. In most cases in Figures E.1 - E.3 only the lower whisker of the box plot varies with the majority of values having a sampling rate of 1 (The FLAT prior with 3 batches being an exception where the posterior is improper).

FIGURE E.1: **Sampling acceptance rates with varying number of kegs and portions (median, interquartile range and range with line joining medians)**

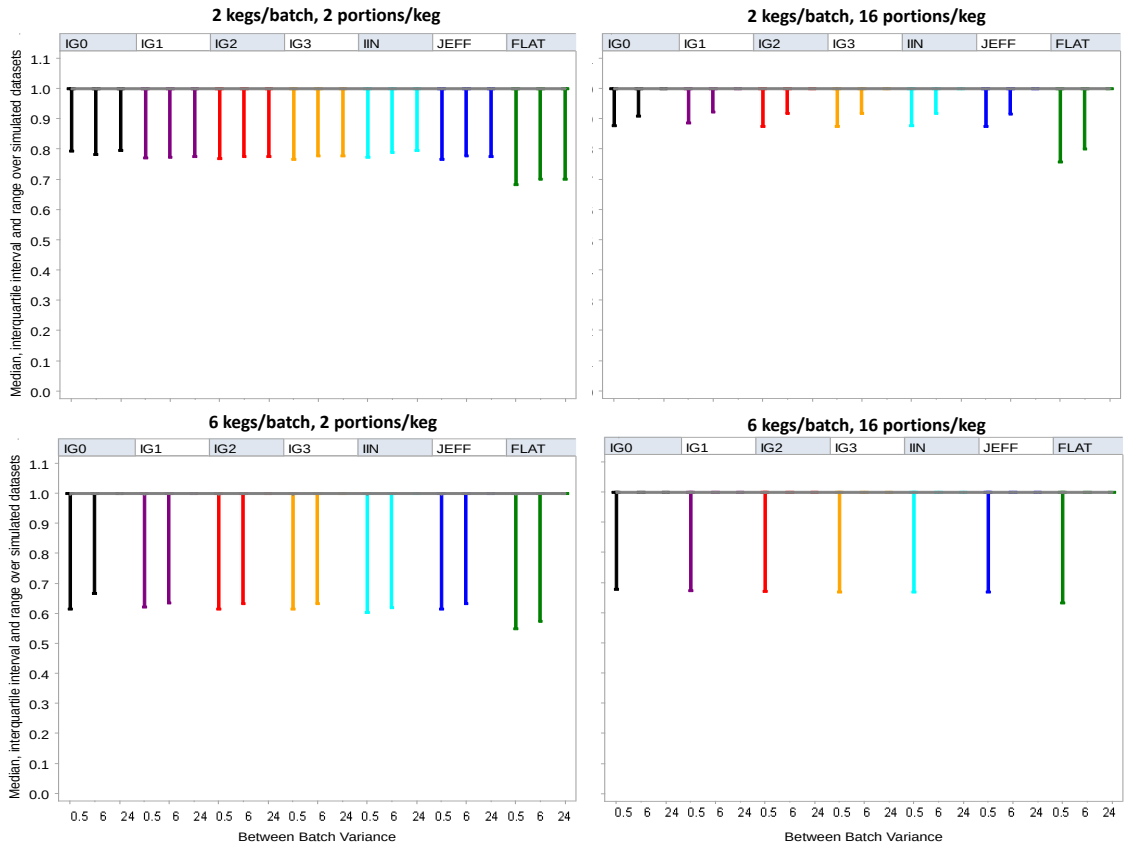


FIGURE E.2: **Sampling acceptance rates with varying number of batches (median, interquartile range and range with line joining medians)**

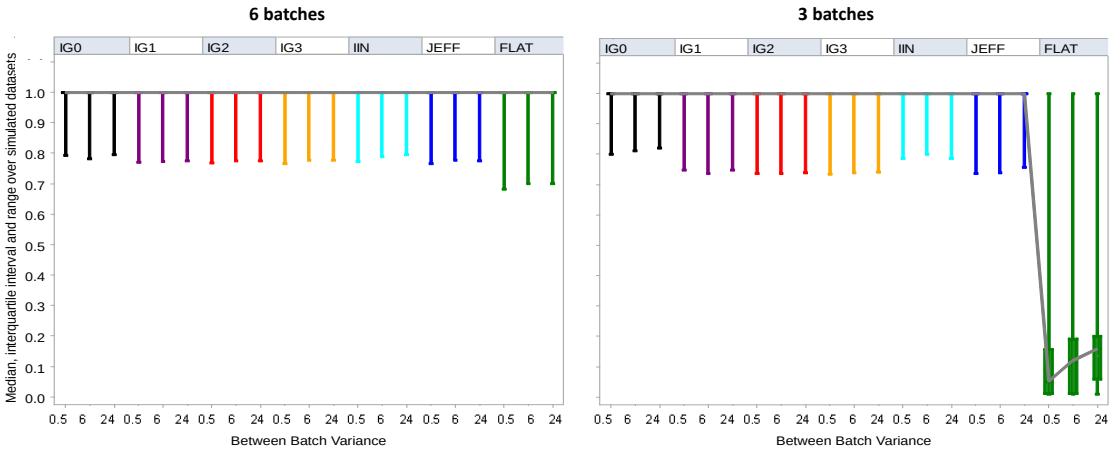
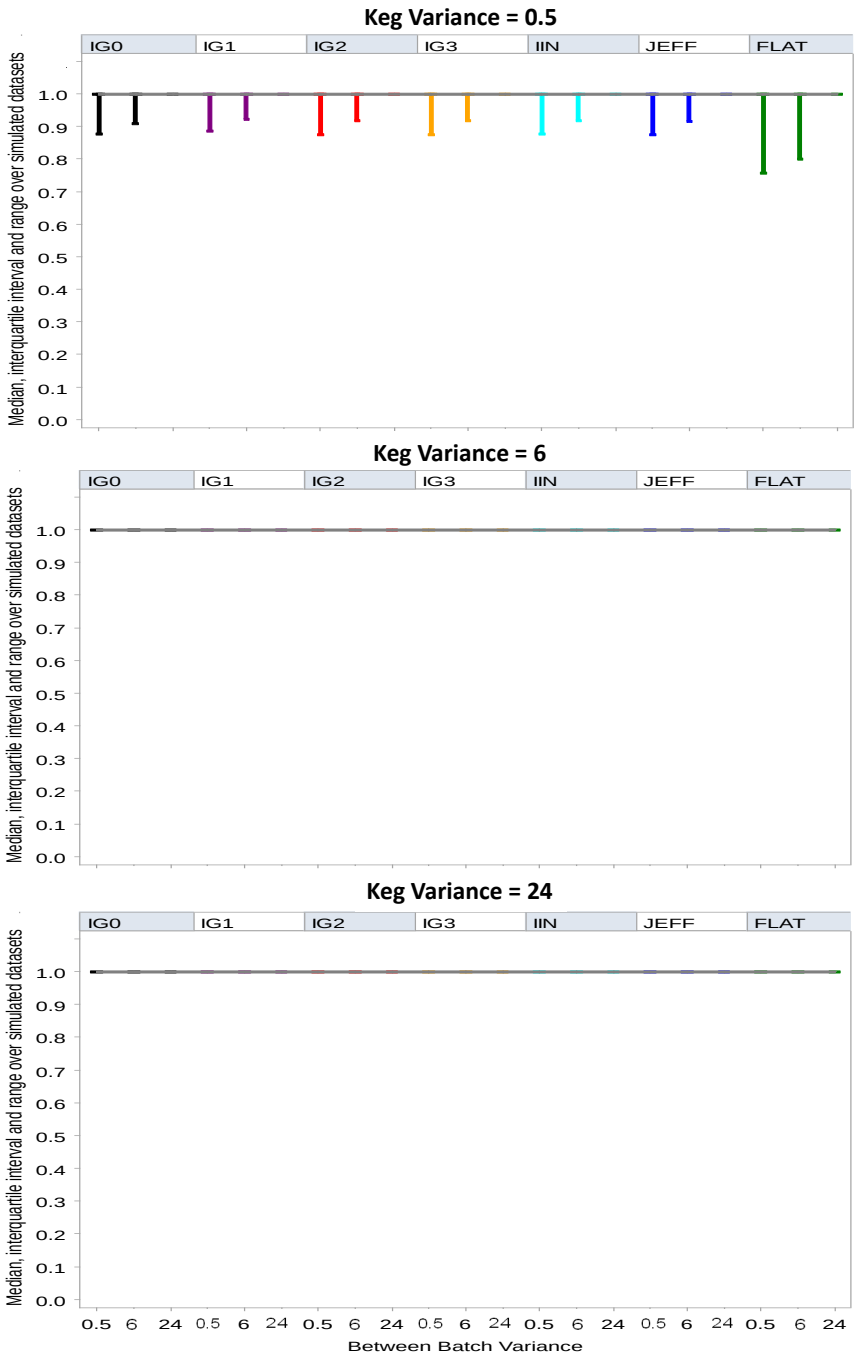


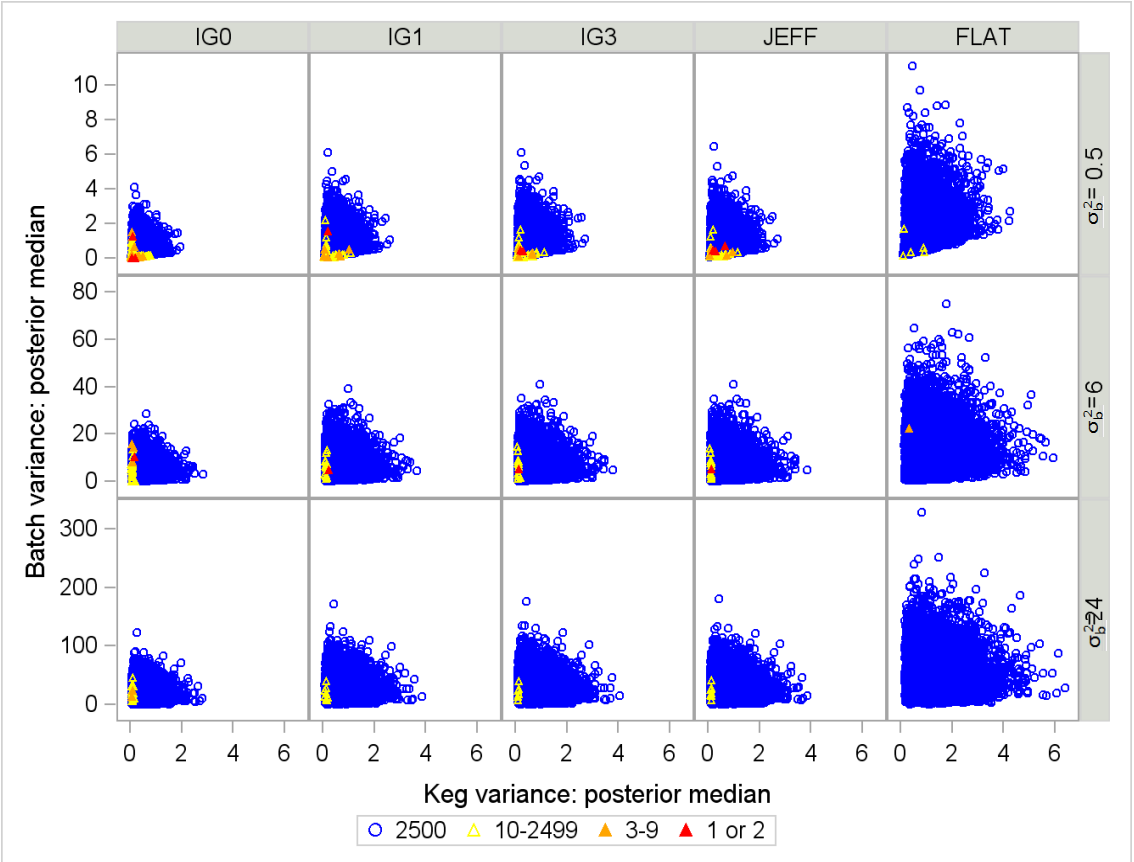
FIGURE E.3: **Sampling acceptance rates with varying keg variance (median, interquartile range and range with line joining medians)**



E.4 More on Association of Stopping with REML Variance Component and Posterior Median Estimates

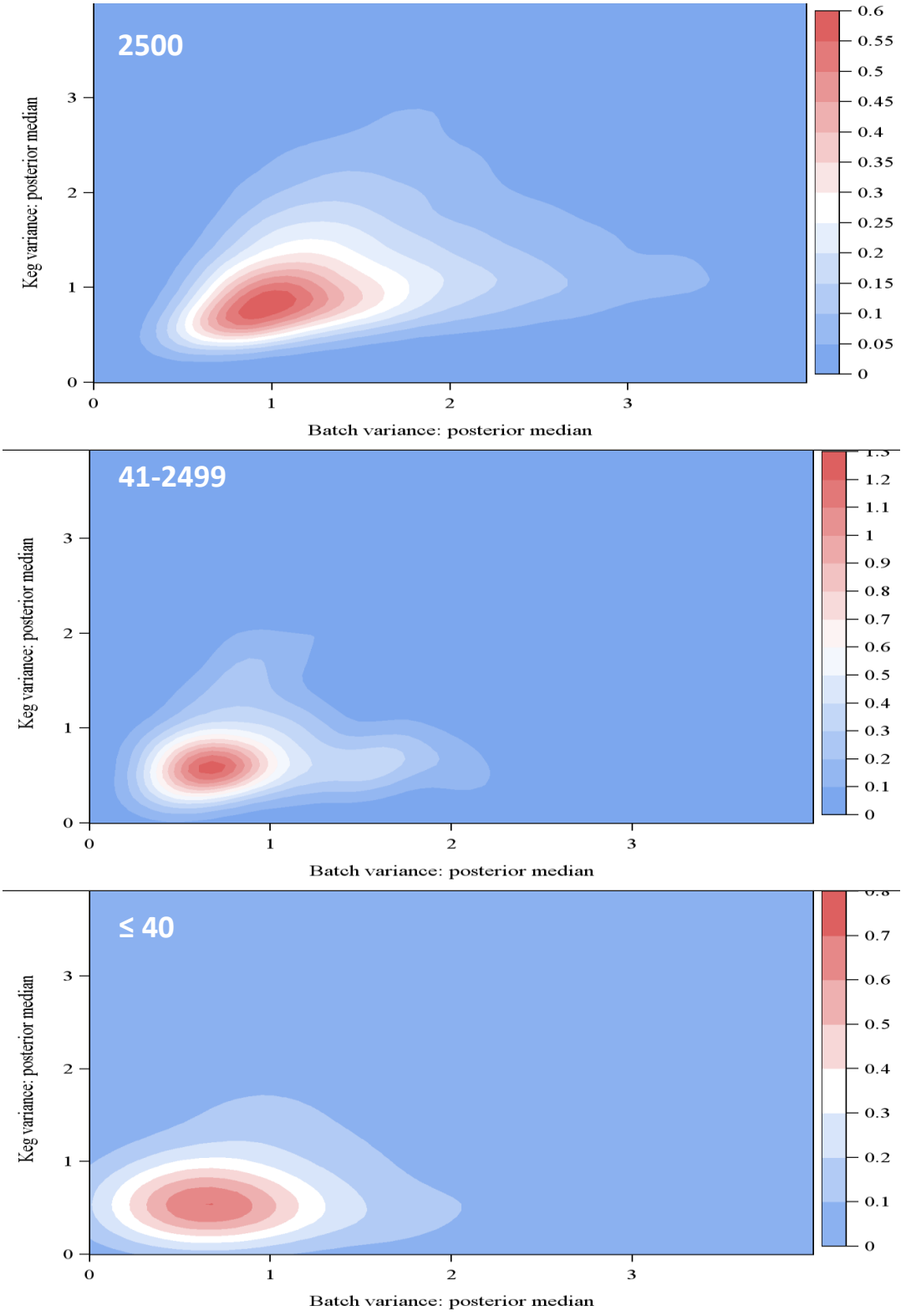
The association of SAS stopping sampling from the posterior with REML Variance Component Estimates and Posterior Median Estimates was investigated in Section 6.4.3.3. This section provides additional plots. Figure E.4 shows a scatterplot of posterior median estimates of batch variance component against keg variance component coloured by the number of posterior samples for the design with 2 kegs/batch and 16 portions/keg. This is provide for comparison with the smaller design with 2 kegs/batch and 2 portions/keg given in Figure 6.20.

FIGURE E.4: **Scatterplot of posterior median estimates of batch variance component against keg variance component coloured by the number of posterior samples - for scenario with 2 kegs/batch and 16 portions/keg**



It cannot be concluded from the scatterplot in Figure 6.21 that the distribution of the points differed according to whether sampling has stopped or not due to the high density of the plotted points. Thus a bivariate density plot of the data is given in Figure E.5 split by datasets where sampling stopped with $\leq$ 40 samples or 40-2499 samples or a full set of 2500 samples was obtained.

FIGURE E.5: **Bivariate density plots of posterior median estimates of batch variance component against keg variance component categorised by the number of posterior samples**



E.5 R program identifying IG distributions with 10 or 20% of marginal distributions above zero

The following code is used to identify inverse gamma distributions for the stratum variances which result in 10, 20 or 50% of the marginal distributions for the batch or keg variance above zero.

```
1  ###################################################################
2  #  Name          : basedist.r
3  #
4  #  Language/Ver : R-2.15.2 or R-3.0.2
5  #       Platform: x86_64-w64-mingw32/x64 (64-bit)
6  #  Operating Sys: X64_VSPRO platform
7  #
8  #  Purpose       : To find base inverse gamma distributions for linear
       combinations of variance components which will produce proposals where
       the variance component distributions have 10, 20 or 50% positive values
9  #
10 #  Input/Output :
11 #       Inputs:      designnos.csv      scenario/design details
12 #                    baseall.csv        default base distributions for
13 #                                  datasets not having full posterior sample
14 #
15 #       Outputs:     basedistn.csv     proposed base distribution, n=1,2,3
16 #                                                    for 10, 20 and 50%
17 #
18 #  Change History:
19 #  Vers.    Programmer    Date              Description
20 #  -----    ----------    ------        ---------------
21 #  1.0      MChatfield    4-MAY-2015     Original version
22 #  2.0      MChatfield    21-JUL-2017    Changed location of temp. files
23 #
24 ###################################################################
25 library(MCMCpack)
26 library(Bessel)
27 # Specify folder from which you will read and where you will write files
28 temp.dir <- "C:/folderlocation/temp/"
29 # Specify the design which determines the linear combinations used
30 designnos<-read.csv(paste(temp.dir,"designnos.csv",sep=""))
31 # Input default proposal distributions for datasets where sampling stopped
32 baseall<-read.csv(paste(temp.dir,"baseall.csv",sep=""))
33 head(baseall)
34 basedist1 <- baseall
35 basedist2 <- baseall
36 basedist3 <- baseall
37
38 # Example proposal distribution and calc of var comps
39 #1 ig 2.6 75 fA strata
40 #2 ig 3.1 50 fB strata
41 #3 ig 6.1 50 Z
```

```
#fB= (fB_strata - fZ_strata)/#fZs/fB;
#fA= (fA_strata - fB_strata)/#fZs/fB * #fAs/fB ;

#Getting the probability for linear combination of inverse gammas
i<-sqrt(as.complex(-1))
lamda.B1 <- as.numeric(1/designnos["no_fZs"])
lamda.B2 <- as.numeric(-1/designnos["no_fZs"])
lamda.A1 <- as.numeric(1/(designnos["no_fZs"]*designnos["no_fBs"]))
lamda.A2 <- as.numeric(-1/(designnos["no_fZs"]*designnos["no_fBs"]))
x<-0  #Value want specified prop. of distribution of var. comp. above (0)
probgt0_p1 <- 0.1 #Specifying how much of the distribution is above x=0
probgt0_p2 <- 0.2 #Specifying how much of the distribution is above x=0
probgt0_p3 <- 0.5 #Specifying how much of the distribution is above x=0
# characteristic function
characteristic.fun <- function(t,gscale,gshape) {2*((-i*t*gscale)^(gshape/
    2))*as.complex(BesselK(2*((-i*t*gscale)^0.5),gshape))/gamma(gshape)}
# integrands for var comps A and B
integrandA.fun <- function(t,gscaleAf) {Im(exp(-i*t*x)*characteristic.fun(
    lamda.A1*t,gscaleAf,gshapeA)*characteristic.fun(lamda.A2*t,gscaleB,
    gshapeB)/t)}
integrandB.fun <- function(t,gscaleBf) {Im(exp(-i*t*x)*characteristic.fun(
    lamda.B1*t,gscaleBf,gshapeB)*characteristic.fun(lamda.B2*t,gscaleZ,
    gshapeZ)/t)}

# Integrating to get the proportions
probBfn <- function(gscaleB) {
integralB <- integrate(integrandB.fun, lower=0,upper=Inf,gscaleBf=gscaleB)
probB <- 1-(0.5 - integralB$value/pi)
probB
}
probAfn <- function(gscaleA) {
integralA <- integrate(integrandA.fun, lower=0,upper=Inf,gscaleAf=gscaleA)
probA <- 1-(0.5 - integralA$value/pi)
probA
}
# Functions to calc. diff. between proportion and the desired proportion
probBminfn_p1 <- function(gscaleB) {
integralB <- integrate(integrandB.fun, lower=0,upper=Inf,gscaleBf=gscaleB)
probB <- 1-(0.5 - integralB$value/pi)
probBmin <- (probB-probgt0_p1)^2
probBmin
}
probBminfn_p2 <- function(gscaleB) {
integralB <- integrate(integrandB.fun, lower=0,upper=Inf,gscaleBf=gscaleB)
probB <- 1-(0.5 - integralB$value/pi)
probBmin <- (probB-probgt0_p2)^2
probBmin
}
probBminfn_p3 <- function(gscaleB) {
integralB <- integrate(integrandB.fun, lower=0,upper=Inf,gscaleBf=gscaleB)
probB <- 1-(0.5 - integralB$value/pi)
```

```
probBmin <- (probB-probgt0_p3)^2
probBmin
}
probAminfn_p1 <- function(gscaleA) {
integralA <- integrate(integrandA.fun, lower=0,upper=Inf,gscaleAf=gscaleA)
probA <- 1-(0.5 - integralA$value/pi)
probAmin <- (probA-probgt0_p1)^2
probAmin
}
probAminfn_p2 <- function(gscaleA) {
integralA <- integrate(integrandA.fun, lower=0,upper=Inf,gscaleAf=gscaleA)
probA <- 1-(0.5 - integralA$value/pi)
probAmin <- (probA-probgt0_p2)^2
probAmin
}
probAminfn_p3 <- function(gscaleA) {
integralA <- integrate(integrandA.fun, lower=0,upper=Inf,gscaleAf=gscaleA)
probA <- 1-(0.5 - integralA$value/pi)
probAmin <- (probA-probgt0_p3)^2
probAmin
}

totiter <- dim(baseall)[1]/3     # No. of datasets needing proposal distr.

for (iteri in 1:totiter) {
iter<-baseall[baseall$itern==iteri, ][1,5]
baseiteri <- baseall[baseall$itern==iteri, ]
gshapeA <- baseiteri[baseiteri$Density==1,"Parm1"]
gscaleA <- baseiteri[baseiteri$Density==1,"Parm2"]
gshapeB <- baseiteri[baseiteri$Density==2,"Parm1"]
gscaleB <- baseiteri[baseiteri$Density==2,"Parm2"]
gshapeZ <- baseiteri[baseiteri$Density==3,"Parm1"]
gscaleZ <- baseiteri[baseiteri$Density==3,"Parm2"]
#large shape parameters cause problems e.g. gamma(gshape)=10x10^155 if
    gshape=100 so try making smaller but keeping same mean
if (gshapeA>50){
gshapeA.old<-gshapeA
gshapeA<-50
gscaleA<-gscaleA*50/gshapeA.old
baseiteri[baseiteri$Density==1,"Parm1"] <- gshapeA
baseiteri[baseiteri$Density==1,"Parm2"] <- gscaleA
basedist1[basedist1$itern==iteri, ][basedist1[basedist1$itern==iteri, ]$
    Density==1,"Parm1"] <- gshapeA
basedist1[basedist1$itern==iteri, ][basedist1[basedist1$itern==iteri, ]$
    Density==1,"Parm2"] <- gscaleA
basedist2[basedist2$itern==iteri, ][basedist2[basedist2$itern==iteri, ]$
    Density==1,"Parm1"] <- gshapeA
basedist2[basedist2$itern==iteri, ][basedist2[basedist2$itern==iteri, ]$
    Density==1,"Parm2"] <- gscaleA
basedist3[basedist3$itern==iteri, ][basedist3[basedist3$itern==iteri, ]$
    Density==1,"Parm1"] <- gshapeA
```

```
133 basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==iteri,]$
        Density==1,"Parm2"] <- gscaleA
134 }
135 if (gshapeB>50){
136 gshapeB.old<-gshapeB
137 gshapeB<-50
138 gscaleB<-gscaleB*50/gshapeB.old
139 baseiteri[baseiteri$Density==2,"Parm1"] <- gshapeB
140 baseiteri[baseiteri$Density==2,"Parm2"] <- gscaleB
141 basedist1[basedist1$itern==iteri,][basedist1[basedist1$itern==iteri,]$
        Density==2,"Parm1"] <- gshapeB
142 basedist1[basedist1$itern==iteri,][basedist1[basedist1$itern==iteri,]$
        Density==2,"Parm2"] <- gscaleB
143 basedist2[basedist2$itern==iteri,][basedist2[basedist2$itern==iteri,]$
        Density==2,"Parm1"] <- gshapeB
144 basedist2[basedist2$itern==iteri,][basedist2[basedist2$itern==iteri,]$
        Density==2,"Parm2"] <- gscaleB
145 basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==iteri,]$
        Density==1,"Parm1"] <- gshapeB
146 basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==iteri,]$
        Density==1,"Parm2"] <- gscaleB
147 }
148 if (gshapeZ>50){
149 gshapeZ.old<-gshapeZ
150 gshapeZ<-50
151 gscaleZ<-gscaleZ*50/gshapeZ.old
152 baseiteri[baseiteri$Density==3,"Parm1"] <- gshapeZ
153 baseiteri[baseiteri$Density==3,"Parm2"] <- gscaleZ
154 basedist1[basedist1$itern==iteri,][basedist1[basedist1$itern==iteri,]$
        Density==3,"Parm1"] <- gshapeZ
155 basedist1[basedist1$itern==iteri,][basedist1[basedist1$itern==iteri,]$
        Density==3,"Parm2"] <- gscaleZ
156 basedist2[basedist2$itern==iteri,][basedist2[basedist2$itern==iteri,]$
        Density==3,"Parm1"] <- gshapeZ
157 basedist2[basedist2$itern==iteri,][basedist2[basedist2$itern==iteri,]$
        Density==3,"Parm2"] <- gscaleZ
158 basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==iteri,]$
        Density==1,"Parm1"] <- gshapeZ
159 basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==iteri,]$
        Density==1,"Parm2"] <- gscaleZ
160 }
161
162 # Now optimise to find proposal distribution with desired proportions
163 # Note this gives warning messages about Nelder-Mead being unreliable but
        it works and the code checks the answer
164 # Changing scaleB where prop. of distrib. above zero is < p1 (usually 0.1)
165 if (tryCatch(probBfn(gscaleB)< probgt0_p1,
166 error=function(cond) {
167         message(cond)
168         # Choose a return value in case of error
169         return(F)
```

```
170 }))
171 {
172         if (gscaleB<2) gscaleB<-2
173         resultB <- tryCatch(
174         {optim(par=c(gscaleB),fn=probBminfn_p1)}
175         , error=function(cond) {
176         message(cond)
177         # Choose a return value in case of error
178         return(list(par=gscaleB,err=cond))
179 })
180
181 problem.probBfn <- tryCatch(if (probBfn(resultB$par)>-1) {}
182 , error=function(cond) {
183         message(cond)
184         # Choose a return value in case of error
185         return(list(par=T,err=cond))
186 })
187
188 if (is.null(problem.probBfn)) {
189 if (abs(probBfn(resultB$par)-probgt0_p1)<0.01) {
190         gscaleB <- resultB$par
191         basedist1[basedist1$itern==iteri,][basedist1[basedist1$itern==
    iteri,]$Density==2,"Parm2"] <- gscaleB
192         } else {
193         if (exists("err",where=resultB)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultB$err,sep="")))} else {errmess<-""}
194         B_p1_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=resultB$par,probgt0=probgt0_p1,actualprob=probBfn(resultB$par),
    errmess)
195         } } else if (problem.probBfn$par == T) {
196         errmess<-gsub("\n","",(paste("Error message: ",problem.probBfn$err,
    sep="")))
197         B_p1_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=NA,probgt0=probgt0_p2,actualprob=NA,errmess)
198 }}
199
200 # Changing scaleA where prop. of distrib. above zero is < p1 (usually 0.1)
201 if (tryCatch(probAfn(gscaleA)< probgt0_p1,
202 error=function(cond) {
203         message(cond)
204         # Choose a return value in case of error
205         return(F)
206 }))
207 {
208         if (gscaleA<2) gscaleA<-2
209         resultA <- tryCatch(
210         {optim(par=c(gscaleA),fn=probAminfn_p1)}
211         , error=function(cond) {
212         message(cond)
213         # Choose a return value in case of error
214         return(list(par=gscaleA,err=cond))
```

```
})

problem.probAfn <- tryCatch(if (probAfn(resultA$par)>-1) {}
,  error=function(cond) {
        message(cond)
        # Choose a return value in case of error
        return(list(par=T, err=cond))
})

if (is.null(problem.probAfn)) {
if (abs(probAfn(resultA$par)-probgt0_p1)<0.01) {
        gscaleA <- resultA$par
        basedist1[basedist1$itern==iteri , ][basedist1[basedist1$itern==
    iteri , ]$Density==1,"Parm2"] <- gscaleA
        } else {
        if (exists("err",where=resultA)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultA$err,sep="")))} else {errmess<-""}
        A_p1_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=resultA$par,probgt0=probgt0_p1,actualprob=probAfn(resultA$par),
    errmess)
        }
        } else if (problem.probAfn$par == T) {
        errmess<-gsub("\n","",(paste("Error message: ",problem.probAfn$err,
    sep="")))
        A_p1_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=NA,probgt0=probgt0_p1,actualprob=NA,errmess)
}}

# Changing scaleB where prop. of distrib. above zero is < p2 (usually 0.2)
if (tryCatch(probBfn(gscaleB)< probgt0_p2,
error=function(cond) {
        message(cond)
        # Choose a return value in case of error
        return(F)
}))
{
        if (gscaleB<2) gscaleB<-2
        resultB <- tryCatch(
        {optim(par=c(gscaleB),fn=probBminfn_p2)}
        ,  error=function(cond) {
        message(cond)
        # Choose a return value in case of error
        return(list(par=gscaleB,err=cond))
})

problem.probBfn <- tryCatch(if (probBfn(resultB$par)>-1) {}
,  error=function(cond) {
        message(cond)
        # Choose a return value in case of error
        return(list(par=T,err=cond))
})
```

```
260
261 if (is.null(problem.probBfn)) {
262 if (abs(probBfn(resultB$par)-probgt0_p2)<0.01) {
263 gscaleB <- resultB$par
264 basedist2[basedist2$itern==iteri, ][basedist2[basedist2$itern==iteri, ]$
    Density==2,"Parm2"] <- gscaleB
265 } else {
266         if (exists("err",where=resultB)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultB$err,sep="")))} else {errmess<-""}
267         B_p2_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=resultB$par,probgt0=probgt0_p2,actualprob=probBfn(resultB$par),
    errmess)
268         }
269         } else if (problem.probBfn$par == T) {
270         errmess<-gsub("\n","",(paste("Error message: ",problem.probBfn$err,
    sep="")))
271         B_p2_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=NA,probgt0=probgt0_p2,actualprob=NA,errmess)
272 }}
273
274 #Changing scaleA where prop. of distrib. above zero is < p2 (usually 0.2)
275 if (tryCatch(probAfn(gscaleA)< probgt0_p2,
276 error=function(cond) {
277         message(cond)
278         # Choose a return value in case of error
279         return(F)
280 }))
281 {
282 if (gscaleA<2) gscaleA<-2
283 resultA <- tryCatch(
284 {optim(par=c(gscaleA),fn=probAminfn_p2)}
285 ,  error=function(cond) {
286 message(cond)
287 # Choose a return value in case of error
288 return(list(par=gscaleA,err=cond))
289 })
290
291 problem.probAfn <- tryCatch(if (probAfn(resultA$par)>-1) {}
292 ,  error=function(cond) {
293         message(cond)
294         # Choose a return value in case of error
295         return(list(par=T,err=cond))
296 })
297
298 if (is.null(problem.probAfn)) {
299 if (abs(probAfn(resultA$par)-probgt0_p2)<0.01) {
300         gscaleA <- resultA$par
301         basedist2[basedist2$itern==iteri, ][basedist2[basedist2$itern==
    iteri, ]$Density==1,"Parm2"] <- gscaleA
302         } else {
```

```
303         if (exists("err",where=resultA)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultA$err,sep="")))} else {errmess<-""}
304         A_p2_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=resultA$par,probgt0=probgt0_p2,actualprob=probAfn(resultA$par),
    errmess)
305         }
306         } else if (problem.probAfn$par == T) {
307         errmess<-gsub("\n","",(paste("Error message: ",problem.probAfn$err,
    sep="")))
308         A_p2_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=NA,probgt0=probgt0_p2,actualprob=NA,errmess)
309 }}
310
311 # Changing scaleB where prop. of distrib. above zero is < p3 (usually 0.5)
312 if (tryCatch(probBfn(gscaleB)< probgt0_p3,
313 error=function(cond) {
314         message(cond)
315         # Choose a return value in case of error
316         return(F)
317 }))
318 {
319 if (gscaleB<2) gscaleB<-2
320 resultB <- tryCatch(
321 {optim(par=c(gscaleB),fn=probBminfn_p3)}
322 ,  error=function(cond) {
323         message(cond)
324         # Choose a return value in case of error
325         return(list(par=gscaleB,err=cond))
326 }
327 )
328
329 problem.probBfn <- tryCatch(if (probBfn(resultB$par)>-1) {}
330 ,  error=function(cond) {
331         message(cond)
332         # Choose a return value in case of error
333         return(list(par=T,err=cond))
334 })
335
336 if (is.null(problem.probBfn)) {
337 if (abs(probBfn(resultB$par)-probgt0_p3)<0.01) {
338 gscaleB <- resultB$par
339 basedist3[basedist3$itern==iteri, ][basedist3[basedist3$itern==iteri, ]$
    Density==2,"Parm2"] <- gscaleB
340 } else {
341         if (exists("err",where=resultB)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultB$err,sep="")))} else {errmess<-""}
342         B_p3_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=resultB$par,probgt0=probgt0_p3,actualprob=probBfn(resultB$par),
    errmess)
343         }
344         } else if (problem.probBfn$par == T) {
```

```
345         errmess<-gsub("\n","",(paste("Error message: ",problem.probBfn$err,
    sep="")))
346         B_p3_optfail <- data.frame(iteri,iter,gscale=gscaleB,param="fB",
    result=NA,probgt0=probgt0_p3,actualprob=NA,errmess)
347 }}
348
349 # Changing scaleA  where prop. of distrib. above zero is < p3 (usually 0.5)
350 if (tryCatch(probAfn(gscaleA)< probgt0_p3,
351 error=function(cond) {
352         message(cond)
353         # Choose a return value in case of error
354         return(F)
355 }))
356 {
357 if (gscaleA<2) gscaleA<-2
358 resultA <- tryCatch(
359 {optim(par=c(gscaleA),fn=probAminfn_p3)}
360 ,  error=function(cond) {
361         message(cond)
362         # Choose a return value in case of error
363         return(list(par=gscaleA,err=cond))
364 })
365
366 problem.probAfn <- tryCatch(if (probAfn(resultA$par)>-1) {}
367 ,  error=function(cond) {
368         message(cond)
369         # Choose a return value in case of error
370         return(list(par=T,err=cond))
371 })
372
373 if (is.null(problem.probAfn)) {
374 if (abs(probAfn(resultA$par)-probgt0_p3)<0.01) {
375         gscaleA <- resultA$par
376         basedist3[basedist3$itern==iteri,][basedist3[basedist3$itern==
    iteri,]$Density==1,"Parm2"] <- gscaleA
377         } else {
378         if (exists("err",where=resultA)) {errmess<-gsub("\n","",(paste("
    Error message: ",resultA$err,sep="")))} else {errmess<-""}
379         A_p3_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=resultA$par,probgt0=probgt0_p3,actualprob=probAfn(resultA$par),
    errmess)
380         }
381         } else if (problem.probAfn$par == T) {
382         errmess<-gsub("\n","",(paste("Error message: ",problem.probAfn$err,
    sep="")))
383         A_p3_optfail <- data.frame(iteri,iter,gscale=gscaleA,param="fA",
    result=NA,probgt0=probgt0_p3,actualprob=NA,errmess)
384 }}
385
386 # Checks to see if optimisation fails
```

```
387 checkoptfail<-c(exists("A_p1_optfail"),exists("A_p2_optfail"),exists("A_p3_
        optfail"),exists("B_p1_optfail"),exists("B_p2_optfail"),exists("B_p3_
        optfail"))
388 if (any(checkoptfail)) {
389 if (!exists("optfail")) optfail<-NULL
390         if (checkoptfail[1]) optfail<-rbind(optfail,A_p1_optfail)
391         if (checkoptfail[2]) optfail<-rbind(optfail,A_p2_optfail)
392         if (checkoptfail[3]) optfail<-rbind(optfail,A_p3_optfail)
393         if (checkoptfail[4]) optfail<-rbind(optfail,B_p1_optfail)
394         if (checkoptfail[5]) optfail<-rbind(optfail,B_p2_optfail)
395         if (checkoptfail[6]) optfail<-rbind(optfail,B_p3_optfail)
396 }}
397 #Write out the new proposal distributions for probgt0_p1, probgt0_p2 or
        probgt0_p3 proportion of variance component proposals above zero
398 if (exists("optfail")) write.csv(optfail, paste(temp.dir,"optfail.csv",sep=
        ""),row.names=F)
399 write.csv(basedist1, paste(temp.dir,"basedist1.csv",sep=""),row.names=F)
400 write.csv(basedist2, paste(temp.dir,"basedist2.csv",sep=""),row.names=F)
401 write.csv(basedist3, paste(temp.dir,"basedist3.csv",sep=""),row.names=F)
```

E.6 Autocorrelation of Batch Variance Components for Example Datasets in SAS

In Section 6.4.5 the autocorrelation of posterior samples was examined for the keg variance for the default proposal distribution and for modified proposal distributions. This section contains the corresponding graphs for the batch variance.

The autocorrelation of the posterior samples for the batch variance for example datasets where the default proposal distribution achieved a full sample from the posterior is shown in Figure E.6. The autocorrelations for the batch variance samples for example datasets where the proposal distributions were modified to aim for 10, 20 or 50% of the marginal keg or batch variance proposal distributions above zero (denoted P10, P20 and P50 respectively), and where the scale parameters for the IG for the batch and keg variance strata are changed to 50 (denoted S50) are shown in Figures E.7 - E.10.

FIGURE E.6: **Autocorrelations for batch variance for example datasets with default proposal distributions**

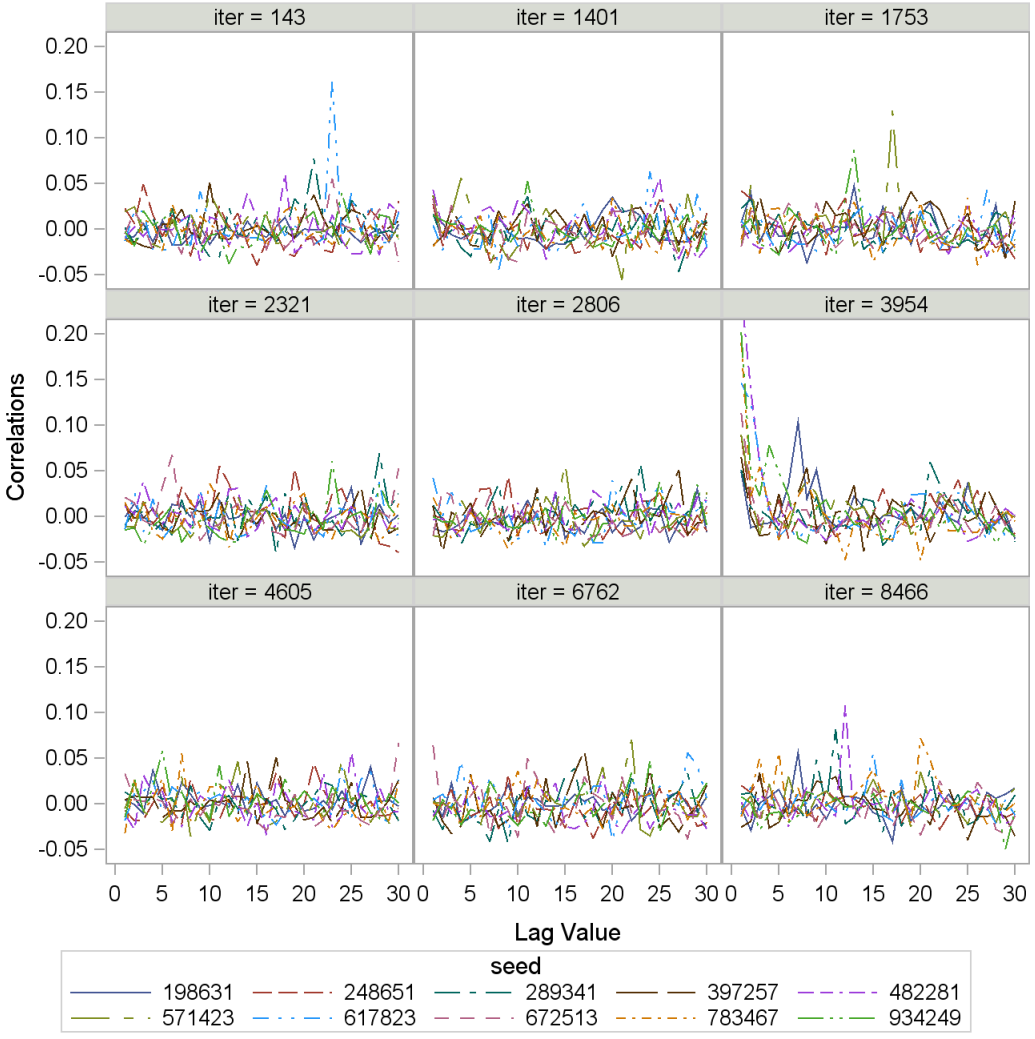


FIGURE E.7: **Autocorrelations for batch variance for example datasets with P10 proposal distributions**

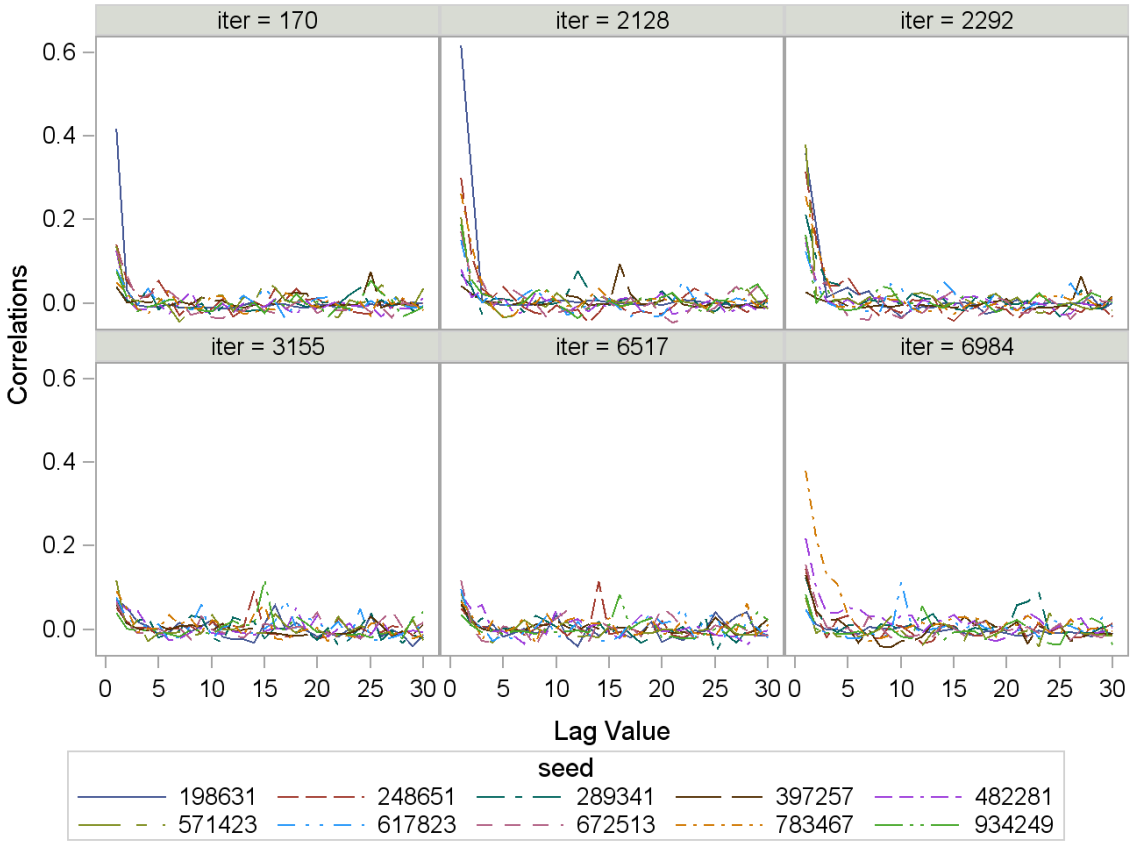


FIGURE E.8: **Autocorrelations for batch variance for example datasets with P20 proposal distributions**

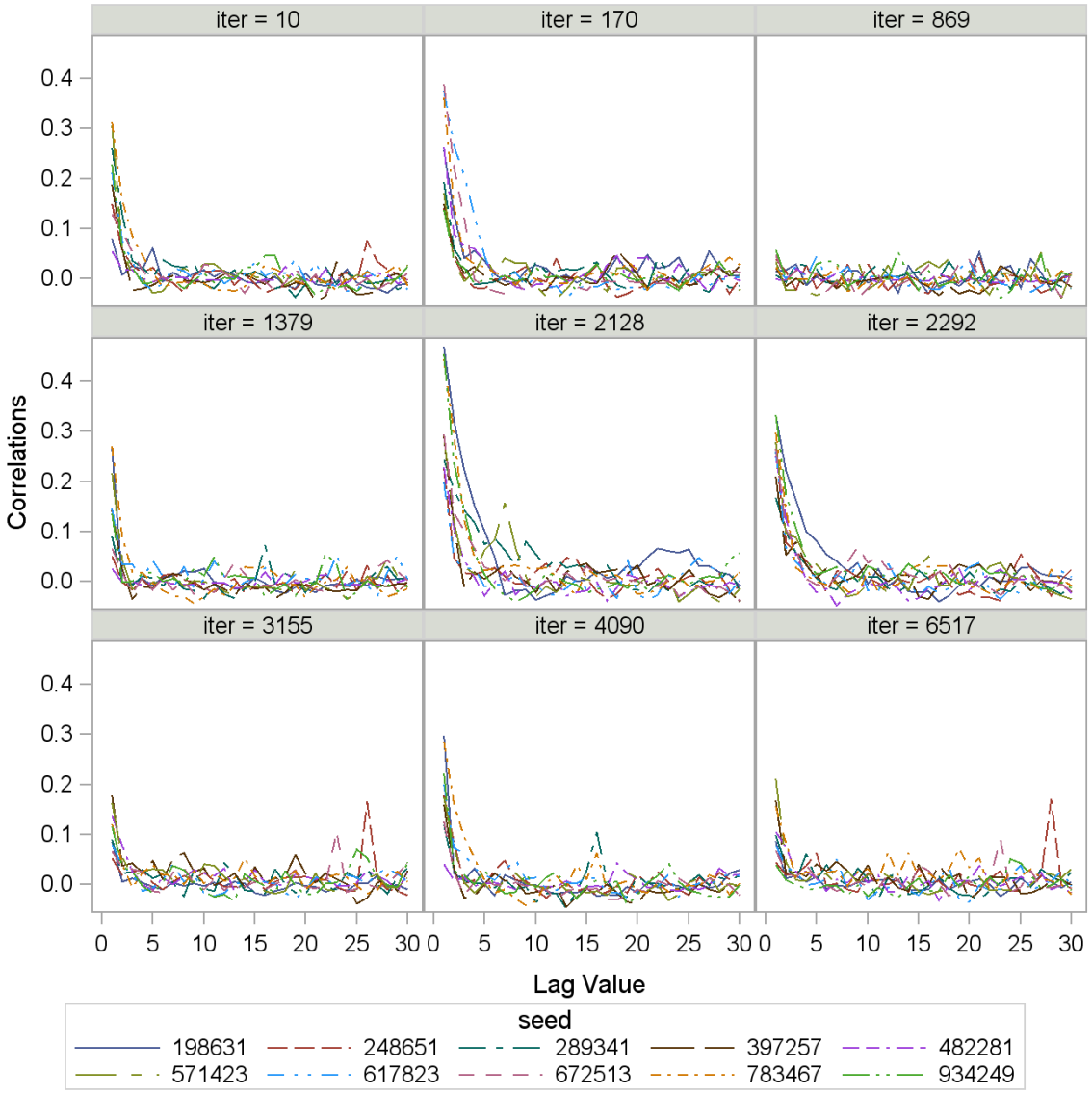


FIGURE E.9: **Autocorrelations for batch variance for example datasets with P50 proposal distributions**

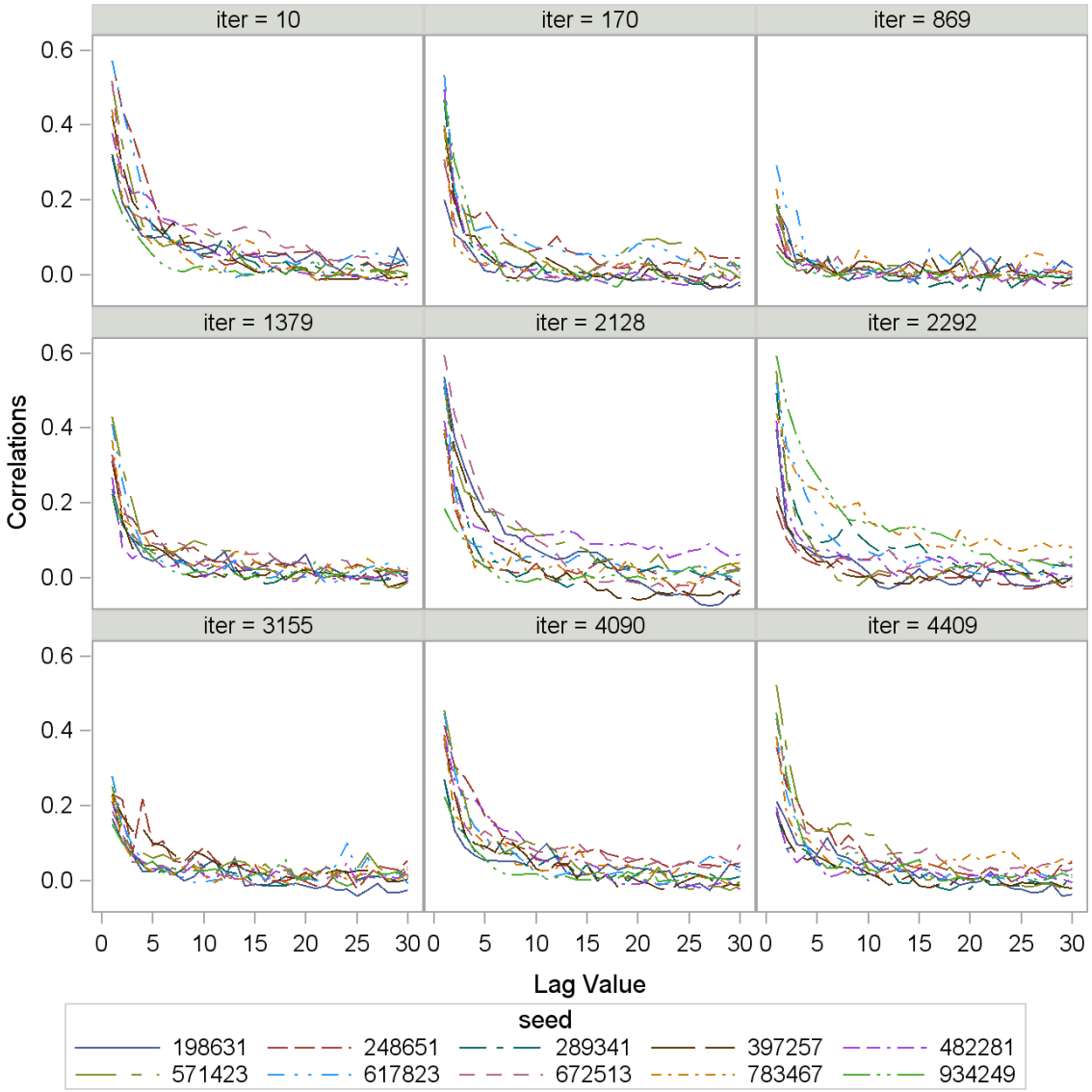
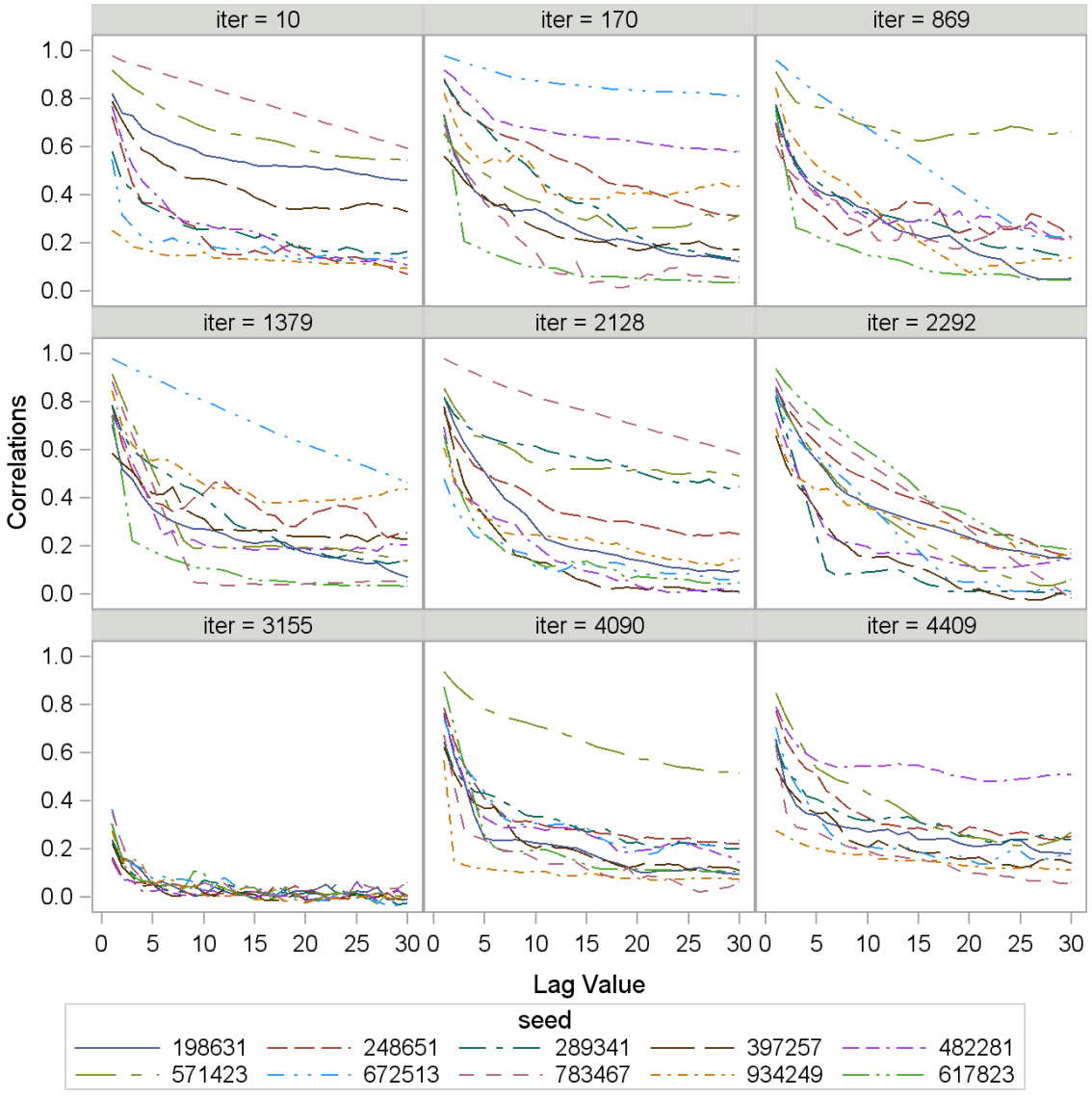


FIGURE E.10: **Autocorrelations for batch variance for example datasets with S50 proposal distributions**



Appendix F

Coverage Results

The coverage of the credible intervals and the proportions of datasets with the true value outside the lower or upper CrI limit, are summarised in Table F.1 for the inverse gamma priors investigated in Section 7.2.1 (excepting those investigated more fully in Section 7.2.2). The results are provided for varying credible interval types for total, batch, keg and portion variances for scenario k2p16_0.5 and $\sigma_b^2 = 0.5$, 6 and 24.

For all scenarios and priors and credible interval types investigated in Section 7.2.2 the coverage of the credible intervals, and the proportions of datasets with the true value outside the lower or upper CrI limit, are summarised in Tables F.2 to F.5. The results are provided for varying credible interval types for total, batch, keg and portion variances for $\sigma_b^2 = 0.5$, 6 and 24.

TABLE F.1: **Coverage and proportion outside limits for total variance and variance components over σ_b^2=0.5, 6, 24 for inverse gamma priors**

			Total Variance			Batch Variance			Keg Variance			Portion Variance		
			Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside	
Prior	σ_b^2	Int.		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
I1E3	0.5	hpds	0.917	0.083	0.000	0.996	0.004	0.000	0.978	0.022	0.000	0.953	0.026	0.021
I1E3	0.5	hpdv	0.933	0.067	0.000	1.000	0.000	0.000	0.994	0.006	0.000	0.955	0.021	0.024
I1E3	0.5	pct	0.795	0.206	0.000	0.986	0.014	0.000	0.960	0.040	0.000	0.953	0.030	0.017
I1E6	0.5	hpds	0.640	0.360	0.000	0.995	0.005	0.000	0.850	0.150	0.000	0.953	0.025	0.022
I1E6	0.5	hpdv	0.690	0.310	0.000	1.000	0.000	0.000	0.952	0.048	0.000	0.955	0.020	0.024
I1E6	0.5	pct	0.337	0.663	0.000	0.975	0.025	0.000	0.735	0.266	0.000	0.953	0.031	0.017
I1O3	0.5	hpds	0.210	0.790	0.000	0.995	0.005	0.000	0.377	0.623	0.000	0.953	0.026	0.021
I1O3	0.5	hpdv	0.276	0.725	0.000	1.000	0.000	0.000	0.701	0.299	0.000	0.958	0.019	0.023
I1O3	0.5	pct	0.043	0.957	0.000	0.965	0.035	0.000	0.108	0.893	0.000	0.954	0.030	0.016
I1O6	0.5	hpds	0.000	1.000	0.000	0.976	0.024	0.000	0.000	1.000	0.000	0.954	0.025	0.021
I1O6	0.5	hpdv	0.000	1.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.957	0.019	0.024
I1O6	0.5	pct	0.000	1.000	0.000	0.588	0.413	0.000	0.000	1.000	0.000	0.954	0.030	0.016
I2O3	0.5	hpds	0.013	0.987	0.000	0.958	0.042	0.000	0.006	0.994	0.000	0.954	0.024	0.022
I2O3	0.5	hpdv	0.023	0.977	0.000	1.000	0.000	0.000	0.203	0.797	0.000	0.956	0.019	0.025
I2O3	0.5	pct	0.001	0.999	0.000	0.880	0.120	0.000	0.000	1.000	0.000	0.954	0.030	0.016
I2O6	0.5	hpds	0.000	1.000	0.000	0.265	0.735	0.000	0.000	1.000	0.000	0.954	0.024	0.022
I2O6	0.5	hpdv	0.000	1.000	0.000	1.000	0.001	0.000	0.000	1.000	0.000	0.956	0.020	0.024
I2O6	0.5	pct	0.000	1.000	0.000	0.015	0.986	0.000	0.000	1.000	0.000	0.953	0.031	0.016
IG0	0.5	hpds	0.924	0.002	0.073	0.961	0.039	0.000	0.674	0.001	0.325	0.949	0.017	0.034
IG0	0.5	hpdv	0.923	0.001	0.076	0.991	0.009	0.000	0.650	0.000	0.350	0.947	0.013	0.040
IG0	0.5	pct	0.951	0.009	0.040	0.931	0.069	0.000	0.739	0.001	0.260	0.953	0.019	0.028
IhE6	0.5	hpds	0.948	0.052	0.000	0.996	0.004	0.000	0.985	0.015	0.000	0.954	0.024	0.022
IhE6	0.5	hpdv	0.957	0.043	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.955	0.020	0.025
IhE6	0.5	pct	0.845	0.155	0.000	0.983	0.017	0.000	0.969	0.031	0.000	0.954	0.030	0.016
I1E3	6	hpds	0.972	0.004	0.025	0.957	0.002	0.042	0.978	0.022	0.000	0.954	0.025	0.021
I1E3	6	hpdv	0.972	0.002	0.026	0.946	0.000	0.054	0.994	0.007	0.000	0.955	0.022	0.023
I1E3	6	pct	0.976	0.016	0.008	0.977	0.006	0.018	0.946	0.054	0.000	0.954	0.030	0.016
I1E6	6	hpds	0.994	0.006	0.000	0.998	0.001	0.001	0.867	0.133	0.000	0.953	0.026	0.020
I1E6	6	hpdv	0.997	0.004	0.000	0.997	0.000	0.002	0.960	0.040	0.000	0.954	0.023	0.023
I1E6	6	pct	0.967	0.033	0.000	0.995	0.006	0.000	0.708	0.292	0.000	0.952	0.031	0.016
I1O3	6	hpds	0.988	0.012	0.000	0.999	0.001	0.000	0.429	0.571	0.000	0.954	0.026	0.020
I1O3	6	hpdv	0.993	0.007	0.000	1.000	0.000	0.000	0.752	0.249	0.000	0.956	0.021	0.023
I1O3	6	pct	0.940	0.060	0.000	0.995	0.005	0.000	0.111	0.889	0.000	0.953	0.031	0.016
I1O6	6	hpds	0.911	0.089	0.000	0.999	0.001	0.000	0.000	1.000	0.000	0.955	0.025	0.020
I1O6	6	hpdv	0.941	0.059	0.000	1.000	0.000	0.000	0.001	0.999	0.000	0.955	0.022	0.023
I1O6	6	pct	0.717	0.283	0.000	0.995	0.005	0.000	0.000	1.000	0.000	0.953	0.031	0.016
I2O3	6	hpds	0.992	0.008	0.000	1.000	0.000	0.000	0.010	0.990	0.000	0.954	0.025	0.021
I2O3	6	hpdv	0.994	0.006	0.000	1.000	0.000	0.000	0.260	0.740	0.000	0.955	0.021	0.023
I2O3	6	pct	0.967	0.034	0.000	0.999	0.001	0.000	0.000	1.000	0.000	0.952	0.032	0.016
I2O6	6	hpds	0.801	0.199	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.954	0.025	0.021
I2O6	6	hpdv	0.871	0.129	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.954	0.022	0.024
I2O6	6	pct	0.527	0.473	0.000	0.998	0.002	0.000	0.000	1.000	0.000	0.953	0.031	0.016
IG0	6	hpds	0.887	0.001	0.112	0.922	0.004	0.074	0.802	0.001	0.196	0.947	0.019	0.034
IG0	6	hpdv	0.884	0.001	0.115	0.917	0.002	0.081	0.781	0.000	0.219	0.947	0.014	0.039
IG0	6	pct	0.928	0.007	0.065	0.954	0.011	0.035	0.847	0.003	0.150	0.950	0.023	0.027
IhE6	6	hpds	0.977	0.004	0.019	0.968	0.002	0.030	0.983	0.017	0.000	0.954	0.025	0.021
IhE6	6	hpdv	0.978	0.003	0.019	0.962	0.001	0.037	0.995	0.005	0.000	0.954	0.022	0.024
IhE6	6	pct	0.968	0.026	0.005	0.978	0.011	0.011	0.956	0.044	0.000	0.953	0.030	0.017
I1E3	24	hpds	0.879	0.001	0.120	0.882	0.001	0.118	0.980	0.020	0.000	0.953	0.024	0.023
I1E3	24	hpdv	0.876	0.000	0.124	0.866	0.000	0.134	0.995	0.005	0.000	0.954	0.020	0.026
I1E3	24	pct	0.922	0.008	0.070	0.915	0.006	0.078	0.943	0.057	0.000	0.954	0.028	0.017
I1E6	24	hpds	0.912	0.001	0.086	0.913	0.001	0.087	0.887	0.113	0.000	0.953	0.024	0.023
I1E6	24	hpdv	0.910	0.001	0.089	0.894	0.000	0.106	0.966	0.034	0.000	0.955	0.020	0.026
I1E6	24	pct	0.947	0.010	0.043	0.937	0.006	0.057	0.712	0.288	0.000	0.954	0.029	0.017
I1O3	24	hpds	0.943	0.002	0.055	0.939	0.001	0.061	0.476	0.524	0.000	0.952	0.025	0.023
I1O3	24	hpdv	0.942	0.000	0.058	0.918	0.000	0.081	0.780	0.220	0.000	0.954	0.021	0.025
I1O3	24	pct	0.965	0.012	0.023	0.958	0.006	0.036	0.110	0.890	0.000	0.954	0.029	0.017
I1O6	24	hpds	0.993	0.004	0.003	0.988	0.001	0.012	0.000	1.000	0.000	0.954	0.024	0.022
I1O6	24	hpdv	0.994	0.002	0.004	0.977	0.000	0.023	0.002	0.999	0.000	0.955	0.021	0.025
I1O6	24	pct	0.980	0.020	0.000	0.992	0.006	0.002	0.000	1.000	0.000	0.953	0.030	0.018
I2O3	24	hpds	0.763	0.000	0.237	0.735	0.000	0.265	0.013	0.987	0.000	0.954	0.023	0.024
I2O3	24	hpdv	0.755	0.000	0.245	0.693	0.000	0.307	0.282	0.718	0.000	0.953	0.020	0.027
I2O3	24	pct	0.853	0.002	0.145	0.792	0.000	0.207	0.000	1.000	0.000	0.953	0.029	0.018
I2O6	24	hpds	0.970	0.001	0.029	0.928	0.000	0.072	0.000	1.000	0.000	0.953	0.023	0.024
I2O6	24	hpdv	0.966	0.000	0.034	0.875	0.000	0.125	0.000	1.000	0.000	0.953	0.020	0.027
I2O6	24	pct	0.994	0.004	0.002	0.958	0.000	0.042	0.000	1.000	0.000	0.953	0.029	0.018
IG0	24	hpds	0.842	0.001	0.157	0.855	0.002	0.143	0.861	0.002	0.137	0.949	0.016	0.035
IG0	24	hpdv	0.838	0.001	0.161	0.835	0.001	0.164	0.847	0.001	0.152	0.947	0.014	0.040
IG0	24	pct	0.898	0.007	0.095	0.896	0.008	0.096	0.894	0.008	0.098	0.955	0.019	0.027
IhE6	24	hpds	0.925	0.002	0.073	0.928	0.002	0.070	0.986	0.014	0.000	0.953	0.024	0.024
IhE6	24	hpdv	0.925	0.001	0.075	0.919	0.001	0.080	0.996	0.004	0.000	0.953	0.020	0.026
IhE6	24	pct	0.948	0.014	0.038	0.945	0.012	0.043	0.954	0.046	0.000	0.954	0.028	0.018

TABLE F.2: **Coverage and proportion outside limits for batch variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
			Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside	
σ_b^2	Prior	Int.		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
0.5	FLAT	hpds	0.056	0.944	0.000	0.999	0.001	0.000	0.998	0.002	0.000	0.985	0.015	0.000	0.951	0.049	0.000	0.990	0.010	0.000	0.993	0.007	0.000
0.5	FLAT	hpdv	0.056	0.944	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.999	0.001	0.000	0.997	0.003	0.000	0.997	0.003	0.000	0.998	0.002	0.000
0.5	FLAT	pct	0.000	1.000	0.000	0.961	0.039	0.000	0.971	0.030	0.000	0.818	0.182	0.000	0.263	0.737	0.000	0.910	0.090	0.000	0.930	0.070	0.000
0.5	HCY	hpdl	0.998	0.002	0.000	0.996	0.004	0.000	0.990	0.010	0.000	0.974	0.026	0.000	0.947	0.053	0.000	0.971	0.029	0.000	0.983	0.018	0.000
0.5	HCY	hpds	1.000	0.000	0.000	1.000	0.000	0.000	0.998	0.002	0.000	0.993	0.008	0.000	0.993	0.008	0.000	0.981	0.006	0.013	0.996	0.005	0.000
0.5	HCY	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.999	0.001	0.000	0.986	0.001	0.013	0.999	0.001	0.000
0.5	HCY	pct	0.999	0.001	0.000	1.000	0.001	0.000	0.991	0.009	0.000	0.986	0.015	0.000	0.978	0.023	0.000	0.958	0.041	0.001	0.984	0.017	0.000
0.5	HT3	hpdl	0.997	0.004	0.000	0.996	0.004	0.000	0.989	0.012	0.000	0.974	0.026	0.000	0.945	0.056	0.000	0.971	0.030	0.000	0.983	0.018	0.000
0.5	HT3	hpds	1.000	0.000	0.000	1.000	0.000	0.000	0.998	0.002	0.000	0.993	0.008	0.000	0.993	0.007	0.000	0.983	0.007	0.011	0.996	0.005	0.000
0.5	HT3	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.001	0.000	0.988	0.001	0.011	0.999	0.001	0.000
0.5	HT3	pct	0.999	0.002	0.000	1.000	0.000	0.000	0.991	0.010	0.000	0.983	0.017	0.000	0.975	0.025	0.000	0.959	0.041	0.001	0.984	0.016	0.000
0.5	ICCU	hpdl	0.998	0.002	0.000	0.994	0.007	0.000	0.989	0.012	0.000	0.940	0.060	0.000	0.902	0.099	0.000	0.961	0.040	0.000	0.974	0.026	0.000
0.5	ICCU	hpds	1.000	0.000	0.000	1.000	0.000	0.000	0.998	0.002	0.000	0.987	0.013	0.000	0.986	0.015	0.000	0.989	0.012	0.000	0.994	0.006	0.000
0.5	ICCU	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.999	0.001	0.000	0.999	0.002	0.000	0.999	0.002	0.000
0.5	ICCU	pct	0.999	0.001	0.000	0.999	0.002	0.000	0.991	0.009	0.000	0.963	0.038	0.000	0.944	0.057	0.000	0.950	0.051	0.000	0.977	0.023	0.000
0.5	IG1	hpds	1.000	0.001	0.000	0.997	0.003	0.000	0.997	0.003	0.000	0.968	0.032	0.000	0.914	0.086	0.000	0.973	0.004	0.023	0.996	0.004	0.000
0.5	IG1	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.001	0.000	0.996	0.004	0.000	0.990	0.010	0.000	0.975	0.002	0.024	0.999	0.001	0.000
0.5	IG1	pct	0.978	0.022	0.000	0.985	0.015	0.000	0.986	0.014	0.000	0.915	0.085	0.000	0.676	0.324	0.000	0.974	0.021	0.004	0.985	0.015	0.000
0.5	IG2	hpds	1.000	0.000	0.000	0.997	0.003	0.000	0.998	0.003	0.000	0.967	0.033	0.000	0.913	0.087	0.000	0.978	0.004	0.018	0.996	0.004	0.000
0.5	IG2	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.001	0.000	0.996	0.004	0.000	0.992	0.008	0.000	0.980	0.002	0.019	0.999	0.001	0.000
0.5	IG2	pct	0.973	0.028	0.000	0.985	0.015	0.000	0.986	0.014	0.000	0.912	0.088	0.000	0.660	0.340	0.000	0.975	0.023	0.003	0.983	0.017	0.000
0.5	IG3	hpds	1.000	0.000	0.000	0.997	0.003	0.000	0.997	0.003	0.000	0.968	0.032	0.000	0.913	0.087	0.000	0.979	0.005	0.017	0.996	0.004	0.000
0.5	IG3	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.992	0.008	0.000	0.981	0.002	0.018	0.999	0.001	0.000
0.5	IG3	pct	0.971	0.029	0.000	0.985	0.015	0.000	0.985	0.015	0.000	0.911	0.089	0.000	0.660	0.340	0.000	0.975	0.023	0.003	0.983	0.017	0.000
0.5	IIN	hpds	1.000	0.000	0.000	0.998	0.002	0.000	0.996	0.004	0.000	0.968	0.032	0.000	0.914	0.086	0.000	0.978	0.005	0.017	0.996	0.004	0.000
0.5	IIN	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.997	0.003	0.000	0.991	0.009	0.000	0.980	0.002	0.018	0.999	0.001	0.000
0.5	IIN	pct	0.981	0.019	0.000	0.989	0.011	0.000	0.986	0.014	0.000	0.910	0.090	0.000	0.659	0.341	0.000	0.975	0.023	0.002	0.984	0.016	0.000
0.5	JEFF	hpds	1.000	0.000	0.000	0.997	0.003	0.000	0.997	0.003	0.000	0.968	0.032	0.000	0.912	0.088	0.000	0.978	0.004	0.018	0.996	0.004	0.000
0.5	JEFF	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.991	0.009	0.000	0.980	0.002	0.018	0.999	0.001	0.000
0.5	JEFF	pct	0.971	0.029	0.000	0.985	0.015	0.000	0.986	0.014	0.000	0.911	0.089	0.000	0.658	0.342	0.000	0.974	0.023	0.003	0.983	0.017	0.000
0.5	UNI	hpdl	0.994	0.006	0.000	0.994	0.007	0.000	0.988	0.013	0.000	0.961	0.039	0.000	0.922	0.078	0.000	0.970	0.031	0.000	0.983	0.018	0.000
0.5	UNI	hpds	0.998	0.003	0.000	1.000	0.000	0.000	0.998	0.003	0.000	0.991	0.009	0.000	0.950	0.050	0.000	0.982	0.006	0.012	0.996	0.005	0.000
0.5	UNI	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.987	0.001	0.012	0.999	0.001	0.000
0.5	UNI	pct	0.999	0.001	0.000	1.000	0.001	0.000	0.991	0.010	0.000	0.984	0.016	0.000	0.976	0.024	0.000	0.958	0.042	0.001	0.983	0.017	0.000
6	FLAT	hpds	0.205	0.795	0.000	0.999	0.002	0.000	0.994	0.004	0.001	0.997	0.003	0.000	0.994	0.006	0.000	0.987	0.011	0.003	0.988	0.011	0.001
6	FLAT	hpdv	0.205	0.795	0.000	1.000	0.000	0.000	0.998	0.001	0.001	1.000	0.000	0.000	1.000	0.000	0.000	0.994	0.003	0.003	0.995	0.004	0.001
6	FLAT	pct	0.241	0.759	0.000	0.979	0.021	0.000	0.937	0.063	0.000	0.964	0.036	0.000	0.937	0.063	0.000	0.899	0.101	0.001	0.913	0.087	0.000
6	HCY	hpdl	0.999	0.002	0.000	0.989	0.011	0.000	0.979	0.018	0.004	0.988	0.012	0.000	0.987	0.013	0.000	0.970	0.022	0.009	0.973	0.024	0.004
6	HCY	hpds	1.000	0.000	0.000	0.986	0.002	0.013	0.963	0.005	0.032	0.991	0.004	0.006	0.997	0.004	0.000	0.980	0.006	0.015	0.966	0.006	0.029
6	HCY	hpdv	1.000	0.000	0.000	0.988	0.000	0.013	0.967	0.001	0.032	0.994	0.000	0.006	1.000	0.001	0.000	0.979	0.002	0.020	0.971	0.001	0.029
6	HCY	pct	0.999	0.001	0.000	0.989	0.011	0.000	0.959	0.030	0.011	0.988	0.012	0.001	0.993	0.007	0.000	0.952	0.039	0.010	0.950	0.039	0.012
6	HT3	hpdl	0.998	0.003	0.000	0.990	0.011	0.000	0.976	0.021	0.004	0.987	0.014	0.000	0.986	0.015	0.000	0.966	0.026	0.009	0.970	0.026	0.004
6	HT3	hpds	1.000	0.000	0.000	0.987	0.003	0.011	0.964	0.006	0.031	0.989	0.006	0.006	0.996	0.004	0.000	0.979	0.006	0.015	0.966	0.007	0.028
6	HT3	hpdv	1.000	0.000	0.000	0.989	0.000	0.011	0.969	0.001	0.031	0.994	0.000	0.006	1.000	0.001	0.000	0.981	0.002	0.018	0.972	0.001	0.028
6	HT3	pct	0.999	0.001	0.000	0.989	0.011	0.000	0.958	0.032	0.011	0.987	0.012	0.001	0.993	0.008	0.000	0.950	0.041	0.009	0.949	0.040	0.012
6	ICCU	hpdl	0.999	0.001	0.001	0.997	0.004	0.000	0.989	0.007	0.005	0.994	0.007	0.000	0.994	0.007	0.000	0.980	0.008	0.013	0.985	0.009	0.007
6	ICCU	hpds	0.998	0.000	0.003	0.972	0.000	0.028	0.962	0.002	0.036	0.992	0.001	0.008	0.999	0.002	0.000	0.974	0.003	0.024	0.962	0.003	0.036
6	ICCU	hpdv	0.997	0.000	0.003	0.969	0.000	0.031	0.957	0.000	0.043	0.992	0.000	0.009	1.000	0.001	0.000	0.971	0.001	0.029	0.959	0.000	0.042
6	ICCU	pct	0.999	0.001	0.001	0.992	0.003	0.006	0.976	0.011	0.014	0.993	0.006	0.001	0.996	0.005	0.000	0.973	0.016	0.012	0.973	0.016	0.012
6	IG1	hpds	1.000	0.000	0.000	0.984	0.002	0.015	0.948	0.003	0.049	0.986	0.005	0.010	0.987	0.013	0.000	0.946	0.003	0.051	0.949	0.005	0.046
6	IG1	hpdv	1.000	0.000	0.000	0.984	0.000	0.015	0.944	0.001	0.055	0.988	0.001	0.010	0.998	0.002	0.000	0.944	0.001	0.055	0.949	0.001	0.050
6	IG1	pct	0.991	0.009	0.000	0.988	0.010	0.002	0.955	0.017	0.028	0.980	0.018	0.003	0.962	0.038	0.000	0.952	0.021	0.028	0.955	0.022	0.023
6	IG2	hpds	1.000	0.000	0.000	0.988	0.002	0.010	0.954	0.003	0.043	0.988	0.005	0.007	0.986	0.014	0.000	0.953	0.004	0.044	0.956	0.005	0.039
6	IG2	hpdv	1.000	0.000	0.000	0.989	0.000	0.011	0.951	0.001	0.048	0.991	0.002	0.007	0.998	0.002	0.000	0.950	0.001	0.048	0.956	0.001	0.043
6	IG2	pct	0.990	0.010	0.000	0.989	0.010	0.001	0.959	0.018	0.023	0.980	0.019	0.002	0.961	0.039	0.000	0.954	0.023	0.024	0.957	0.025	0.018
6	IG3	hpds	1.000	0.000	0.000	0.989	0.002	0.010	0.954	0.003	0.043	0.988	0.005	0.007	0.987	0.013	0.000	0.952	0.004	0.043	0.956	0.006	0.039
6	IG3	hpdv	1.000	0.000	0.000	0.989	0.000	0.011	0.951	0.001	0.048	0.991	0.001	0.008	0.998	0.002	0.000	0.951	0.002	0.047	0.957	0.001	0.041
6	IG3	pct	0.990	0.010	0.000	0.989	0.010	0.001	0.959	0.019	0.022	0.978	0.020	0.002	0.962	0.038	0.000	0.954	0.024	0.022	0.958	0.024	0.018
6	IIN	hpds	1.000	0.000	0.000	0.990	0.002	0.009	0.954	0.003	0.043	0.987	0.005	0.008	0.987	0.013	0.000	0.953	0.004	0.043	0.956	0.005	0.039
6	IIN	hpdv	1.000	0.000	0.000	0.991	0.000	0.009	0.951	0.001	0.047	0.991	0.001	0.008	0.998	0.002	0.000	0.951	0.001	0.048	0.956	0.001	0.042
6	IIN	pct	0.992	0.008	0.000	0.991	0.009	0.001	0.958	0.019	0.023	0.979	0.020	0.001	0.961	0.039	0.000	0.953	0.025	0.023	0.957	0.024	0.019
6	JEFF	hpds	1.000	0.000	0.000	0.988	0.002	0.010	0.955	0.003	0.042	0.988	0.005	0.007	0.987	0.013	0.000	0.953	0.004	0.043	0.956	0.006	0.039
6	JEFF	hpdv	1.000	0.000	0.000	0.989	0.000	0.010	0.952	0.001	0.046	0.991	0.001	0.008	0.998	0.002	0.000	0.951	0.002	0.048	0.957	0.001	0.042
6	JEFF	pct	0.990	0.010	0.000	0.989	0.010	0.001	0.959	0.019	0.022	0.979	0.020	0.001	0.961	0.039	0.000	0.954	0.023	0.023	0.957	0.025	0.018
6	UNI	hpdl	0.996	0.005	0.000	0.979	0.022	0.000	0.940	0.057	0.004	0.974	0.027	0.000	0.979	0.022	0.000	0.933	0.061	0.007	0.930	0.067	0.003
6	UNI	hpds	0.996	0.005	0.000	0.977	0.015	0.009	0.953	0.021	0.027	0.973	0.022	0.005	0.980	0.021	0.000	0.962	0.024	0.015	0.949	0.027	0.025
6	UNI	hpdv	0.997	0.004	0.000	0.986	0.006	0.009	0.967	0.007	0.027	0.988	0.008	0.005	0.985	0.016	0.000	0.978	0.007	0.016	0.965	0.010	0.025
6	UNI	pct	0.999	0.001	0.000	0.989	0.011	0.000	0.956	0.034	0.010	0.987	0.012	0.001	0.993	0.007	0.000	0.947	0.045	0.009	0.948	0.043	0.010
24	FLAT	hpds	0.360	0.640	0.000	0.995	0.005	0.000	0.990	0.008	0.003	0.996	0.004	0.000	0.997	0.003	0.000	0.985	0.012	0.003	0.985	0.012	0.003
24	FLAT	hpdv	0.360	0.640	0.000	0.999	0.001	0.000	0.995	0.002	0.003	0.999	0.001	0.000	1.000	0.000	0.000	0.993	0.004	0.003	0.993	0.004	0.003
24	FLAT	pct	0.504	0.496	0.000	0.951	0.049	0.000	0.914	0.086	0.001	0.948	0.052	0.000	0.962	0.038	0.000	0.894	0.106	0.001	0.903	0.096	0.001
24	HCY	hpdl	0.998	0.003	0.000	0.987	0.009	0.004	0.972	0.021	0.007	0.987	0.013	0.001	0.994	0.007	0.000	0.968	0.015	0.017	0.973	0.018	0.009
24	HCY	hpds	1.000	0.000	0.000	0.953	0.004	0.044	0.957	0.007	0.036	0.961	0.004	0.036	0.987	0.002	0.011	0.971	0.003	0.027	0.969	0.006	0.026
24	HCY	hpdv	1.000	0.000	0.000	0.956	0.001	0.044	0.960	0.000	0.040	0.965	0.000	0.036	0.989	0.000	0.011	0.969	0.000	0.031	0.972	0.001	0.028
24	HCY	pct	0.996	0.004	0.000	0.964	0.014	0.023	0.954	0.031	0.016	0.975	0.017	0.009	0.993	0.007	0.001	0.959	0.028	0.014	0.959	0.027	0.014
24	HT3	hpdl	0.997	0.004	0.000	0.987	0.010	0.003	0.973	0.021	0.006	0.986	0.013	0.002	0.993	0.008	0.000	0.967	0.018	0.016	0.969	0.021	0.011
24	HT3	hpds	1.000	0.000	0.000	0.955	0.005	0.041	0.960	0.008	0.033	0.964	0.004	0.032	0.989	0.002	0.010	0.969	0.005	0.026	0.970	0.007	0.024
24	HT3	hpdv	1.000	0.000	0.000	0.959	0.001	0.041	0.963	0.000	0.038	0.968	0.000	0.032	0.991	0.000	0.010	0.971	0.001	0.029	0.972	0.002	0.027
24	HT3	pct	0.997	0.003	0.000	0.968	0.014	0.019	0.955	0.032	0.014	0.976	0.016	0.009	0.993	0.006	0.001	0.960	0.027	0.013	0.959	0.028	0.013
24	ICCU	hpdl	0.977	0.000	0.023	0.953	0.005	0.043	0.944	0.006	0.051	0.964	0.005	0.031	0.985	0.002	0.013	0.936	0.004	0.061	0.936	0.006	0.059
24	ICCU	hpds	0.897	0.000	0.104	0.885	0.001	0.115	0.913	0.001	0.087	0.891	0.002	0.108	0.887	0.001	0.113	0.924	0.001	0.076	0.912	0.001	0.087
24	ICCU	hpdv	0.892	0.000	0.109	0.867	0.000	0.134	0.896	0.000	0.104	0.875	0.000	0.125	0.878	0.000	0.122	0.918	0.000	0.082	0.905	0.001	0.094
24	ICCU	pct	0.967	0.001	0.033	0.919	0.005	0.077	0.933	0.010	0.058	0.938	0.008	0.055	0.953	0.002	0.046	0.940	0.010	0.051	0.943	0.009	0.049
24	IG1	hpds	0.994	0.001	0.006	0.945	0.004	0.052	0.949	0.003	0.049	0.951	0.004	0.045	0.988	0.004	0.008	0.944	0.004	0.052	0.944	0.005	0.051
24	IG1	hpdv	0.994	0.000	0.006	0.942	0.001	0.057	0.945	0.001	0.054	0.950	0.001	0.049	0.991	0.001	0.008	0.945	0.002	0.054	0.944	0.002	0.055
24	IG1	pct	0.987	0.013	0.000	0.959	0.015	0.027	0.950	0.020	0.029	0.962	0.016	0.022	0.979	0.019	0.002	0.951	0.022	0.028	0.948	0.023	0.029
24	IG2	hpds	0.998	0.000	0.001	0.951	0.005	0.044	0.956	0.003	0.041	0.958	0.004	0.038	0.991	0.003	0.005	0.951	0.004	0.045	0.951	0.006	0.044
24	IG2	hpdv	0.999	0.000	0.002	0.951	0.001	0.047	0.951	0.001	0.048	0.958	0.001	0.041	0.993	0.001	0.006	0.951	0.001	0.048	0.951	0.002	0.048
24	IG2	pct	0.986	0.014	0.000	0.962	0.016	0.022	0.952	0.024	0.024	0.966	0.017	0.017	0.979	0.020	0.002	0.954	0.024	0.022	0.950	0.026	0.025
24	IG3	hpds	0.998	0.001	0.001	0.953	0.005	0.042	0.955	0.004	0.041	0.959	0.004	0.037	0.991	0.004	0.005	0.952	0.004	0.045	0.951	0.006	0.043
24	IG3	hpdv	0.999	0.000	0.001	0.952	0.001	0.047	0.952	0.001	0.047	0.959	0.001	0.040	0.993	0.001	0.006	0.952	0.002	0.046	0.951	0.002	0.047
24	IG3	pct	0.986	0.014	0.000	0.962	0.017	0.022	0.953	0.022	0.025	0.967	0.017	0.016	0.978	0.020	0.002	0.954	0.024	0.022	0.950	0.026	0.024
24	IIN	hpds	1.000	0.000	0.000	0.952	0.004	0.044	0.955	0.004	0.041	0.958	0.004	0.038	0.991	0.004	0.005	0.951	0.004	0.045	0.952	0.006	0.042
24	IIN	hpdv	1.000	0.000	0.000	0.953	0.001	0.046	0.953	0.001	0.046	0.959	0.001	0.040	0.994	0.001	0.005	0.953	0.001	0.046	0.951	0.002	0.047
24	IIN	pct	0.987	0.013	0.000	0.963	0.016	0.021	0.954	0.023	0.024	0.967	0.018	0.016	0.978	0.021	0.002	0.953	0.025	0.022	0.950	0.026	0.024
24	JEFF	hpds	0.999	0.000	0.001	0.953	0.004	0.043	0.956	0.003	0.041	0.959	0.004	0.037	0.990	0.005	0.006	0.951	0.004	0.045	0.952	0.005	0.043
24	JEFF	hpdv	0.999	0.000	0.001	0.952	0.001	0.047	0.952	0.001	0.047	0.959	0.001	0.040	0.993	0.001	0.006	0.952	0.001	0.047	0.951	0.002	0.047
24	JEFF	pct	0.986	0.015	0.000	0.964	0.016	0.020	0.953	0.023	0.024	0.967	0.017	0.016	0.978	0.021	0.002	0.955	0.023	0.022	0.949	0.026	0.024
24	UNI	hpdl	0.998	0.002	0.000	0.979	0.021	0.001	0.946	0.049	0.005	0.977	0.022	0.002	0.989	0.011	0.000	0.941	0.046	0.014	0.953	0.041	0.006
24	UNI	hpds	0.998	0.002	0.000	0.944	0.021	0.035	0.928	0.049	0.024	0.956	0.022	0.023	0.986	0.011	0.003	0.936	0.046	0.019	0.940	0.041	0.020
24	UNI	hpdv	0.998	0.002	0.000	0.944	0.021	0.035	0.921	0.049	0.031	0.956	0.022	0.023	0.986	0.011	0.003	0.934	0.046	0.021	0.938	0.041	0.022
24	UNI	pct	1.000	0.001	0.000	0.979	0.007	0.015	0.972	0.021	0.008	0.987	0.009	0.005	0.997	0.002	0.001	0.973	0.018	0.010	0.970	0.020	0.010

TABLE F.3: **Coverage and proportion outside limits for keg variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
			Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside	
σ_b^2	Prior	Int.		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
0.5	FLAT	hpds	0.679	0.321	0.000	0.973	0.027	0.000	0.982	0.018	0.000	0.979	0.009	0.012	0.976	0.008	0.017	0.950	0.034	0.017	0.958	0.042	0.000
0.5	FLAT	hpdv	0.707	0.293	0.000	0.998	0.003	0.000	0.995	0.005	0.000	0.980	0.004	0.016	0.978	0.003	0.019	0.959	0.014	0.027	0.987	0.014	0.000
0.5	FLAT	pct	0.357	0.643	0.000	0.937	0.063	0.000	0.951	0.049	0.000	0.957	0.036	0.006	0.959	0.035	0.006	0.947	0.041	0.012	0.962	0.038	0.000
0.5	HCY	hpdl	0.985	0.015	0.000	0.986	0.014	0.000	0.980	0.020	0.000	0.984	0.007	0.010	0.959	0.006	0.035	0.973	0.022	0.005	0.980	0.020	0.000
0.5	HCY	hpds	0.997	0.003	0.000	0.995	0.006	0.000	0.992	0.003	0.006	0.965	0.003	0.032	0.947	0.002	0.052	0.935	0.013	0.053	0.988	0.012	0.000
0.5	HCY	hpdv	1.000	0.000	0.000	1.000	0.001	0.000	0.994	0.001	0.006	0.956	0.002	0.042	0.939	0.002	0.060	0.939	0.008	0.054	0.994	0.007	0.000
0.5	HCY	pct	0.994	0.007	0.000	0.994	0.006	0.000	0.982	0.019	0.000	0.969	0.010	0.022	0.962	0.010	0.028	0.948	0.021	0.032	0.991	0.010	0.000
0.5	HT3	hpdl	0.985	0.016	0.000	0.987	0.014	0.000	0.981	0.020	0.000	0.982	0.007	0.011	0.963	0.006	0.032	0.971	0.023	0.006	0.979	0.022	0.000
0.5	HT3	hpds	0.997	0.003	0.000	0.995	0.006	0.000	0.992	0.003	0.006	0.966	0.003	0.031	0.949	0.003	0.049	0.933	0.015	0.053	0.988	0.013	0.000
0.5	HT3	hpdv	1.000	0.000	0.000	1.000	0.001	0.000	0.994	0.001	0.006	0.957	0.002	0.041	0.944	0.002	0.055	0.939	0.008	0.053	0.994	0.006	0.000
0.5	HT3	pct	0.994	0.007	0.000	0.993	0.007	0.000	0.981	0.019	0.000	0.969	0.010	0.022	0.963	0.011	0.027	0.947	0.021	0.032	0.991	0.010	0.000
0.5	ICCU	hpdl	0.960	0.041	0.000	0.968	0.032	0.000	0.973	0.028	0.000	0.973	0.004	0.023	0.908	0.002	0.090	0.969	0.030	0.001	0.942	0.058	0.000
0.5	ICCU	hpds	0.996	0.005	0.000	0.987	0.013	0.000	0.995	0.005	0.000	0.953	0.002	0.045	0.887	0.002	0.112	0.966	0.016	0.019	0.965	0.035	0.000
0.5	ICCU	hpdv	1.000	0.001	0.000	0.998	0.002	0.000	0.999	0.001	0.000	0.939	0.001	0.060	0.874	0.001	0.126	0.960	0.009	0.032	0.990	0.011	0.000
0.5	ICCU	pct	0.982	0.019	0.000	0.980	0.021	0.000	0.974	0.027	0.000	0.970	0.005	0.026	0.924	0.004	0.073	0.961	0.027	0.013	0.968	0.032	0.000
0.5	IG1	hpds	0.986	0.014	0.000	0.984	0.016	0.000	0.988	0.003	0.010	0.872	0.002	0.127	0.848	0.002	0.151	0.951	0.015	0.034	0.967	0.033	0.000
0.5	IG1	hpdv	0.998	0.002	0.000	0.997	0.003	0.000	0.988	0.001	0.011	0.855	0.001	0.145	0.835	0.001	0.164	0.941	0.006	0.053	0.986	0.014	0.000
0.5	IG1	pct	0.961	0.039	0.000	0.975	0.025	0.000	0.992	0.007	0.001	0.910	0.004	0.086	0.894	0.004	0.102	0.951	0.021	0.028	0.969	0.031	0.000
0.5	IG2	hpds	0.985	0.015	0.000	0.984	0.016	0.000	0.990	0.003	0.007	0.887	0.002	0.111	0.866	0.002	0.132	0.951	0.016	0.033	0.967	0.033	0.000
0.5	IG2	hpdv	0.998	0.002	0.000	0.997	0.003	0.000	0.991	0.001	0.008	0.871	0.001	0.128	0.855	0.001	0.144	0.943	0.008	0.050	0.987	0.013	0.000
0.5	IG2	pct	0.956	0.044	0.000	0.974	0.026	0.000	0.992	0.007	0.001	0.920	0.006	0.074	0.906	0.005	0.090	0.952	0.022	0.027	0.968	0.032	0.000
0.5	IG3	hpds	0.984	0.016	0.000	0.984	0.016	0.000	0.990	0.003	0.007	0.890	0.002	0.109	0.866	0.002	0.132	0.951	0.017	0.032	0.966	0.034	0.000
0.5	IG3	hpdv	0.998	0.002	0.000	0.996	0.004	0.000	0.991	0.001	0.008	0.873	0.001	0.127	0.857	0.001	0.142	0.944	0.007	0.049	0.986	0.014	0.000
0.5	IG3	pct	0.957	0.043	0.000	0.974	0.026	0.000	0.992	0.008	0.001	0.921	0.005	0.074	0.908	0.004	0.088	0.952	0.022	0.026	0.968	0.032	0.000
0.5	IIN	hpds	0.998	0.002	0.000	0.996	0.005	0.000	0.990	0.003	0.007	0.888	0.002	0.110	0.867	0.002	0.132	0.951	0.016	0.033	0.981	0.019	0.000
0.5	IIN	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	0.991	0.001	0.008	0.874	0.001	0.125	0.856	0.001	0.143	0.942	0.007	0.050	0.994	0.006	0.000
0.5	IIN	pct	0.987	0.014	0.000	0.991	0.009	0.000	0.991	0.008	0.001	0.921	0.006	0.073	0.907	0.005	0.088	0.951	0.022	0.027	0.983	0.017	0.000
0.5	JEFF	hpds	0.987	0.014	0.000	0.983	0.017	0.000	0.990	0.003	0.007	0.890	0.002	0.108	0.868	0.002	0.130	0.951	0.016	0.033	0.965	0.036	0.000
0.5	JEFF	hpdv	0.998	0.002	0.000	0.997	0.003	0.000	0.991	0.001	0.008	0.873	0.001	0.126	0.857	0.001	0.142	0.944	0.008	0.049	0.986	0.014	0.000
0.5	JEFF	pct	0.957	0.043	0.000	0.976	0.025	0.000	0.990	0.009	0.001	0.921	0.005	0.074	0.907	0.005	0.088	0.951	0.022	0.027	0.968	0.032	0.000
0.5	UNI	hpdl*	0.977	0.024	0.000	0.985	0.015	0.000	0.981	0.020	0.000	0.980	0.010	0.011	0.948	0.026	0.027	0.974	0.022	0.005	0.980	0.020	0.000
0.5	UNI	hpds	0.995	0.006	0.000	0.995	0.006	0.000	0.992	0.003	0.006	0.968	0.004	0.029	0.933	0.026	0.042	0.934	0.013	0.053	0.988	0.013	0.000
0.5	UNI	hpdv	1.000	0.001	0.000	0.999	0.001	0.000	0.994	0.001	0.006	0.959	0.002	0.039	0.932	0.022	0.047	0.939	0.008	0.054	0.994	0.006	0.000
0.5	UNI	pct	0.992	0.009	0.000	0.994	0.007	0.000	0.981	0.019	0.000	0.968	0.012	0.021	0.969	0.010	0.022	0.947	0.021	0.032	0.990	0.010	0.000
6	FLAT	hpds	0.677	0.323	0.000	0.957	0.044	0.000	0.974	0.027	0.000	0.973	0.020	0.008	0.977	0.011	0.012	0.947	0.038	0.015	0.954	0.046	0.000
6	FLAT	hpdv	0.706	0.294	0.000	0.997	0.004	0.000	0.993	0.007	0.000	0.983	0.008	0.009	0.981	0.005	0.014	0.962	0.013	0.025	0.980	0.020	0.000
6	FLAT	pct	0.352	0.648	0.000	0.897	0.103	0.000	0.911	0.089	0.000	0.936	0.061	0.003	0.952	0.043	0.005	0.947	0.043	0.010	0.955	0.045	0.000
6	HCY	hpdl	0.979	0.022	0.000	0.972	0.029	0.000	0.967	0.034	0.000	0.962	0.033	0.006	0.964	0.012	0.025	0.972	0.025	0.004	0.979	0.021	0.000
6	HCY	hpds	0.998	0.003	0.000	0.993	0.008	0.000	0.988	0.013	0.000	0.960	0.015	0.025	0.961	0.005	0.034	0.922	0.017	0.062	0.989	0.011	0.000
6	HCY	hpdv	1.000	0.000	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.965	0.005	0.030	0.960	0.001	0.040	0.929	0.009	0.063	0.998	0.002	0.000
6	HCY	pct	0.991	0.010	0.000	0.988	0.013	0.000	0.968	0.033	0.000	0.946	0.040	0.015	0.962	0.018	0.021	0.944	0.025	0.032	0.993	0.008	0.000
6	HT3	hpdl	0.975	0.025	0.000	0.969	0.031	0.000	0.966	0.035	0.000	0.960	0.034	0.007	0.965	0.013	0.023	0.968	0.026	0.007	0.982	0.019	0.000
6	HT3	hpds	0.997	0.003	0.000	0.992	0.009	0.000	0.989	0.012	0.000	0.960	0.017	0.024	0.963	0.005	0.032	0.922	0.017	0.062	0.989	0.011	0.000
6	HT3	hpdv	1.000	0.001	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.969	0.005	0.027	0.963	0.002	0.036	0.929	0.008	0.063	0.998	0.002	0.000
6	HT3	pct	0.991	0.010	0.000	0.986	0.015	0.000	0.968	0.033	0.000	0.945	0.040	0.015	0.962	0.019	0.020	0.944	0.025	0.032	0.993	0.007	0.000
6	ICCU	hpdl	0.949	0.052	0.000	0.932	0.069	0.000	0.949	0.052	0.000	0.967	0.021	0.013	0.939	0.005	0.057	0.966	0.033	0.002	0.960	0.041	0.000
6	ICCU	hpds	0.996	0.004	0.000	0.984	0.016	0.000	0.982	0.019	0.000	0.964	0.009	0.027	0.916	0.003	0.082	0.965	0.020	0.016	0.976	0.024	0.000
6	ICCU	hpdv	0.999	0.001	0.000	1.000	0.000	0.000	0.996	0.005	0.000	0.959	0.004	0.038	0.906	0.000	0.094	0.958	0.014	0.029	0.994	0.006	0.000
6	ICCU	pct	0.971	0.030	0.000	0.967	0.034	0.000	0.951	0.049	0.000	0.959	0.025	0.017	0.945	0.007	0.049	0.961	0.031	0.009	0.977	0.023	0.000
6	IG1	hpds	0.979	0.022	0.000	0.968	0.033	0.000	0.992	0.008	0.000	0.924	0.004	0.072	0.875	0.001	0.124	0.951	0.017	0.032	0.956	0.044	0.000
6	IG1	hpdv	0.996	0.004	0.000	0.995	0.005	0.000	0.998	0.002	0.000	0.916	0.001	0.083	0.864	0.000	0.135	0.944	0.006	0.051	0.979	0.021	0.000
6	IG1	pct	0.946	0.054	0.000	0.949	0.051	0.000	0.981	0.019	0.000	0.942	0.012	0.046	0.913	0.007	0.080	0.953	0.023	0.024	0.958	0.042	0.000
6	IG2	hpds	0.978	0.022	0.000	0.968	0.032	0.000	0.992	0.008	0.000	0.932	0.005	0.063	0.889	0.002	0.109	0.951	0.019	0.030	0.957	0.043	0.000
6	IG2	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.998	0.002	0.000	0.925	0.002	0.073	0.881	0.001	0.118	0.945	0.006	0.049	0.979	0.021	0.000
6	IG2	pct	0.940	0.060	0.000	0.947	0.053	0.000	0.979	0.021	0.000	0.945	0.014	0.041	0.924	0.007	0.069	0.953	0.024	0.023	0.957	0.043	0.000
6	IG3	hpds	0.977	0.023	0.000	0.968	0.032	0.000	0.991	0.009	0.000	0.934	0.004	0.062	0.890	0.002	0.107	0.952	0.018	0.030	0.957	0.043	0.000
6	IG3	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.998	0.002	0.000	0.926	0.002	0.073	0.881	0.001	0.118	0.945	0.007	0.049	0.980	0.020	0.000
6	IG3	pct	0.941	0.059	0.000	0.946	0.054	0.000	0.978	0.022	0.000	0.946	0.014	0.040	0.925	0.007	0.068	0.954	0.024	0.022	0.958	0.042	0.000
6	IIN	hpds	0.997	0.003	0.000	0.988	0.012	0.000	0.992	0.008	0.000	0.933	0.005	0.062	0.891	0.002	0.107	0.951	0.018	0.031	0.971	0.029	0.000
6	IIN	hpdv	1.000	0.000	0.000	0.999	0.001	0.000	0.998	0.002	0.000	0.926	0.002	0.072	0.882	0.000	0.118	0.944	0.007	0.049	0.989	0.011	0.000
6	IIN	pct	0.979	0.021	0.000	0.976	0.024	0.000	0.979	0.021	0.000	0.946	0.014	0.040	0.925	0.007	0.067	0.955	0.023	0.022	0.973	0.027	0.000
6	JEFF	hpds	0.981	0.019	0.000	0.968	0.032	0.000	0.991	0.009	0.000	0.935	0.005	0.060	0.891	0.002	0.108	0.953	0.018	0.029	0.957	0.043	0.000
6	JEFF	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.997	0.003	0.000	0.926	0.002	0.072	0.882	0.001	0.117	0.946	0.006	0.049	0.981	0.019	0.000
6	JEFF	pct	0.941	0.059	0.000	0.947	0.053	0.000	0.979	0.021	0.000	0.946	0.014	0.040	0.924	0.008	0.068	0.955	0.023	0.022	0.959	0.041	0.000
6	UNI	hpdl	0.956	0.045	0.000	0.969	0.031	0.000	0.966	0.034	0.000	0.955	0.040	0.005	0.940	0.041	0.020	0.971	0.027	0.003	0.981	0.020	0.000
6	UNI	hpds	0.992	0.008	0.000	0.992	0.009	0.000	0.989	0.012	0.000	0.954	0.022	0.025	0.931	0.041	0.029	0.922	0.016	0.062	0.988	0.012	0.000
6	UNI	hpdv	0.999	0.001	0.000	1.000	0.000	0.000	0.996	0.004	0.000	0.965	0.008	0.027	0.933	0.038	0.030	0.928	0.010	0.063	0.998	0.002	0.000
6	UNI	pct	0.990	0.010	0.000	0.987	0.014	0.000	0.967	0.033	0.000	0.941	0.045	0.015	0.970	0.016	0.015	0.944	0.026	0.031	0.992	0.009	0.000
24	FLAT	hpds	0.708	0.292	0.000	0.959	0.041	0.000	0.977	0.023	0.000	0.971	0.024	0.005	0.975	0.018	0.007	0.942	0.043	0.015	0.946	0.054	0.000
24	FLAT	hpdv	0.736	0.264	0.000	0.996	0.004	0.000	0.993	0.007	0.000	0.984	0.010	0.006	0.984	0.008	0.008	0.957	0.018	0.025	0.978	0.022	0.000
24	FLAT	pct	0.292	0.708	0.000	0.877	0.124	0.000	0.908	0.092	0.000	0.916	0.083	0.002	0.935	0.062	0.003	0.942	0.048	0.010	0.946	0.054	0.000
24	HCY	hpdl	0.967	0.034	0.000	0.969	0.032	0.000	0.972	0.029	0.000	0.964	0.034	0.003	0.963	0.027	0.011	0.958	0.037	0.006	0.978	0.022	0.000
24	HCY	hpds	0.995	0.006	0.000	0.987	0.014	0.000	0.993	0.007	0.000	0.965	0.017	0.019	0.972	0.011	0.018	0.924	0.022	0.054	0.985	0.015	0.000
24	HCY	hpdv	1.000	0.000	0.000	0.998	0.002	0.000	0.999	0.002	0.000	0.974	0.005	0.022	0.977	0.003	0.020	0.933	0.011	0.056	0.996	0.005	0.000
24	HCY	pct	0.980	0.020	0.000	0.980	0.020	0.000	0.969	0.031	0.000	0.946	0.042	0.013	0.960	0.033	0.007	0.933	0.036	0.032	0.989	0.011	0.000
24	HT3	hpdl	0.964	0.036	0.000	0.967	0.034	0.000	0.971	0.029	0.000	0.962	0.034	0.004	0.962	0.028	0.011	0.958	0.037	0.005	0.978	0.023	0.000
24	HT3	hpds	0.995	0.006	0.000	0.987	0.013	0.000	0.993	0.008	0.000	0.966	0.017	0.018	0.972	0.011	0.018	0.923	0.022	0.055	0.985	0.016	0.000
24	HT3	hpdv	1.000	0.000	0.000	0.998	0.002	0.000	0.999	0.001	0.000	0.973	0.007	0.021	0.978	0.004	0.019	0.931	0.012	0.058	0.996	0.005	0.000
24	HT3	pct	0.981	0.019	0.000	0.979	0.022	0.000	0.968	0.032	0.000	0.944	0.044	0.012	0.959	0.035	0.007	0.933	0.036	0.032	0.989	0.011	0.000
24	ICCU	hpdl	0.928	0.073	0.000	0.941	0.059	0.000	0.952	0.048	0.000	0.972	0.022	0.007	0.957	0.015	0.029	0.952	0.045	0.003	0.959	0.041	0.000
24	ICCU	hpds	0.994	0.007	0.000	0.986	0.015	0.000	0.989	0.012	0.000	0.969	0.010	0.021	0.956	0.007	0.038	0.952	0.031	0.018	0.974	0.027	0.000
24	ICCU	hpdv	1.000	0.000	0.000	0.999	0.002	0.000	0.998	0.002	0.000	0.972	0.002	0.026	0.954	0.003	0.044	0.956	0.015	0.030	0.992	0.008	0.000
24	ICCU	pct	0.961	0.039	0.000	0.962	0.038	0.000	0.948	0.052	0.000	0.962	0.026	0.013	0.956	0.022	0.023	0.947	0.042	0.012	0.974	0.026	0.000
24	IG1	hpds	0.977	0.023	0.000	0.966	0.034	0.000	0.994	0.006	0.000	0.942	0.008	0.050	0.921	0.005	0.075	0.947	0.022	0.031	0.949	0.051	0.000
24	IG1	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.999	0.001	0.000	0.940	0.003	0.057	0.917	0.002	0.081	0.941	0.008	0.051	0.978	0.022	0.000
24	IG1	pct	0.928	0.073	0.000	0.940	0.061	0.000	0.979	0.021	0.000	0.952	0.020	0.029	0.945	0.012	0.043	0.949	0.028	0.023	0.950	0.050	0.000
24	IG2	hpds	0.976	0.024	0.000	0.967	0.033	0.000	0.993	0.007	0.000	0.948	0.008	0.044	0.930	0.005	0.065	0.947	0.024	0.029	0.948	0.052	0.000
24	IG2	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.999	0.001	0.000	0.945	0.003	0.052	0.928	0.002	0.070	0.943	0.009	0.048	0.980	0.020	0.000
24	IG2	pct	0.923	0.077	0.000	0.937	0.063	0.000	0.978	0.022	0.000	0.954	0.022	0.024	0.949	0.013	0.038	0.948	0.029	0.023	0.949	0.051	0.000
24	IG3	hpds	0.977	0.023	0.000	0.967	0.033	0.000	0.994	0.006	0.000	0.948	0.009	0.043	0.932	0.005	0.063	0.949	0.022	0.029	0.949	0.051	0.000
24	IG3	hpdv	0.997	0.003	0.000	0.995	0.005	0.000	0.998	0.002	0.000	0.945	0.003	0.051	0.928	0.002	0.070	0.943	0.009	0.048	0.980	0.020	0.000
24	IG3	pct	0.922	0.078	0.000	0.937	0.064	0.000	0.978	0.022	0.000	0.956	0.021	0.023	0.948	0.015	0.037	0.949	0.029	0.022	0.950	0.050	0.000
24	IIN	hpds	0.996	0.005	0.000	0.987	0.013	0.000	0.993	0.007	0.000	0.948	0.009	0.043	0.932	0.005	0.063	0.948	0.022	0.030	0.969	0.031	0.000
24	IIN	hpdv	1.000	0.000	0.000	0.999	0.002	0.000	0.999	0.002	0.000	0.946	0.003	0.050	0.928	0.002	0.070	0.943	0.009	0.048	0.990	0.011	0.000
24	IIN	pct	0.967	0.033	0.000	0.969	0.031	0.000	0.978	0.022	0.000	0.954	0.022	0.024	0.948	0.014	0.038	0.949	0.028	0.023	0.970	0.030	0.000
24	JEFF	hpds	0.978	0.022	0.000	0.966	0.034	0.000	0.993	0.007	0.000	0.948	0.010	0.043	0.931	0.005	0.064	0.948	0.024	0.029	0.949	0.051	0.000
24	JEFF	hpdv	0.998	0.003	0.000	0.994	0.006	0.000	0.999	0.001	0.000	0.946	0.003	0.051	0.929	0.002	0.070	0.944	0.009	0.047	0.979	0.021	0.000
24	JEFF	pct	0.923	0.077	0.000	0.938	0.062	0.000	0.978	0.022	0.000	0.957	0.021	0.023	0.949	0.014	0.038	0.949	0.029	0.023	0.950	0.050	0.000
24	UNI	hpdl	0.943	0.057	0.000	0.962	0.038	0.000	0.973	0.028	0.000	0.941	0.057	0.003	0.936	0.058	0.007	0.959	0.037	0.005	0.978	0.023	0.000
24	UNI	hpds	0.978	0.023	0.000	0.986	0.015	0.000	0.994	0.007	0.000	0.960	0.024	0.017	0.930	0.058	0.013	0.922	0.023	0.056	0.984	0.016	0.000
24	UNI	hpdv	1.000	0.001	0.000	0.998	0.002	0.000	0.999	0.001	0.000	0.970	0.012	0.019	0.930	0.056	0.015	0.932	0.012	0.057	0.996	0.004	0.000
24	UNI	pct	0.979	0.022	0.000	0.978	0.022	0.000	0.967	0.033	0.000	0.941	0.049	0.011	0.966	0.029	0.006	0.933	0.037	0.030	0.988	0.012	0.000

TABLE F.4: **Coverage and proportion outside limits for portion variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
			Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside	
σ_b^2	Prior	Int.		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
0.5	FLAT	hpds	0.945	0.045	0.010	0.960	0.009	0.031	0.952	0.025	0.023	0.951	0.025	0.024	0.947	0.027	0.026	0.948	0.030	0.022	0.951	0.006	0.043
0.5	FLAT	hpdv	0.960	0.030	0.011	0.959	0.005	0.036	0.953	0.021	0.026	0.953	0.020	0.027	0.949	0.022	0.029	0.948	0.028	0.024	0.946	0.004	0.050
0.5	FLAT	pct	0.914	0.082	0.004	0.962	0.020	0.018	0.952	0.029	0.019	0.953	0.028	0.019	0.949	0.031	0.021	0.949	0.030	0.020	0.959	0.009	0.032
0.5	HCY	hpdl	0.958	0.017	0.025	0.954	0.011	0.036	0.958	0.020	0.022	0.942	0.029	0.030	0.952	0.026	0.023	0.953	0.025	0.022	0.964	0.012	0.024
0.5	HCY	hpds	0.958	0.007	0.035	0.943	0.007	0.051	0.962	0.015	0.023	0.942	0.027	0.032	0.952	0.023	0.026	0.954	0.024	0.023	0.960	0.009	0.032
0.5	HCY	hpdv	0.961	0.002	0.038	0.940	0.004	0.056	0.964	0.014	0.023	0.944	0.024	0.032	0.953	0.021	0.027	0.954	0.024	0.023	0.961	0.007	0.033
0.5	HCY	pct	0.958	0.025	0.018	0.961	0.013	0.027	0.961	0.023	0.017	0.943	0.032	0.026	0.953	0.028	0.020	0.955	0.027	0.019	0.966	0.013	0.021
0.5	HT3	hpdl	0.958	0.017	0.025	0.953	0.011	0.037	0.959	0.020	0.022	0.943	0.028	0.030	0.951	0.026	0.023	0.954	0.025	0.022	0.965	0.013	0.023
0.5	HT3	hpds	0.961	0.007	0.033	0.943	0.007	0.051	0.961	0.017	0.023	0.943	0.026	0.032	0.952	0.024	0.025	0.952	0.024	0.024	0.960	0.011	0.030
0.5	HT3	hpdv	0.962	0.002	0.037	0.942	0.003	0.056	0.962	0.015	0.024	0.944	0.024	0.033	0.953	0.020	0.028	0.954	0.024	0.023	0.961	0.006	0.033
0.5	HT3	pct	0.959	0.025	0.017	0.960	0.013	0.028	0.961	0.022	0.017	0.941	0.033	0.027	0.953	0.028	0.020	0.955	0.027	0.019	0.965	0.014	0.022
0.5	ICCU	hpdl	0.932	0.008	0.061	0.908	0.005	0.088	0.956	0.011	0.034	0.943	0.027	0.031	0.954	0.026	0.021	0.946	0.022	0.032	0.932	0.004	0.065
0.5	ICCU	hpds	0.918	0.003	0.080	0.893	0.003	0.105	0.954	0.009	0.038	0.942	0.023	0.036	0.955	0.023	0.023	0.947	0.020	0.034	0.920	0.002	0.078
0.5	ICCU	hpdv	0.911	0.001	0.089	0.885	0.001	0.115	0.953	0.008	0.040	0.942	0.021	0.037	0.955	0.022	0.023	0.949	0.019	0.033	0.915	0.002	0.084
0.5	ICCU	pct	0.941	0.010	0.050	0.919	0.007	0.075	0.959	0.013	0.029	0.946	0.027	0.028	0.952	0.028	0.020	0.952	0.022	0.027	0.939	0.005	0.057
0.5	IG1	hpds	0.828	0.001	0.171	0.848	0.002	0.151	0.950	0.017	0.033	0.952	0.018	0.030	0.948	0.021	0.032	0.948	0.025	0.027	0.904	0.002	0.094
0.5	IG1	hpdv	0.818	0.000	0.182	0.835	0.001	0.165	0.949	0.015	0.036	0.952	0.015	0.033	0.948	0.017	0.036	0.949	0.022	0.029	0.893	0.002	0.105
0.5	IG1	pct	0.883	0.004	0.113	0.892	0.003	0.105	0.953	0.021	0.026	0.953	0.023	0.024	0.949	0.025	0.026	0.951	0.025	0.024	0.923	0.003	0.075
0.5	IG2	hpds	0.848	0.001	0.151	0.860	0.002	0.138	0.950	0.018	0.033	0.951	0.020	0.029	0.948	0.021	0.031	0.949	0.025	0.026	0.908	0.002	0.090
0.5	IG2	hpdv	0.838	0.001	0.161	0.848	0.000	0.152	0.949	0.015	0.036	0.951	0.016	0.034	0.947	0.018	0.035	0.949	0.023	0.028	0.897	0.002	0.102
0.5	IG2	pct	0.898	0.005	0.096	0.900	0.003	0.096	0.953	0.021	0.026	0.954	0.023	0.023	0.948	0.026	0.026	0.951	0.026	0.022	0.926	0.003	0.071
0.5	IG3	hpds	0.849	0.001	0.150	0.861	0.002	0.137	0.949	0.019	0.032	0.951	0.021	0.028	0.945	0.022	0.032	0.949	0.025	0.026	0.909	0.002	0.089
0.5	IG3	hpdv	0.839	0.001	0.160	0.849	0.001	0.151	0.949	0.015	0.036	0.951	0.017	0.033	0.947	0.018	0.036	0.950	0.023	0.028	0.899	0.001	0.100
0.5	IG3	pct	0.900	0.005	0.094	0.900	0.004	0.097	0.953	0.021	0.026	0.954	0.023	0.022	0.949	0.026	0.025	0.951	0.026	0.023	0.926	0.003	0.071
0.5	IIN	hpds	0.995	0.003	0.002	0.967	0.003	0.030	0.955	0.022	0.023	0.954	0.024	0.021	0.950	0.027	0.023	0.950	0.028	0.022	0.951	0.003	0.045
0.5	IIN	hpdv	0.996	0.001	0.003	0.962	0.002	0.036	0.955	0.018	0.027	0.956	0.019	0.025	0.952	0.021	0.027	0.950	0.027	0.024	0.944	0.003	0.053
0.5	IIN	pct	0.992	0.008	0.000	0.977	0.006	0.017	0.955	0.027	0.018	0.953	0.029	0.018	0.951	0.031	0.018	0.952	0.029	0.019	0.962	0.005	0.033
0.5	JEFF	hpds	0.848	0.001	0.151	0.860	0.002	0.138	0.949	0.018	0.033	0.951	0.020	0.029	0.947	0.022	0.032	0.950	0.025	0.026	0.907	0.003	0.090
0.5	JEFF	hpdv	0.836	0.001	0.164	0.849	0.001	0.151	0.949	0.015	0.036	0.951	0.016	0.033	0.947	0.017	0.036	0.950	0.022	0.028	0.898	0.002	0.101
0.5	JEFF	pct	0.898	0.005	0.097	0.902	0.003	0.095	0.953	0.021	0.026	0.954	0.023	0.023	0.949	0.026	0.025	0.951	0.026	0.023	0.927	0.003	0.070
0.5	UNI	hpdl	0.958	0.016	0.026	0.953	0.011	0.037	0.959	0.020	0.022	0.944	0.029	0.028	0.952	0.026	0.023	0.953	0.025	0.022	0.965	0.012	0.024
0.5	UNI	hpds	0.958	0.007	0.036	0.942	0.007	0.051	0.961	0.016	0.024	0.943	0.026	0.032	0.951	0.023	0.027	0.954	0.023	0.024	0.961	0.010	0.030
0.5	UNI	hpdv	0.961	0.002	0.037	0.940	0.004	0.056	0.962	0.015	0.024	0.944	0.023	0.033	0.952	0.021	0.027	0.953	0.023	0.025	0.961	0.006	0.034
0.5	UNI	pct	0.959	0.025	0.017	0.961	0.013	0.027	0.962	0.023	0.016	0.943	0.032	0.026	0.953	0.028	0.020	0.955	0.027	0.019	0.965	0.014	0.022
6	FLAT	hpds	0.939	0.050	0.011	0.965	0.010	0.026	0.950	0.029	0.022	0.949	0.029	0.022	0.948	0.027	0.026	0.947	0.031	0.023	0.949	0.009	0.042
6	FLAT	hpdv	0.952	0.036	0.012	0.966	0.006	0.029	0.953	0.023	0.024	0.950	0.024	0.026	0.950	0.022	0.028	0.948	0.028	0.024	0.945	0.006	0.049
6	FLAT	pct	0.903	0.093	0.005	0.962	0.025	0.013	0.952	0.031	0.017	0.949	0.034	0.017	0.950	0.029	0.021	0.949	0.030	0.022	0.957	0.012	0.030
6	HCY	hpdl	0.962	0.017	0.022	0.961	0.016	0.023	0.948	0.026	0.026	0.953	0.021	0.027	0.952	0.022	0.027	0.950	0.027	0.024	0.956	0.017	0.028
6	HCY	hpds	0.963	0.007	0.031	0.959	0.008	0.034	0.945	0.025	0.031	0.951	0.020	0.029	0.950	0.021	0.030	0.951	0.025	0.025	0.954	0.013	0.034
6	HCY	hpdv	0.963	0.003	0.034	0.958	0.004	0.039	0.946	0.023	0.031	0.953	0.017	0.031	0.951	0.019	0.031	0.951	0.024	0.025	0.951	0.011	0.039
6	HCY	pct	0.957	0.027	0.017	0.966	0.020	0.015	0.947	0.031	0.023	0.953	0.024	0.024	0.952	0.024	0.025	0.951	0.029	0.021	0.959	0.017	0.025
6	HT3	hpdl	0.962	0.017	0.022	0.960	0.018	0.023	0.947	0.027	0.026	0.952	0.022	0.027	0.952	0.021	0.028	0.950	0.027	0.023	0.956	0.017	0.028
6	HT3	hpds	0.964	0.007	0.030	0.957	0.009	0.034	0.948	0.024	0.029	0.951	0.019	0.030	0.950	0.021	0.030	0.950	0.025	0.025	0.954	0.012	0.034
6	HT3	hpdv	0.964	0.003	0.034	0.956	0.004	0.041	0.948	0.022	0.031	0.951	0.019	0.031	0.952	0.019	0.029	0.953	0.024	0.024	0.951	0.010	0.040
6	HT3	pct	0.957	0.027	0.017	0.966	0.020	0.015	0.946	0.031	0.023	0.954	0.024	0.023	0.952	0.024	0.025	0.950	0.029	0.022	0.959	0.017	0.025
6	ICCU	hpdl	0.947	0.011	0.043	0.958	0.008	0.034	0.947	0.022	0.032	0.951	0.022	0.028	0.952	0.023	0.026	0.950	0.023	0.027	0.945	0.008	0.048
6	ICCU	hpds	0.941	0.004	0.055	0.949	0.003	0.049	0.947	0.020	0.034	0.953	0.019	0.029	0.952	0.021	0.028	0.951	0.022	0.028	0.936	0.005	0.059
6	ICCU	hpdv	0.941	0.001	0.059	0.941	0.001	0.059	0.948	0.017	0.036	0.954	0.016	0.031	0.952	0.021	0.028	0.952	0.021	0.027	0.933	0.004	0.064
6	ICCU	pct	0.952	0.015	0.034	0.961	0.011	0.029	0.947	0.025	0.029	0.953	0.023	0.024	0.951	0.026	0.023	0.952	0.025	0.024	0.948	0.010	0.043
6	IG1	hpds	0.856	0.001	0.143	0.879	0.003	0.118	0.947	0.020	0.033	0.948	0.023	0.028	0.947	0.021	0.032	0.948	0.024	0.028	0.910	0.003	0.087
6	IG1	hpdv	0.848	0.000	0.152	0.868	0.001	0.130	0.945	0.018	0.037	0.949	0.019	0.032	0.947	0.018	0.036	0.949	0.022	0.029	0.901	0.002	0.097
6	IG1	pct	0.903	0.004	0.093	0.917	0.005	0.078	0.951	0.023	0.026	0.949	0.028	0.023	0.951	0.023	0.026	0.950	0.025	0.025	0.927	0.005	0.068
6	IG2	hpds	0.872	0.002	0.126	0.887	0.003	0.110	0.946	0.020	0.034	0.948	0.023	0.028	0.948	0.021	0.031	0.948	0.026	0.027	0.914	0.003	0.083
6	IG2	hpdv	0.865	0.001	0.134	0.877	0.002	0.121	0.946	0.017	0.037	0.948	0.019	0.033	0.948	0.017	0.035	0.948	0.024	0.028	0.905	0.003	0.093
6	IG2	pct	0.914	0.005	0.081	0.923	0.006	0.071	0.952	0.023	0.026	0.950	0.028	0.022	0.950	0.024	0.025	0.950	0.026	0.024	0.931	0.005	0.064
6	IG3	hpds	0.873	0.001	0.125	0.888	0.003	0.109	0.948	0.020	0.033	0.948	0.024	0.028	0.948	0.021	0.031	0.950	0.025	0.026	0.913	0.004	0.083
6	IG3	hpdv	0.865	0.001	0.134	0.880	0.001	0.119	0.946	0.018	0.036	0.949	0.019	0.032	0.948	0.017	0.035	0.949	0.023	0.028	0.906	0.003	0.092
6	IG3	pct	0.914	0.005	0.081	0.924	0.006	0.070	0.953	0.023	0.025	0.949	0.028	0.023	0.951	0.024	0.025	0.951	0.026	0.024	0.931	0.005	0.064
6	IIN	hpds	0.997	0.002	0.001	0.981	0.004	0.015	0.952	0.024	0.024	0.951	0.028	0.021	0.952	0.025	0.023	0.950	0.028	0.023	0.954	0.006	0.040
6	IIN	hpdv	0.998	0.001	0.001	0.979	0.002	0.019	0.953	0.021	0.026	0.953	0.023	0.024	0.953	0.020	0.027	0.950	0.026	0.024	0.949	0.004	0.047
6	IIN	pct	0.992	0.008	0.000	0.983	0.009	0.008	0.954	0.027	0.018	0.952	0.032	0.016	0.953	0.028	0.019	0.952	0.029	0.020	0.962	0.008	0.029
6	JEFF	hpds	0.872	0.002	0.126	0.887	0.003	0.110	0.948	0.019	0.033	0.947	0.024	0.030	0.949	0.020	0.031	0.949	0.025	0.027	0.913	0.004	0.083
6	JEFF	hpdv	0.864	0.000	0.136	0.878	0.001	0.120	0.947	0.017	0.036	0.950	0.019	0.031	0.948	0.017	0.035	0.949	0.023	0.028	0.906	0.002	0.092
6	JEFF	pct	0.912	0.005	0.083	0.923	0.006	0.071	0.953	0.023	0.025	0.948	0.029	0.023	0.951	0.024	0.026	0.950	0.027	0.024	0.931	0.005	0.064
6	UNI	hpdl	0.960	0.018	0.022	0.962	0.015	0.023	0.947	0.028	0.026	0.951	0.022	0.028	0.950	0.022	0.028	0.950	0.027	0.024	0.956	0.016	0.029
6	UNI	hpds	0.963	0.007	0.031	0.957	0.008	0.035	0.946	0.025	0.030	0.950	0.021	0.030	0.952	0.020	0.029	0.950	0.026	0.025	0.954	0.013	0.034
6	UNI	hpdv	0.965	0.002	0.033	0.956	0.004	0.041	0.947	0.023	0.031	0.952	0.017	0.031	0.952	0.018	0.031	0.951	0.024	0.025	0.953	0.009	0.039
6	UNI	pct	0.957	0.027	0.016	0.966	0.020	0.015	0.945	0.032	0.024	0.954	0.023	0.023	0.952	0.024	0.025	0.950	0.029	0.021	0.959	0.017	0.025
24	FLAT	hpds	0.937	0.052	0.011	0.966	0.010	0.024	0.951	0.027	0.022	0.950	0.029	0.021	0.948	0.027	0.025	0.944	0.031	0.025	0.951	0.007	0.042
24	FLAT	hpdv	0.953	0.035	0.012	0.967	0.006	0.027	0.951	0.023	0.026	0.954	0.022	0.024	0.952	0.021	0.028	0.944	0.028	0.027	0.947	0.005	0.048
24	FLAT	pct	0.895	0.101	0.005	0.961	0.026	0.013	0.951	0.031	0.019	0.949	0.033	0.018	0.948	0.032	0.020	0.946	0.030	0.025	0.957	0.010	0.033
24	HCY	hpdl	0.962	0.018	0.020	0.960	0.014	0.027	0.965	0.018	0.018	0.947	0.028	0.025	0.946	0.025	0.029	0.944	0.024	0.032	0.956	0.013	0.032
24	HCY	hpds	0.968	0.004	0.028	0.960	0.005	0.036	0.966	0.014	0.020	0.950	0.024	0.027	0.950	0.020	0.031	0.945	0.022	0.034	0.951	0.011	0.039
24	HCY	hpdv	0.969	0.001	0.031	0.959	0.003	0.039	0.968	0.012	0.020	0.952	0.021	0.027	0.953	0.016	0.031	0.949	0.020	0.032	0.953	0.007	0.041
24	HCY	pct	0.959	0.026	0.016	0.963	0.019	0.019	0.965	0.021	0.015	0.947	0.031	0.022	0.945	0.029	0.026	0.945	0.027	0.029	0.957	0.015	0.029
24	HT3	hpdl	0.962	0.018	0.020	0.961	0.012	0.028	0.965	0.018	0.018	0.946	0.029	0.026	0.945	0.026	0.030	0.945	0.023	0.033	0.955	0.012	0.034
24	HT3	hpds	0.968	0.004	0.028	0.959	0.006	0.036	0.966	0.016	0.019	0.949	0.024	0.028	0.948	0.022	0.031	0.946	0.021	0.034	0.952	0.010	0.038
24	HT3	hpdv	0.968	0.001	0.032	0.958	0.002	0.041	0.968	0.012	0.021	0.952	0.021	0.028	0.952	0.017	0.032	0.948	0.020	0.033	0.952	0.008	0.041
24	HT3	pct	0.960	0.026	0.015	0.963	0.019	0.019	0.966	0.021	0.014	0.948	0.031	0.022	0.945	0.030	0.026	0.945	0.027	0.029	0.955	0.015	0.031
24	ICCU	hpdl	0.962	0.017	0.021	0.966	0.011	0.024	0.967	0.014	0.019	0.948	0.030	0.023	0.944	0.032	0.025	0.944	0.022	0.035	0.951	0.010	0.040
24	ICCU	hpds	0.972	0.003	0.026	0.961	0.005	0.035	0.966	0.012	0.023	0.949	0.025	0.026	0.948	0.026	0.027	0.943	0.019	0.038	0.950	0.006	0.044
24	ICCU	hpdv	0.970	0.000	0.030	0.959	0.003	0.039	0.966	0.011	0.024	0.951	0.023	0.027	0.953	0.021	0.027	0.945	0.019	0.037	0.946	0.004	0.051
24	ICCU	pct	0.957	0.027	0.017	0.966	0.017	0.017	0.965	0.018	0.017	0.947	0.034	0.020	0.942	0.037	0.022	0.947	0.024	0.030	0.952	0.011	0.038
24	IG1	hpds	0.869	0.001	0.130	0.893	0.003	0.105	0.946	0.018	0.036	0.952	0.020	0.029	0.948	0.021	0.032	0.945	0.025	0.030	0.912	0.003	0.085
24	IG1	hpdv	0.863	0.001	0.137	0.882	0.002	0.116	0.945	0.016	0.039	0.953	0.016	0.032	0.948	0.017	0.035	0.945	0.023	0.032	0.904	0.002	0.094
24	IG1	pct	0.914	0.005	0.081	0.924	0.006	0.070	0.951	0.021	0.029	0.953	0.024	0.023	0.950	0.025	0.026	0.946	0.025	0.028	0.929	0.004	0.066
24	IG2	hpds	0.885	0.002	0.114	0.900	0.003	0.097	0.945	0.019	0.036	0.952	0.020	0.029	0.947	0.021	0.032	0.943	0.026	0.032	0.915	0.003	0.082
24	IG2	hpdv	0.877	0.001	0.122	0.891	0.002	0.108	0.945	0.016	0.039	0.953	0.015	0.032	0.948	0.017	0.036	0.944	0.024	0.032	0.908	0.002	0.090
24	IG2	pct	0.924	0.005	0.071	0.929	0.006	0.065	0.950	0.022	0.028	0.954	0.024	0.022	0.950	0.025	0.025	0.946	0.026	0.028	0.933	0.005	0.063
24	IG3	hpds	0.884	0.001	0.115	0.899	0.003	0.098	0.947	0.018	0.035	0.952	0.020	0.028	0.948	0.021	0.031	0.944	0.025	0.031	0.915	0.003	0.082
24	IG3	hpdv	0.877	0.001	0.123	0.889	0.002	0.109	0.945	0.016	0.039	0.952	0.016	0.031	0.949	0.016	0.035	0.944	0.024	0.032	0.908	0.002	0.090
24	IG3	pct	0.924	0.006	0.071	0.929	0.006	0.065	0.950	0.021	0.028	0.953	0.024	0.023	0.951	0.024	0.025	0.947	0.026	0.027	0.933	0.004	0.063
24	IIN	hpds	0.997	0.003	0.000	0.982	0.005	0.014	0.952	0.022	0.026	0.955	0.024	0.022	0.952	0.025	0.023	0.944	0.029	0.027	0.952	0.005	0.043
24	IIN	hpdv	0.999	0.001	0.000	0.982	0.003	0.016	0.953	0.019	0.029	0.956	0.020	0.024	0.953	0.021	0.027	0.945	0.027	0.028	0.949	0.003	0.047
24	IIN	pct	0.990	0.010	0.000	0.984	0.010	0.006	0.954	0.026	0.020	0.952	0.030	0.018	0.951	0.031	0.018	0.946	0.030	0.024	0.963	0.006	0.031
24	JEFF	hpds	0.882	0.002	0.116	0.900	0.003	0.097	0.946	0.018	0.036	0.953	0.019	0.028	0.949	0.019	0.032	0.945	0.026	0.029	0.915	0.003	0.082
24	JEFF	hpdv	0.874	0.001	0.125	0.891	0.002	0.108	0.943	0.017	0.040	0.952	0.016	0.032	0.949	0.016	0.035	0.945	0.024	0.032	0.907	0.002	0.091
24	JEFF	pct	0.922	0.005	0.073	0.928	0.006	0.066	0.951	0.021	0.029	0.954	0.023	0.023	0.951	0.024	0.026	0.947	0.026	0.027	0.934	0.005	0.062
24	UNI	hpdl	0.963	0.019	0.019	0.961	0.013	0.027	0.965	0.018	0.018	0.947	0.028	0.026	0.946	0.026	0.029	0.945	0.024	0.032	0.953	0.014	0.034
24	UNI	hpds	0.969	0.005	0.027	0.959	0.006	0.036	0.969	0.014	0.018	0.949	0.024	0.028	0.950	0.020	0.031	0.944	0.022	0.035	0.951	0.011	0.039
24	UNI	hpdv	0.968	0.001	0.031	0.958	0.003	0.040	0.968	0.012	0.020	0.951	0.022	0.028	0.952	0.018	0.030	0.947	0.020	0.033	0.953	0.007	0.041
24	UNI	pct	0.959	0.026	0.016	0.963	0.019	0.019	0.966	0.021	0.014	0.947	0.031	0.022	0.946	0.029	0.026	0.946	0.026	0.029	0.955	0.015	0.030

TABLE F.5: **Coverage and proportion outside limits for total variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
			Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside		Cov.	Outside	
σ_b^2	Prior	Int.		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
0.5	FLAT	hpds	0.038	0.962	0.000	0.895	0.105	0.000	0.967	0.033	0.000	0.950	0.050	0.000	0.941	0.059	0.000	0.982	0.018	0.000	0.962	0.038	0.000
0.5	FLAT	hpdv	0.038	0.962	0.000	0.943	0.057	0.000	0.978	0.022	0.000	0.974	0.026	0.000	0.973	0.028	0.000	0.988	0.012	0.000	0.977	0.023	0.000
0.5	FLAT	pct	0.009	0.991	0.000	0.627	0.373	0.000	0.804	0.196	0.000	0.734	0.266	0.000	0.732	0.268	0.000	0.873	0.127	0.000	0.814	0.186	0.000
0.5	HCY	hpdl	0.897	0.104	0.000	0.940	0.060	0.000	0.984	0.016	0.001	0.965	0.033	0.002	0.959	0.038	0.004	0.977	0.017	0.006	0.971	0.028	0.001
0.5	HCY	hpds	0.969	0.031	0.000	0.968	0.032	0.000	0.993	0.007	0.001	0.984	0.014	0.003	0.980	0.017	0.004	0.983	0.011	0.006	0.981	0.019	0.001
0.5	HCY	hpdv	0.993	0.007	0.000	0.986	0.014	0.001	0.995	0.005	0.001	0.988	0.009	0.004	0.989	0.008	0.004	0.987	0.007	0.007	0.987	0.013	0.001
0.5	HCY	pct	0.806	0.194	0.000	0.894	0.107	0.000	0.926	0.074	0.001	0.925	0.075	0.001	0.938	0.061	0.002	0.932	0.065	0.004	0.935	0.065	0.001
0.5	HT3	hpdl	0.882	0.118	0.000	0.939	0.061	0.000	0.983	0.017	0.001	0.963	0.036	0.002	0.957	0.040	0.003	0.978	0.016	0.006	0.972	0.027	0.001
0.5	HT3	hpds	0.963	0.037	0.000	0.967	0.033	0.000	0.992	0.008	0.001	0.980	0.017	0.004	0.978	0.019	0.003	0.983	0.011	0.007	0.979	0.020	0.001
0.5	HT3	hpdv	0.988	0.012	0.000	0.988	0.013	0.000	0.996	0.004	0.001	0.988	0.009	0.004	0.989	0.008	0.004	0.987	0.007	0.007	0.987	0.013	0.001
0.5	HT3	pct	0.801	0.199	0.000	0.890	0.110	0.000	0.925	0.075	0.001	0.923	0.077	0.001	0.934	0.065	0.002	0.931	0.065	0.004	0.933	0.067	0.001
0.5	ICCU	hpdl	0.926	0.075	0.000	0.948	0.052	0.001	0.968	0.032	0.000	0.972	0.027	0.002	0.965	0.017	0.019	0.975	0.023	0.002	0.969	0.031	0.001
0.5	ICCU	hpds	0.972	0.027	0.001	0.968	0.030	0.003	0.983	0.017	0.000	0.984	0.015	0.002	0.973	0.008	0.020	0.979	0.019	0.003	0.976	0.024	0.001
0.5	ICCU	hpdv	0.990	0.010	0.001	0.984	0.012	0.004	0.991	0.009	0.000	0.990	0.009	0.002	0.975	0.003	0.023	0.982	0.016	0.003	0.983	0.016	0.001
0.5	ICCU	pct	0.887	0.113	0.000	0.915	0.085	0.000	0.909	0.091	0.000	0.946	0.054	0.001	0.966	0.027	0.008	0.923	0.077	0.001	0.937	0.063	0.000
0.5	IG1	hpds	0.992	0.008	0.000	0.981	0.017	0.003	0.984	0.011	0.005	0.981	0.010	0.009	0.978	0.012	0.011	0.980	0.010	0.011	0.979	0.017	0.005
0.5	IG1	hpdv	0.997	0.003	0.000	0.989	0.008	0.003	0.986	0.008	0.006	0.986	0.004	0.010	0.983	0.005	0.012	0.980	0.008	0.012	0.982	0.012	0.006
0.5	IG1	pct	0.917	0.083	0.000	0.941	0.059	0.000	0.955	0.043	0.003	0.951	0.045	0.004	0.953	0.044	0.003	0.958	0.037	0.006	0.952	0.045	0.002
0.5	IG2	hpds	0.989	0.011	0.000	0.978	0.020	0.002	0.984	0.012	0.004	0.982	0.011	0.007	0.978	0.014	0.008	0.981	0.010	0.009	0.976	0.019	0.004
0.5	IG2	hpdv	0.995	0.005	0.000	0.988	0.010	0.003	0.986	0.010	0.004	0.986	0.006	0.008	0.985	0.006	0.009	0.982	0.008	0.010	0.983	0.013	0.005
0.5	IG2	pct	0.894	0.106	0.000	0.930	0.070	0.000	0.951	0.047	0.002	0.945	0.052	0.003	0.946	0.051	0.002	0.956	0.039	0.005	0.947	0.051	0.002
0.5	IG3	hpds	0.990	0.010	0.000	0.979	0.020	0.002	0.982	0.014	0.004	0.983	0.011	0.006	0.979	0.013	0.008	0.981	0.011	0.009	0.976	0.019	0.004
0.5	IG3	hpdv	0.996	0.004	0.000	0.990	0.008	0.002	0.985	0.010	0.005	0.987	0.006	0.007	0.985	0.006	0.008	0.982	0.009	0.009	0.982	0.014	0.005
0.5	IG3	pct	0.892	0.108	0.000	0.929	0.071	0.000	0.950	0.048	0.002	0.945	0.052	0.003	0.947	0.051	0.002	0.955	0.040	0.005	0.947	0.051	0.002
0.5	IIN	hpds	0.984	0.016	0.000	0.971	0.029	0.000	0.981	0.016	0.003	0.982	0.012	0.006	0.979	0.014	0.008	0.982	0.011	0.008	0.973	0.025	0.003
0.5	IIN	hpdv	0.995	0.005	0.000	0.984	0.016	0.000	0.985	0.012	0.003	0.987	0.007	0.007	0.986	0.006	0.008	0.982	0.010	0.008	0.980	0.017	0.003
0.5	IIN	pct	0.820	0.180	0.000	0.896	0.104	0.000	0.944	0.055	0.001	0.941	0.056	0.003	0.946	0.052	0.002	0.953	0.042	0.004	0.935	0.064	0.001
0.5	JEFF	hpds	0.989	0.011	0.000	0.978	0.020	0.002	0.983	0.013	0.004	0.983	0.011	0.006	0.978	0.014	0.008	0.981	0.010	0.009	0.978	0.018	0.004
0.5	JEFF	hpdv	0.996	0.004	0.000	0.987	0.010	0.002	0.986	0.010	0.004	0.988	0.006	0.007	0.987	0.005	0.008	0.983	0.008	0.009	0.983	0.013	0.005
0.5	JEFF	pct	0.890	0.110	0.000	0.930	0.070	0.000	0.950	0.048	0.002	0.944	0.053	0.003	0.946	0.052	0.002	0.955	0.040	0.005	0.947	0.052	0.002
0.5	UNI	hpdl	0.771	0.229	0.000	0.936	0.064	0.000	0.985	0.015	0.001	0.950	0.049	0.002	0.926	0.073	0.002	0.977	0.018	0.006	0.973	0.027	0.001
0.5	UNI	hpds	0.885	0.115	0.000	0.967	0.033	0.000	0.992	0.008	0.001	0.975	0.023	0.002	0.953	0.045	0.003	0.983	0.011	0.006	0.981	0.018	0.001
0.5	UNI	hpdv	0.960	0.041	0.000	0.987	0.014	0.000	0.995	0.005	0.001	0.987	0.011	0.003	0.965	0.033	0.003	0.989	0.006	0.006	0.987	0.012	0.001
0.5	UNI	pct	0.792	0.209	0.000	0.887	0.113	0.000	0.924	0.076	0.001	0.916	0.084	0.001	0.926	0.073	0.001	0.930	0.067	0.004	0.933	0.067	0.001
6	FLAT	hpds	0.076	0.924	0.000	0.953	0.048	0.000	0.986	0.014	0.001	0.970	0.030	0.000	0.953	0.047	0.000	0.987	0.011	0.002	0.982	0.018	0.001
6	FLAT	hpdv	0.077	0.923	0.000	0.978	0.022	0.000	0.994	0.006	0.001	0.987	0.014	0.000	0.979	0.021	0.000	0.993	0.005	0.002	0.992	0.008	0.001
6	FLAT	pct	0.046	0.954	0.000	0.755	0.245	0.000	0.856	0.144	0.000	0.796	0.204	0.000	0.762	0.238	0.000	0.896	0.103	0.001	0.859	0.141	0.000
6	HCY	hpdl	0.949	0.051	0.000	0.963	0.036	0.002	0.966	0.019	0.016	0.964	0.031	0.006	0.963	0.033	0.005	0.973	0.013	0.015	0.961	0.024	0.016
6	HCY	hpds	0.988	0.012	0.000	0.984	0.015	0.002	0.975	0.009	0.017	0.982	0.012	0.007	0.982	0.014	0.005	0.979	0.005	0.016	0.972	0.010	0.018
6	HCY	hpdv	0.997	0.004	0.000	0.992	0.004	0.004	0.980	0.002	0.019	0.990	0.003	0.007	0.990	0.005	0.005	0.983	0.002	0.016	0.978	0.003	0.020
6	HCY	pct	0.882	0.119	0.000	0.931	0.070	0.000	0.939	0.054	0.008	0.934	0.065	0.002	0.940	0.059	0.002	0.948	0.044	0.009	0.935	0.057	0.009
6	HT3	hpdl	0.935	0.066	0.000	0.960	0.039	0.001	0.962	0.023	0.016	0.959	0.036	0.005	0.958	0.038	0.004	0.972	0.014	0.015	0.961	0.026	0.014
6	HT3	hpds	0.983	0.018	0.000	0.982	0.017	0.002	0.974	0.010	0.017	0.979	0.015	0.006	0.980	0.016	0.005	0.980	0.005	0.016	0.972	0.011	0.017
6	HT3	hpdv	0.996	0.005	0.000	0.991	0.007	0.003	0.981	0.003	0.017	0.989	0.005	0.007	0.989	0.007	0.005	0.983	0.002	0.016	0.977	0.004	0.020
6	HT3	pct	0.877	0.123	0.000	0.926	0.074	0.000	0.938	0.055	0.008	0.930	0.069	0.002	0.938	0.061	0.002	0.945	0.046	0.009	0.933	0.059	0.009
6	ICCU	hpdl	0.972	0.024	0.004	0.972	0.020	0.008	0.971	0.011	0.019	0.977	0.013	0.011	0.972	0.010	0.019	0.970	0.006	0.024	0.966	0.012	0.023
6	ICCU	hpds	0.987	0.008	0.006	0.981	0.008	0.012	0.974	0.005	0.022	0.984	0.005	0.012	0.972	0.005	0.024	0.973	0.003	0.025	0.968	0.007	0.026
6	ICCU	hpdv	0.993	0.002	0.006	0.982	0.003	0.015	0.977	0.002	0.022	0.987	0.002	0.012	0.974	0.001	0.026	0.974	0.001	0.026	0.970	0.002	0.029
6	ICCU	pct	0.956	0.043	0.002	0.966	0.031	0.004	0.963	0.029	0.009	0.967	0.029	0.005	0.975	0.016	0.010	0.970	0.019	0.011	0.964	0.026	0.011
6	IG1	hpds	0.996	0.004	0.000	0.978	0.011	0.012	0.954	0.004	0.041	0.977	0.006	0.017	0.977	0.010	0.014	0.946	0.003	0.051	0.962	0.006	0.032
6	IG1	hpdv	0.999	0.001	0.000	0.982	0.005	0.013	0.954	0.003	0.043	0.980	0.003	0.017	0.982	0.003	0.015	0.947	0.001	0.052	0.961	0.003	0.035
6	IG1	pct	0.953	0.047	0.000	0.953	0.042	0.004	0.950	0.028	0.022	0.957	0.035	0.008	0.955	0.040	0.005	0.950	0.024	0.026	0.952	0.031	0.017
6	IG2	hpds	0.996	0.004	0.000	0.980	0.012	0.008	0.960	0.004	0.036	0.980	0.006	0.014	0.979	0.011	0.010	0.952	0.004	0.044	0.966	0.007	0.027
6	IG2	hpdv	0.998	0.002	0.000	0.985	0.006	0.009	0.961	0.003	0.036	0.982	0.003	0.015	0.985	0.005	0.011	0.953	0.002	0.045	0.967	0.004	0.029
6	IG2	pct	0.942	0.059	0.000	0.949	0.048	0.003	0.951	0.031	0.018	0.955	0.040	0.006	0.950	0.046	0.004	0.953	0.026	0.022	0.952	0.035	0.013
6	IG3	hpds	0.995	0.005	0.000	0.980	0.012	0.008	0.960	0.005	0.035	0.980	0.007	0.014	0.979	0.011	0.009	0.952	0.005	0.044	0.967	0.007	0.026
6	IG3	hpdv	0.998	0.002	0.000	0.984	0.006	0.010	0.961	0.003	0.036	0.983	0.003	0.014	0.985	0.005	0.010	0.954	0.002	0.044	0.969	0.003	0.028
6	IG3	pct	0.940	0.060	0.000	0.949	0.049	0.002	0.951	0.032	0.017	0.954	0.040	0.005	0.951	0.046	0.004	0.952	0.027	0.021	0.952	0.035	0.013
6	IIN	hpds	0.993	0.007	0.000	0.981	0.015	0.004	0.961	0.005	0.034	0.980	0.007	0.013	0.979	0.012	0.009	0.953	0.004	0.043	0.968	0.008	0.024
6	IIN	hpdv	0.998	0.002	0.000	0.989	0.007	0.004	0.962	0.003	0.035	0.983	0.003	0.014	0.986	0.004	0.009	0.954	0.002	0.044	0.970	0.005	0.025
6	IIN	pct	0.918	0.082	0.000	0.939	0.061	0.001	0.950	0.033	0.017	0.954	0.041	0.005	0.950	0.047	0.004	0.950	0.028	0.022	0.950	0.039	0.011
6	JEFF	hpds	0.997	0.004	0.000	0.980	0.012	0.008	0.960	0.005	0.035	0.978	0.009	0.013	0.980	0.011	0.009	0.952	0.005	0.044	0.966	0.008	0.027
6	JEFF	hpdv	0.998	0.002	0.000	0.985	0.006	0.009	0.961	0.002	0.036	0.983	0.003	0.014	0.986	0.004	0.010	0.954	0.001	0.045	0.968	0.004	0.029
6	JEFF	pct	0.938	0.062	0.000	0.949	0.049	0.003	0.951	0.032	0.017	0.955	0.040	0.005	0.950	0.047	0.004	0.952	0.027	0.021	0.952	0.036	0.013
6	UNI	hpdl	0.862	0.138	0.000	0.913	0.087	0.001	0.933	0.053	0.014	0.917	0.080	0.004	0.932	0.066	0.003	0.945	0.042	0.014	0.929	0.058	0.013
6	UNI	hpds	0.928	0.072	0.000	0.956	0.043	0.002	0.957	0.028	0.016	0.947	0.050	0.004	0.950	0.047	0.004	0.969	0.018	0.014	0.954	0.032	0.015
6	UNI	hpdv	0.967	0.033	0.000	0.977	0.021	0.003	0.971	0.014	0.016	0.969	0.027	0.005	0.967	0.030	0.004	0.978	0.009	0.014	0.970	0.014	0.017
6	UNI	pct	0.875	0.125	0.000	0.920	0.081	0.000	0.934	0.060	0.007	0.924	0.076	0.001	0.939	0.061	0.001	0.943	0.049	0.009	0.930	0.063	0.008
24	FLAT	hpds	0.171	0.829	0.000	0.974	0.027	0.000	0.987	0.012	0.002	0.981	0.019	0.000	0.972	0.028	0.000	0.986	0.012	0.003	0.984	0.014	0.003
24	FLAT	hpdv	0.171	0.829	0.000	0.991	0.009	0.000	0.995	0.003	0.002	0.993	0.007	0.000	0.987	0.013	0.000	0.993	0.004	0.003	0.993	0.005	0.003
24	FLAT	pct	0.181	0.819	0.000	0.844	0.156	0.000	0.890	0.109	0.001	0.845	0.155	0.000	0.807	0.193	0.000	0.892	0.108	0.001	0.890	0.109	0.001
24	HCY	hpdl	0.979	0.022	0.000	0.965	0.017	0.019	0.951	0.017	0.032	0.969	0.021	0.011	0.972	0.022	0.007	0.961	0.014	0.026	0.960	0.018	0.023
24	HCY	hpds	0.997	0.003	0.000	0.972	0.005	0.024	0.959	0.007	0.034	0.980	0.008	0.013	0.979	0.013	0.009	0.971	0.003	0.027	0.969	0.007	0.025
24	HCY	hpdv	1.000	0.001	0.000	0.973	0.002	0.026	0.965	0.001	0.035	0.986	0.002	0.013	0.986	0.005	0.010	0.971	0.000	0.029	0.973	0.001	0.026
24	HCY	pct	0.956	0.045	0.000	0.953	0.037	0.011	0.948	0.040	0.013	0.956	0.039	0.005	0.960	0.039	0.002	0.958	0.029	0.014	0.955	0.033	0.013
24	HT3	hpdl	0.972	0.028	0.000	0.962	0.020	0.018	0.950	0.020	0.031	0.967	0.024	0.010	0.972	0.025	0.004	0.962	0.015	0.024	0.958	0.020	0.023
24	HT3	hpds	0.996	0.005	0.000	0.976	0.004	0.021	0.961	0.008	0.032	0.979	0.010	0.011	0.979	0.014	0.008	0.970	0.005	0.026	0.970	0.007	0.024
24	HT3	hpdv	0.999	0.001	0.000	0.977	0.002	0.022	0.966	0.002	0.033	0.987	0.002	0.012	0.986	0.007	0.008	0.972	0.001	0.028	0.974	0.001	0.025
24	HT3	pct	0.953	0.047	0.000	0.953	0.037	0.010	0.948	0.041	0.012	0.955	0.040	0.005	0.960	0.039	0.002	0.957	0.031	0.013	0.954	0.034	0.013
24	ICCU	hpdl	0.962	0.005	0.034	0.931	0.004	0.065	0.907	0.008	0.086	0.933	0.006	0.061	0.947	0.009	0.045	0.923	0.003	0.074	0.921	0.007	0.073
24	ICCU	hpds	0.956	0.002	0.043	0.919	0.002	0.080	0.907	0.002	0.092	0.930	0.003	0.068	0.940	0.005	0.055	0.922	0.001	0.078	0.916	0.002	0.083
24	ICCU	hpdv	0.956	0.001	0.044	0.917	0.001	0.083	0.907	0.000	0.094	0.927	0.002	0.072	0.937	0.000	0.064	0.920	0.000	0.081	0.913	0.001	0.087
24	ICCU	pct	0.968	0.015	0.018	0.941	0.013	0.047	0.938	0.012	0.051	0.960	0.013	0.028	0.962	0.014	0.025	0.939	0.010	0.052	0.942	0.014	0.045
24	IG1	hpds	0.993	0.002	0.005	0.956	0.007	0.038	0.946	0.003	0.051	0.958	0.005	0.037	0.977	0.007	0.016	0.943	0.004	0.053	0.946	0.006	0.049
24	IG1	hpdv	0.994	0.001	0.005	0.957	0.003	0.041	0.946	0.001	0.052	0.960	0.002	0.038	0.980	0.003	0.017	0.945	0.002	0.054	0.947	0.002	0.051
24	IG1	pct	0.970	0.029	0.001	0.953	0.029	0.018	0.949	0.023	0.028	0.954	0.027	0.019	0.962	0.032	0.006	0.952	0.022	0.027	0.946	0.026	0.028
24	IG2	hpds	0.996	0.003	0.002	0.962	0.007	0.032	0.952	0.004	0.045	0.964	0.005	0.031	0.980	0.007	0.013	0.950	0.003	0.047	0.952	0.006	0.043
24	IG2	hpdv	0.997	0.001	0.002	0.964	0.003	0.033	0.953	0.002	0.046	0.966	0.002	0.032	0.983	0.004	0.014	0.951	0.002	0.048	0.955	0.002	0.044
24	IG2	pct	0.963	0.037	0.000	0.952	0.033	0.015	0.950	0.027	0.023	0.956	0.031	0.014	0.959	0.037	0.005	0.954	0.024	0.022	0.947	0.029	0.024
24	IG3	hpds	0.997	0.002	0.001	0.961	0.008	0.031	0.953	0.004	0.044	0.965	0.005	0.030	0.980	0.008	0.012	0.952	0.004	0.045	0.952	0.006	0.042
24	IG3	hpdv	0.998	0.001	0.001	0.964	0.003	0.032	0.954	0.001	0.045	0.967	0.002	0.031	0.984	0.003	0.013	0.953	0.001	0.046	0.954	0.002	0.044
24	IG3	pct	0.964	0.036	0.000	0.952	0.034	0.015	0.950	0.027	0.023	0.956	0.031	0.014	0.959	0.037	0.004	0.954	0.024	0.022	0.948	0.029	0.023
24	IIN	hpds	0.998	0.002	0.000	0.964	0.008	0.028	0.953	0.004	0.044	0.965	0.006	0.029	0.979	0.009	0.012	0.952	0.003	0.045	0.954	0.006	0.040
24	IIN	hpdv	0.999	0.001	0.000	0.967	0.004	0.030	0.953	0.002	0.045	0.967	0.003	0.031	0.983	0.004	0.013	0.953	0.001	0.046	0.956	0.002	0.042
24	IIN	pct	0.959	0.042	0.000	0.951	0.036	0.013	0.951	0.027	0.022	0.955	0.032	0.014	0.958	0.037	0.005	0.953	0.025	0.022	0.947	0.031	0.022
24	JEFF	hpds	0.997	0.002	0.001	0.963	0.007	0.030	0.953	0.004	0.043	0.965	0.005	0.029	0.980	0.008	0.012	0.951	0.004	0.045	0.953	0.006	0.042
24	JEFF	hpdv	0.998	0.001	0.002	0.965	0.003	0.032	0.954	0.001	0.045	0.967	0.003	0.031	0.984	0.003	0.013	0.952	0.001	0.046	0.955	0.002	0.043
24	JEFF	pct	0.963	0.037	0.000	0.952	0.034	0.014	0.951	0.027	0.022	0.955	0.031	0.014	0.958	0.038	0.005	0.955	0.024	0.021	0.948	0.030	0.023
24	UNI	hpdl	0.970	0.030	0.000	0.949	0.036	0.015	0.929	0.051	0.021	0.959	0.034	0.008	0.969	0.029	0.003	0.936	0.045	0.019	0.938	0.044	0.019
24	UNI	hpds	0.981	0.019	0.000	0.955	0.028	0.018	0.929	0.048	0.023	0.967	0.025	0.009	0.971	0.025	0.004	0.937	0.044	0.020	0.939	0.042	0.020
24	UNI	hpdv	0.993	0.007	0.000	0.964	0.018	0.019	0.932	0.042	0.027	0.973	0.019	0.009	0.976	0.020	0.005	0.939	0.042	0.020	0.943	0.037	0.021
24	UNI	pct	0.973	0.027	0.000	0.967	0.026	0.008	0.965	0.027	0.008	0.968	0.029	0.004	0.975	0.024	0.002	0.973	0.018	0.010	0.967	0.024	0.010

Appendix G

Credible Intervals

Contents

G.1 R Code for Plotting CrIs on the icc Scale

The following code is used to create Figure 8.1d) which uses an icc scale.

```
library(ggplot2)
library(scales)

# Function to create icc scale
varcomps_icc = function(scale.val) trans_new("varcomps",
function(x,...) {
        sapply(x,
        function(z,scale.val) {
                if (is.na(z)) {z <- z} else
                {z <- z/(z+scale.val)}
        },scale.val=scale.val)
}
,
function(x,...) {
        sapply(x,
        function(z,scale.val) {
                if (is.na(z)) {z <- z} else
                {z <- z*scale.val/(1-z)}
        },scale.val=scale.val)
}
)

#Put data in here. Need variables with Median, Lower, Upper (for median and
     CrI limits) and grouping variable prior
plotdata.example<- XXXXXX

#Plotting median credible intervals for keg variance using icc scale -
    normalised to 6
breaks.example=c(0.00001,0.001,0.1,0.5,1,2,4,6,8,10,20,50,100)
ci.k <- ggplot(data=plotdata.example)
ci.k <- ci.k+ geom_linerange(aes(x=prior, ymin=Lower, ymax=Upper,colour=
    prior), size=1.5) # High/Low bar
ci.k <- ci.k+ guides(colour=FALSE)
ci.k <- ci.k+ geom_point(aes(x=prior, y=Median,colour=prior), size=10,
    shape="--")
ci.k <- ci.k+ geom_hline(aes(yintercept=trueval),linetype="dashed")
ci.k <- ci.k+ theme(axis.text.x = element_text(angle=45, hjust=1, vjust=1))
ci.k <- ci.k+ scale_x_discrete(limits=c("UNI","HCY","HT3","IIN","IG2","JEFF
    "))
ci.k <- ci.k+ ylab("Median over simulated datasets") + xlab("Prior")
ci.k.icc6 <- ci.k+ scale_y_continuous(trans=varcomps_icc(scale.val=6),
    breaks=breaks.example,labels=prettyNum)
ci.k.icc6 <- ci.k.icc6+ ggtitle("icc scale (normalised to 6)")

ci.k.icc6
```

G.2 Median of CrI Limits and Posterior Median Results

The median lower and upper limits of the hpd intervals of all the simulated datasets (together with the posterior median) are presented in Tables G.1 to G.4. These are provided for $\sigma_b^2 = 0.5$, 6, 24 for all scenarios and for the priors investigated in Section 7.2.2.

TABLE G.1: **Median CrI limits and posterior median for batch variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and selected priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
	σ_b^2	Prior	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU
SD	0.5	FLAT	3.86	130	1654	0.076	1.95	6.41	0.065	1.19	3.66	0.107	2.46	7.99	0.192	4.60	15.0	0.207	1.03	2.96	0.064	1.24	3.82
SD	0.5	HCY	0.002	1.80	8.65	0.002	0.948	3.11	0.002	0.752	2.08	0.003	1.30	3.97	0.004	2.18	6.53	0.069	0.781	1.88	0.002	0.739	2.16
SD	0.5	HT3	0.006	1.86	8.63	0.002	0.961	3.16	0.002	0.755	2.10	0.006	1.32	4.06	0.010	2.22	6.68	0.061	0.781	1.88	0.002	0.739	2.18
SD	0.5	IG2	0.046	1.96	8.96	0.073	1.22	3.19	0.072	0.786	1.89	0.112	1.58	3.98	0.198	2.93	7.44	0.188	0.709	1.60	0.058	0.810	2.02
SD	0.5	IIN	0.047	2.06	9.56	0.072	1.23	3.22	0.072	0.787	1.90	0.111	1.58	4.00	0.194	2.94	7.47	0.188	0.709	1.60	0.058	0.812	2.03
SD	0.5	JEFF	0.046	1.98	9.15	0.073	1.22	3.20	0.072	0.788	1.90	0.111	1.58	4.00	0.196	2.94	7.48	0.189	0.710	1.60	0.058	0.814	2.03
SD	0.5	UNI	7.6E-5	1.92	6.99	6.1E-5	0.981	3.30	9.5E-5	0.759	2.13	7.7E-5	1.36	4.28	1.5E-4	2.34	6.56	0.069	0.785	1.92	9.0E-5	0.748	2.23
SD	6	FLAT	4.53	140	2121	0.258	3.44	10.2	0.840	3.35	9.31	0.262	3.76	11.3	0.259	5.32	16.9	1.26	3.39	9.12	1.09	3.43	9.32
SD	6	HCY	0.004	2.72	10.9	0.008	2.36	5.60	0.847	2.62	5.74	0.007	2.34	5.86	0.006	2.65	7.41	1.24	2.74	5.61	1.04	2.67	5.68
SD	6	HT3	0.010	2.80	10.4	0.016	2.39	5.70	0.853	2.64	5.84	0.013	2.38	5.98	0.013	2.72	7.54	1.25	2.75	5.69	1.05	2.70	5.77
SD	6	IG2	0.088	2.82	12.0	0.359	2.38	5.46	0.975	2.43	5.12	0.355	2.58	5.93	0.275	3.46	8.51	1.18	2.46	4.95	1.02	2.45	5.11
SD	6	IIN	0.083	2.86	12.5	0.327	2.36	5.46	0.971	2.43	5.13	0.351	2.59	5.96	0.270	3.47	8.53	1.18	2.46	4.97	1.01	2.46	5.12
SD	6	JEFF	0.087	2.85	12.4	0.355	2.38	5.49	0.971	2.43	5.12	0.350	2.59	5.96	0.274	3.48	8.56	1.18	2.47	4.97	1.01	2.46	5.13
SD	6	UNI	2.7E-4	2.81	7.53	4.0E-4	2.46	5.86	0.899	2.70	6.04	2.8E-4	2.46	6.09	2.8E-4	2.83	7.07	1.27	2.82	5.88	1.08	2.76	5.95
SD	24	FLAT	6.30	185	3038	1.44	6.60	18.4	2.32	6.68	17.8	1.20	6.76	19.1	0.533	7.51	22.5	2.60	6.73	18.1	2.52	6.76	18.0
SD	24	HCY	0.010	4.72	15.3	1.41	4.81	10.2	2.39	5.10	10.1	0.263	4.80	10.1	0.012	4.30	10.3	2.53	5.15	9.96	2.42	5.10	9.96
SD	24	HT3	0.024	4.76	13.9	1.47	4.85	10.2	2.42	5.15	10.1	0.371	4.85	10.2	0.024	4.40	10.2	2.55	5.20	10.0	2.45	5.13	9.95
SD	24	IG2	0.304	4.90	19.2	1.76	4.81	10.2	2.32	4.89	9.80	1.67	4.86	10.4	0.739	5.18	11.8	2.42	4.92	9.80	2.37	4.92	9.86
SD	24	IIN	0.262	4.91	19.4	1.71	4.80	10.2	2.33	4.90	9.84	1.67	4.87	10.5	0.735	5.18	11.9	2.43	4.92	9.86	2.37	4.93	9.91
SD	24	JEFF	0.292	4.91	19.9	1.76	4.82	10.2	2.33	4.89	9.85	1.67	4.86	10.5	0.735	5.18	11.9	2.42	4.93	9.84	2.38	4.93	9.90
SD	24	UNI	0.909	4.46	8.66	2.31	4.88	8.57	2.89	5.16	8.33	1.88	4.87	8.65	0.611	4.48	8.50	2.98	5.20	8.30	2.90	5.15	8.29
Var	0.5	FLAT	14.9	16917	2736866	0.001	3.80	40.9	7.5E-4	1.41	13.2	0.003	6.07	63.4	0.009	21.2	224	0.004	1.05	8.42	8.2E-4	1.55	14.5
Var	0.5	HCY	<1E-15	3.24	74.9	2.4E-7	0.900	9.67	1.0E-7	0.565	4.32	2.1E-6	1.69	15.8	5.1E-7	4.75	42.7	2.7E-6	0.609	3.37	1.5E-7	0.547	4.68
Var	0.5	HT3	<1E-15	3.45	74.5	9.3E-7	0.923	9.96	2.4E-7	0.570	4.40	1.3E-5	1.74	16.5	3.5E-6	4.94	44.7	2.4E-5	0.610	3.39	2.4E-7	0.546	4.74
Var	0.5	IG2	9.8E-4	3.83	80.2	6.2E-4	1.49	10.0	3.8E-4	0.617	3.46	0.001	2.49	15.5	0.004	8.59	54.1	0.002	0.503	2.38	3.3E-4	0.657	3.99
Var	0.5	IIN	0.001	4.23	91.3	6.0E-4	1.51	10.2	3.7E-4	0.620	3.49	0.001	2.50	15.7	0.004	8.65	54.6	0.002	0.503	2.39	3.4E-4	0.660	4.06
Var	0.5	JEFF	0.001	3.93	83.5	6.0E-4	1.50	10.1	3.9E-4	0.622	3.50	0.001	2.50	15.7	0.004	8.64	54.7	0.002	0.503	2.39	3.4E-4	0.663	4.04
Var	0.5	UNI	1.4E-9	3.69	48.9	5.9E-10	0.962	10.9	4.8E-11	0.576	4.55	4.0E-9	1.85	18.3	8.4E-9	5.46	43.0	7.1E-9	0.616	3.50	1.5E-11	0.559	4.95
Var	6	FLAT	20.5	19654	4497784	0.008	11.8	103	0.060	11.2	82.9	0.010	14.1	125	0.014	28.3	282	1.34	11.5	82.0	0.844	11.8	85.5
Var	6	HCY	3.7E-7	7.40	118	5.4E-6	5.57	31.2	1.0E-4	6.86	30.3	3.4E-6	5.48	34.1	6.5E-7	7.03	54.9	1.20	7.48	30.7	0.622	7.15	31.0
Var	6	HT3	8.6E-6	7.83	109	3.1E-5	5.72	32.3	0.002	6.97	31.3	2.0E-5	5.65	35.6	7.0E-6	7.39	56.8	1.21	7.59	31.5	0.623	7.27	32.0
Var	6	IG2	0.003	7.93	144	0.006	5.66	28.3	0.306	5.89	24.5	0.007	6.66	33.6	0.007	12.0	70.5	1.19	6.06	24.0	0.708	6.02	25.1
Var	6	IIN	0.003	8.21	156	0.005	5.56	28.4	0.286	5.92	24.7	0.006	6.68	33.9	0.007	12.1	70.8	1.19	6.08	24.2	0.720	6.05	25.4
Var	6	JEFF	0.003	8.10	152	0.005	5.69	28.6	0.292	5.92	24.6	0.007	6.70	33.9	0.007	12.1	71.2	1.19	6.09	24.3	0.720	6.06	25.4
Var	6	UNI	1.5E-8	7.92	56.7	2.1E-8	6.06	34.3	5.6E-6	7.31	33.1	1.8E-8	6.04	36.9	1.8E-8	8.01	50.0	1.22	7.94	33.6	0.656	7.62	34.0
Var	24	FLAT	39.7	34349	9226843	0.148	43.6	324	3.35	44.7	311	0.114	45.7	349	0.040	56.4	498	5.73	45.3	323	5.21	45.6	318
Var	24	HCY	9.3E-7	22.3	235	3.7E-4	23.1	93.8	4.12	26.0	96.7	1.2E-4	23.1	97.7	2.2E-5	18.5	104	5.14	26.6	96.5	4.55	26.0	96.1
Var	24	HT3	9.7E-5	22.6	192	0.004	23.6	93.6	4.25	26.5	96.6	8.6E-4	23.6	96.9	1.7E-4	19.4	104	5.23	27.0	96.7	4.64	26.4	95.5
Var	24	IG2	0.021	24.0	364	0.452	23.1	96.2	4.39	23.9	93.7	0.242	23.6	101	0.029	26.8	133	5.17	24.2	94.2	4.65	24.2	94.9
Var	24	IIN	0.017	24.1	375	0.341	23.0	96.1	4.38	24.0	94.5	0.236	23.7	102	0.028	26.8	135	5.13	24.2	95.4	4.69	24.3	96.2
Var	24	JEFF	0.022	24.2	392	0.417	23.2	96.8	4.36	23.9	94.4	0.236	23.7	102	0.028	26.9	135	5.15	24.3	94.9	4.66	24.3	95.6
Var	24	UNI	4.0E-7	19.9	65.4	8.8E-5	23.8	63.7	6.27	26.6	65.4	6.7E-6	23.7	63.9	3.4E-7	20.1	63.7	6.94	27.1	65.2	6.30	26.5	65.0

TABLE G.2: **Median CrI limits and posterior median for keg variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and selected priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
	σ_b^2	Prior	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU
SD	0.5	FLAT	0.441	4.95	26.1	0.134	1.55	3.58	0.168	0.882	1.87	1.37	2.68	4.84	2.85	5.27	9.36	0.442	0.747	1.11	0.175	1.01	1.85
SD	0.5	HCY	0.003	1.30	4.39	0.003	0.908	2.44	0.004	0.679	1.47	1.29	2.41	4.04	2.67	4.66	7.52	0.390	0.716	1.08	0.005	0.727	1.59
SD	0.5	HT3	0.006	1.32	4.49	0.003	0.915	2.46	0.005	0.682	1.47	1.29	2.42	4.06	2.69	4.68	7.58	0.389	0.716	1.08	0.007	0.727	1.59
SD	0.5	IG2	0.082	1.34	3.48	0.110	1.09	2.36	0.103	0.635	1.27	1.21	2.12	3.41	2.55	4.21	6.61	0.400	0.700	1.04	0.164	0.938	1.71
SD	0.5	IIN	0.074	1.38	3.67	0.100	1.08	2.37	0.100	0.636	1.28	1.21	2.12	3.42	2.55	4.22	6.63	0.398	0.699	1.04	0.148	0.905	1.67
SD	0.5	JEFF	0.081	1.35	3.54	0.110	1.10	2.37	0.103	0.638	1.28	1.21	2.13	3.42	2.55	4.22	6.64	0.400	0.700	1.04	0.164	0.940	1.71
SD	0.5	UNI	5.4E-5	1.36	4.69	1.3E-4	0.924	2.50	3.3E-4	0.679	1.48	1.29	2.44	4.14	2.80	4.77	7.53	0.388	0.716	1.08	3.3E-4	0.732	1.59
SD	6	FLAT	0.449	4.82	25.9	0.151	1.80	4.26	0.171	1.01	2.37	1.39	2.87	5.38	2.86	5.40	9.68	0.447	0.754	1.13	0.179	1.04	1.90
SD	6	HCY	0.003	1.48	4.99	0.004	1.14	3.12	0.003	0.755	1.83	1.32	2.62	4.65	2.76	4.88	7.96	0.375	0.703	1.07	0.004	0.731	1.62
SD	6	HT3	0.007	1.50	5.10	0.007	1.16	3.15	0.007	0.758	1.84	1.31	2.63	4.68	2.77	4.90	8.01	0.373	0.703	1.07	0.007	0.732	1.62
SD	6	IG2	0.090	1.50	3.97	0.132	1.30	2.86	0.104	0.732	1.58	1.24	2.28	3.83	2.58	4.33	6.90	0.408	0.712	1.07	0.176	0.988	1.80
SD	6	IIN	0.083	1.54	4.15	0.118	1.29	2.88	0.102	0.733	1.59	1.24	2.28	3.84	2.58	4.34	6.92	0.408	0.711	1.07	0.157	0.955	1.77
SD	6	JEFF	0.087	1.51	4.03	0.131	1.31	2.87	0.105	0.733	1.59	1.24	2.28	3.84	2.58	4.34	6.92	0.409	0.712	1.07	0.175	0.989	1.80
SD	6	UNI	7.5E-5	1.56	5.28	1.5E-4	1.17	3.21	2.8E-4	0.756	1.85	1.32	2.65	4.79	2.93	4.98	7.79	0.374	0.704	1.07	3.2E-4	0.734	1.62
SD	24	FLAT	0.374	4.96	27.1	0.140	1.94	4.95	0.157	1.02	2.50	1.39	3.05	6.12	2.90	5.73	10.6	0.445	0.752	1.12	0.185	1.05	1.92
SD	24	HCY	0.003	1.64	5.84	0.003	1.17	3.46	0.004	0.747	1.83	1.28	2.74	5.26	2.77	5.15	8.78	0.378	0.709	1.07	0.005	0.762	1.64
SD	24	HT3	0.007	1.67	5.97	0.007	1.18	3.49	0.007	0.748	1.83	1.28	2.75	5.31	2.79	5.17	8.81	0.380	0.709	1.07	0.007	0.759	1.65
SD	24	IG2	0.092	1.65	4.58	0.124	1.39	3.25	0.098	0.734	1.63	1.24	2.41	4.31	2.63	4.59	7.56	0.406	0.709	1.06	0.178	1.00	1.82
SD	24	IIN	0.084	1.69	4.79	0.112	1.38	3.28	0.094	0.736	1.63	1.24	2.42	4.32	2.63	4.60	7.59	0.404	0.709	1.06	0.159	0.968	1.79
SD	24	JEFF	0.089	1.66	4.66	0.125	1.40	3.26	0.095	0.737	1.63	1.24	2.41	4.32	2.63	4.60	7.57	0.406	0.710	1.06	0.179	1.00	1.82
SD	24	UNI	9.1E-5	1.77	6.06	1.0E-4	1.20	3.61	2.5E-4	0.752	1.85	1.29	2.79	5.48	3.09	5.25	8.17	0.381	0.709	1.07	3.2E-4	0.769	1.65
Var	0.5	FLAT	0.134	24.5	678	0.001	2.39	12.4	0.001	0.778	3.29	1.56	7.16	22.8	6.94	27.7	85.3	0.141	0.558	1.15	7.5E-4	1.03	3.18
Var	0.5	HCY	2.3E-6	1.69	19.2	5.9E-7	0.824	5.95	3.1E-6	0.461	2.13	1.35	5.81	15.7	6.10	21.7	54.6	0.109	0.512	1.09	6.2E-6	0.528	2.51
Var	0.5	HT3	1.0E-5	1.75	20.2	1.6E-6	0.836	6.07	9.1E-6	0.465	2.14	1.34	5.85	15.9	6.13	21.9	55.5	0.108	0.513	1.09	2.0E-5	0.529	2.51
Var	0.5	IG2	7.3E-4	1.81	11.9	6.0E-4	1.20	5.36	3.8E-4	0.404	1.53	1.25	4.50	11.3	5.79	17.7	42.5	0.117	0.489	1.01	6.3E-4	0.881	2.69
Var	0.5	IIN	7.2E-4	1.89	13.3	5.2E-4	1.16	5.43	3.6E-4	0.404	1.53	1.26	4.50	11.3	5.80	17.8	42.8	0.115	0.488	1.01	5.5E-4	0.819	2.61
Var	0.5	JEFF	7.6E-4	1.83	12.3	5.9E-4	1.20	5.38	3.7E-4	0.407	1.53	1.26	4.52	11.3	5.79	17.8	42.8	0.117	0.490	1.01	6.3E-4	0.884	2.70
Var	0.5	UNI	3.1E-9	1.86	22.0	2.8E-10	0.853	6.26	8.8E-9	0.461	2.16	1.34	5.95	16.5	6.61	22.7	54.5	0.109	0.512	1.10	2.2E-8	0.536	2.53
Var	6	FLAT	0.131	23.2	669	0.002	3.25	17.6	0.002	1.01	5.33	1.57	8.23	28.1	7.02	29.1	91.6	0.145	0.569	1.18	8.3E-4	1.08	3.36
Var	6	HCY	1.6E-6	2.19	24.9	2.8E-6	1.30	9.72	9.7E-7	0.570	3.33	1.33	6.87	20.8	6.37	23.8	61.3	0.097	0.494	1.07	5.7E-6	0.535	2.61
Var	6	HT3	1.0E-5	2.25	26.0	1.6E-5	1.34	9.91	8.4E-6	0.574	3.37	1.34	6.92	21.1	6.43	24.0	61.9	0.097	0.494	1.07	2.4E-5	0.536	2.61
Var	6	IG2	9.4E-4	2.26	15.5	8.9E-4	1.70	7.86	5.0E-4	0.535	2.38	1.27	5.20	14.1	5.86	18.8	46.3	0.119	0.506	1.06	7.5E-4	0.977	2.99
Var	6	IIN	8.8E-4	2.36	17.0	7.8E-4	1.65	8.00	4.7E-4	0.537	2.40	1.27	5.22	14.2	5.88	18.8	46.6	0.118	0.506	1.06	6.3E-4	0.913	2.91
Var	6	JEFF	9.8E-4	2.28	16.0	9.2E-4	1.71	7.92	5.1E-4	0.537	2.40	1.27	5.21	14.2	5.86	18.9	46.6	0.12	0.507	1.06	7.3E-4	0.977	3.00
Var	6	UNI	4.6E-9	2.42	27.8	4.8E-9	1.37	10.3	1.8E-9	0.572	3.41	1.34	7.05	22.1	7.16	24.8	58.3	0.096	0.496	1.07	1.7E-8	0.539	2.62
Var	24	FLAT	0.080	24.6	732	0.002	3.76	23.9	0.002	1.04	6.01	1.58	9.32	36.5	7.15	32.8	109	0.141	0.565	1.18	8.4E-4	1.10	3.42
Var	24	HCY	3.2E-7	2.68	34.0	3.2E-6	1.36	11.9	7.5E-7	0.558	3.34	1.25	7.49	26.7	6.29	26.5	74.5	0.100	0.502	1.08	5.5E-6	0.581	2.70
Var	24	HT3	2.9E-6	2.77	35.6	1.5E-5	1.39	12.2	8.5E-6	0.560	3.35	1.24	7.55	27.2	6.35	26.8	74.8	0.102	0.503	1.08	2.7E-5	0.576	2.71
Var	24	IG2	0.001	2.72	20.7	1.0E-3	1.95	10.2	4.8E-4	0.538	2.53	1.23	5.81	17.9	5.94	21.1	55.4	0.117	0.503	1.05	7.9E-4	1.00	3.07
Var	24	IIN	0.001	2.86	22.7	8.8E-4	1.91	10.5	4.7E-4	0.541	2.54	1.23	5.84	18.0	5.93	21.1	55.9	0.116	0.502	1.05	6.7E-4	0.936	2.97
Var	24	JEFF	0.001	2.75	21.4	9.9E-4	1.96	10.3	4.9E-4	0.543	2.54	1.23	5.83	18.0	5.94	21.2	55.7	0.116	0.504	1.05	7.9E-4	1.00	3.07
Var	24	UNI	2.7E-9	3.12	36.7	4.5E-9	1.45	13.1	1.1E-9	0.565	3.42	1.24	7.79	29.0	7.69	27.6	63.6	0.101	0.503	1.08	1.8E-8	0.592	2.73

TABLE G.3: **Median CrI limits and posterior median for portion variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and selected priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
	σ_b^2	Prior	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU
SD	0.5	FLAT	1.53	2.91	5.26	1.65	2.45	3.52	2.22	2.46	2.73	2.21	2.46	2.73	2.22	2.46	2.73	2.31	2.45	2.60	1.95	2.39	2.90
SD	0.5	HCY	1.37	2.46	4.18	1.64	2.39	3.36	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.42	2.92
SD	0.5	HT3	1.37	2.46	4.18	1.65	2.39	3.35	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.42	2.92
SD	0.5	IG2	1.28	2.08	3.26	1.55	2.17	2.97	2.20	2.44	2.70	2.21	2.45	2.71	2.21	2.45	2.72	2.31	2.45	2.60	1.90	2.32	2.79
SD	0.5	IIN	1.56	2.39	3.58	1.71	2.34	3.13	2.22	2.45	2.72	2.22	2.46	2.73	2.22	2.46	2.73	2.31	2.45	2.60	1.96	2.37	2.84
SD	0.5	JEFF	1.28	2.08	3.25	1.55	2.18	2.97	2.2	2.44	2.70	2.21	2.45	2.71	2.21	2.45	2.72	2.31	2.45	2.60	1.90	2.32	2.79
SD	0.5	UNI	1.37	2.46	4.19	1.64	2.39	3.35	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.42	2.92
SD	6	FLAT	1.55	2.84	5.36	1.66	2.50	3.63	2.22	2.46	2.73	2.22	2.47	2.74	2.22	2.46	2.73	2.32	2.46	2.61	1.96	2.40	2.92
SD	6	HCY	1.39	2.53	4.38	1.63	2.43	3.49	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.46	2.60	1.97	2.43	2.95
SD	6	HT3	1.39	2.53	4.38	1.63	2.43	3.49	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.46	2.60	1.97	2.43	2.95
SD	6	IG2	1.29	2.13	3.38	1.56	2.23	3.10	2.20	2.44	2.70	2.21	2.45	2.72	2.21	2.45	2.72	2.31	2.45	2.60	1.91	2.33	2.82
SD	6	IIN	1.57	2.45	3.68	1.72	2.40	3.26	2.22	2.46	2.72	2.22	2.46	2.73	2.22	2.46	2.73	2.32	2.46	2.61	1.96	2.39	2.87
SD	6	JEFF	1.29	2.13	3.37	1.56	2.23	3.10	2.20	2.44	2.70	2.21	2.45	2.72	2.21	2.45	2.72	2.31	2.45	2.60	1.91	2.33	2.82
SD	6	UNI	1.39	2.53	4.38	1.63	2.42	3.49	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.46	2.60	1.97	2.43	2.95
SD	24	FLAT	1.53	2.76	5.43	1.66	2.52	3.70	2.22	2.46	2.73	2.22	2.46	2.73	2.22	2.46	2.73	2.32	2.45	2.60	1.95	2.40	2.91
SD	24	HCY	1.37	2.53	4.46	1.64	2.45	3.54	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.42	2.92
SD	24	HT3	1.37	2.53	4.46	1.64	2.45	3.54	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.42	2.92
SD	24	IG2	1.29	2.15	3.47	1.56	2.25	3.16	2.20	2.44	2.70	2.21	2.45	2.71	2.21	2.45	2.72	2.31	2.45	2.60	1.91	2.33	2.81
SD	24	IIN	1.58	2.48	3.77	1.73	2.43	3.33	2.22	2.45	2.71	2.22	2.46	2.73	2.22	2.46	2.73	2.31	2.45	2.60	1.96	2.39	2.86
SD	24	JEFF	1.29	2.16	3.47	1.57	2.26	3.16	2.20	2.44	2.70	2.21	2.45	2.71	2.21	2.45	2.72	2.31	2.45	2.60	1.91	2.33	2.81
SD	24	UNI	1.37	2.54	4.47	1.63	2.45	3.54	2.21	2.46	2.72	2.21	2.46	2.72	2.21	2.46	2.72	2.31	2.45	2.60	1.96	2.41	2.92
Var	0.5	FLAT	2.06	8.47	27.4	2.50	6.03	12.1	4.87	6.05	7.39	4.87	6.05	7.40	4.88	6.07	7.42	5.35	6.02	6.77	3.70	5.74	8.29
Var	0.5	HCY	1.60	6.05	17.0	2.50	5.72	10.9	4.86	6.03	7.37	4.85	6.03	7.37	4.85	6.03	7.37	5.32	6.02	6.76	3.76	5.85	8.42
Var	0.5	HT3	1.61	6.06	17.0	2.51	5.73	11.0	4.86	6.03	7.37	4.85	6.03	7.37	4.86	6.03	7.37	5.32	6.02	6.76	3.75	5.85	8.44
Var	0.5	IG2	1.47	4.31	10.4	2.26	4.72	8.62	4.82	5.96	7.27	4.82	5.98	7.32	4.83	6.00	7.34	5.32	6.00	6.74	3.52	5.38	7.66
Var	0.5	IIN	2.22	5.72	12.5	2.77	5.46	9.62	4.88	6.03	7.34	4.88	6.05	7.40	4.89	6.07	7.41	5.34	6.02	6.77	3.74	5.62	7.95
Var	0.5	JEFF	1.47	4.33	10.3	2.27	4.73	8.64	4.82	5.96	7.27	4.82	5.98	7.32	4.83	6.00	7.34	5.32	6.00	6.74	3.52	5.38	7.67
Var	0.5	UNI	1.59	6.07	17.0	2.49	5.72	11.0	4.86	6.03	7.37	4.86	6.03	7.37	4.85	6.03	7.36	5.32	6.02	6.76	3.75	5.85	8.42
Var	6	FLAT	2.10	8.08	28.1	2.53	6.24	12.8	4.88	6.06	7.41	4.89	6.08	7.43	4.88	6.07	7.43	5.36	6.03	6.79	3.74	5.78	8.42
Var	6	HCY	1.63	6.41	18.6	2.43	5.88	11.9	4.87	6.04	7.38	4.88	6.05	7.40	4.87	6.05	7.40	5.33	6.03	6.77	3.79	5.92	8.57
Var	6	HT3	1.62	6.41	18.6	2.41	5.88	11.9	4.87	6.04	7.38	4.87	6.05	7.40	4.87	6.05	7.40	5.33	6.03	6.77	3.79	5.92	8.57
Var	6	IG2	1.49	4.52	11.1	2.27	4.97	9.36	4.82	5.97	7.27	4.83	6.00	7.34	4.83	6.00	7.34	5.33	6.01	6.75	3.56	5.44	7.83
Var	6	IIN	2.26	5.98	13.2	2.80	5.74	10.4	4.88	6.03	7.34	4.89	6.07	7.41	4.89	6.07	7.41	5.36	6.03	6.78	3.78	5.70	8.12
Var	6	JEFF	1.48	4.54	11.1	2.27	4.97	9.37	4.82	5.96	7.27	4.83	6.00	7.34	4.83	6.00	7.35	5.34	6.01	6.76	3.56	5.45	7.82
Var	6	UNI	1.62	6.40	18.6	2.40	5.88	11.8	4.87	6.04	7.38	4.87	6.05	7.40	4.88	6.05	7.39	5.33	6.03	6.77	3.78	5.92	8.55
Var	24	FLAT	2.04	7.60	28.9	2.54	6.37	13.3	4.88	6.05	7.40	4.89	6.07	7.43	4.89	6.08	7.43	5.35	6.02	6.78	3.72	5.77	8.38
Var	24	HCY	1.57	6.42	19.3	2.45	6.00	12.2	4.86	6.03	7.36	4.86	6.04	7.38	4.87	6.04	7.38	5.32	6.02	6.77	3.74	5.84	8.44
Var	24	HT3	1.57	6.41	19.4	2.45	5.99	12.2	4.86	6.03	7.36	4.86	6.04	7.38	4.86	6.04	7.38	5.32	6.02	6.77	3.75	5.84	8.45
Var	24	IG2	1.49	4.63	11.7	2.28	5.08	9.74	4.82	5.95	7.26	4.83	5.99	7.32	4.83	6.00	7.34	5.32	5.99	6.74	3.55	5.43	7.81
Var	24	IIN	2.30	6.14	13.9	2.80	5.88	10.8	4.88	6.02	7.33	4.89	6.06	7.40	4.89	6.07	7.41	5.34	6.02	6.76	3.77	5.70	8.11
Var	24	JEFF	1.48	4.65	11.7	2.28	5.09	9.77	4.82	5.96	7.26	4.83	5.99	7.32	4.83	6.00	7.34	5.32	5.99	6.74	3.55	5.44	7.82
Var	24	UNI	1.57	6.44	19.4	2.44	5.98	12.2	4.86	6.03	7.36	4.86	6.04	7.38	4.87	6.04	7.39	5.32	6.02	6.77	3.74	5.83	8.44

TABLE G.4: **Median CrI limits and posterior median for total variance over σ_b^2=0.5, 6, 24 for evaluated scenarios and selected priors**

			b3k2s2_0.5			k2s2_0.5			k2s16_0.5			k2s16_6			k2s16_24			k6s16_0.5			k6s2_0.5		
	σ_b^2	Prior	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU	CrL	Median	CrU
SD	0.5	FLAT	6.31	132	1654	2.21	3.87	7.63	2.39	2.99	4.69	2.99	4.68	9.36	4.32	7.98	17.2	2.46	2.80	3.91	2.28	3.05	4.77
SD	0.5	HCY	1.82	4.05	10.1	2.04	3.07	4.71	2.35	2.77	3.49	2.90	3.91	5.90	4.06	6.14	9.69	2.45	2.71	3.19	2.25	2.82	3.62
SD	0.5	HT3	1.84	4.11	10.0	2.04	3.07	4.76	2.35	2.77	3.50	2.90	3.93	5.96	4.08	6.19	9.78	2.45	2.71	3.19	2.25	2.82	3.63
SD	0.5	IG2	1.69	3.58	9.84	2.00	2.97	4.53	2.35	2.73	3.31	2.87	3.77	5.53	4.01	6.02	9.71	2.45	2.68	3.04	2.24	2.76	3.49
SD	0.5	IIN	1.89	3.87	10.5	2.11	3.09	4.65	2.36	2.74	3.32	2.88	3.79	5.55	4.01	6.04	9.76	2.46	2.69	3.05	2.27	2.80	3.52
SD	0.5	JEFF	1.69	3.60	10.0	2.00	2.97	4.55	2.35	2.73	3.32	2.87	3.78	5.55	4.01	6.03	9.76	2.45	2.68	3.05	2.24	2.77	3.49
SD	0.5	UNI	2.02	4.17	8.46	2.03	3.09	4.86	2.35	2.77	3.53	2.91	3.96	6.15	4.25	6.31	9.27	2.45	2.71	3.21	2.24	2.82	3.66
SD	6	FLAT	7.05	142	2121	2.69	5.04	11.1	2.78	4.40	9.59	3.32	5.70	12.3	4.58	8.66	19.0	2.83	4.26	9.45	2.72	4.39	9.67
SD	6	HCY	2.14	4.85	12.0	2.53	4.02	6.68	2.74	3.78	6.20	3.23	4.67	7.41	4.33	6.62	10.5	2.79	3.75	6.14	2.67	3.79	6.24
SD	6	HT3	2.18	4.91	11.7	2.53	4.04	6.77	2.74	3.79	6.29	3.24	4.69	7.50	4.35	6.67	10.6	2.79	3.76	6.22	2.67	3.80	6.32
SD	6	IG2	1.98	4.37	12.8	2.44	3.81	6.30	2.71	3.60	5.66	3.17	4.45	7.06	4.24	6.49	10.7	2.78	3.56	5.56	2.63	3.63	5.73
SD	6	IIN	2.15	4.58	13.3	2.53	3.89	6.38	2.72	3.61	5.67	3.18	4.47	7.10	4.25	6.52	10.7	2.78	3.56	5.58	2.66	3.65	5.77
SD	6	JEFF	1.99	4.4	13.2	2.44	3.81	6.33	2.71	3.61	5.67	3.17	4.46	7.09	4.24	6.51	10.7	2.78	3.56	5.58	2.63	3.63	5.76
SD	6	UNI	2.51	4.94	8.93	2.57	4.1	6.91	2.75	3.84	6.44	3.28	4.77	7.60	4.57	6.76	9.64	2.80	3.81	6.38	2.69	3.85	6.46
SD	24	FLAT	8.92	188	3038	3.69	7.67	18.6	3.61	7.27	17.9	4.13	8.16	19.4	5.28	10.5	24.1	3.63	7.21	18.2	3.62	7.29	18.1
SD	24	HCY	2.81	6.51	16.3	3.45	5.88	10.4	3.54	5.76	10.3	4.00	6.38	10.9	5.00	7.90	12.8	3.55	5.75	10.2	3.50	5.74	10.3
SD	24	HT3	2.89	6.56	14.9	3.47	5.93	10.4	3.56	5.80	10.3	4.03	6.43	10.9	5.04	7.95	12.8	3.56	5.80	10.3	3.52	5.77	10.2
SD	24	IG2	2.55	6.09	19.7	3.37	5.70	10.4	3.47	5.55	10.1	3.90	6.16	10.9	4.87	7.83	13.4	3.51	5.55	10.1	3.46	5.58	10.2
SD	24	IIN	2.70	6.25	20.0	3.43	5.77	10.5	3.48	5.56	10.1	3.91	6.17	10.9	4.88	7.84	13.5	3.51	5.55	10.2	3.49	5.61	10.2
SD	24	JEFF	2.55	6.12	20.4	3.36	5.72	10.4	3.47	5.56	10.1	3.90	6.17	10.9	4.87	7.84	13.5	3.51	5.55	10.2	3.46	5.59	10.2
SD	24	UNI	3.35	6.26	9.47	3.88	5.93	8.85	3.87	5.81	8.65	4.39	6.43	9.14	5.38	7.79	10.4	3.87	5.80	8.60	3.85	5.79	8.65
Var	0.5	FLAT	39.8	17340	2737272	4.44	15.0	57.2	5.57	8.94	21.8	8.60	21.9	86.9	17.2	63.7	293	6.01	7.86	15.2	5.01	9.30	22.5
Var	0.5	HCY	2.67	16.4	99.3	3.83	9.40	21.7	5.44	7.67	12.1	8.08	15.3	34.3	15.3	37.7	92.1	5.95	7.36	10.1	4.91	7.94	13.0
Var	0.5	HT3	2.70	16.9	98.8	3.81	9.43	22.1	5.43	7.68	12.2	8.09	15.4	35.1	15.4	38.4	93.8	5.95	7.35	10.2	4.91	7.94	13.0
Var	0.5	IG2	2.52	12.8	95.7	3.71	8.80	20.1	5.45	7.44	10.9	8.01	14.2	30.2	15.1	36.2	92.7	5.99	7.20	9.24	4.88	7.63	12.0
Var	0.5	IIN	3.21	15.0	109	4.14	9.53	21.2	5.5	7.51	11.0	8.07	14.3	30.4	15.2	36.4	93.5	6.00	7.22	9.28	5.04	7.83	12.3
Var	0.5	JEFF	2.54	13.0	99.6	3.70	8.84	20.2	5.45	7.45	10.9	7.99	14.3	30.5	15.1	36.4	93.6	5.99	7.20	9.25	4.88	7.65	12.0
Var	0.5	UNI	3.17	17.4	69.5	3.79	9.54	23.0	5.42	7.69	12.3	8.08	15.7	37.2	16.7	39.8	83.7	5.95	7.36	10.3	4.90	7.96	13.2
Var	6	FLAT	49.7	20137	4497853	6.53	25.4	121	7.46	19.3	91.5	10.6	32.4	150	19.1	75.0	357	7.73	18.2	88.8	7.05	19.3	92.9
Var	6	HCY	3.72	23.6	142	5.78	16.1	43.5	7.20	14.3	37.9	9.92	21.8	54.0	17.3	43.8	108	7.58	14.1	37.3	6.74	14.4	38.4
Var	6	HT3	3.84	24.1	135	5.80	16.3	44.8	7.22	14.4	39.0	9.89	22.0	55.4	17.4	44.5	110	7.58	14.2	38.3	6.75	14.4	39.3
Var	6	IG2	3.44	19.1	162	5.54	14.5	38.9	7.11	13.0	31.7	9.70	19.8	49.2	16.8	42.2	112	7.54	12.7	30.6	6.65	13.1	32.4
Var	6	IIN	4.16	21.0	176	5.94	15.2	39.9	7.18	13.0	31.9	9.74	20.0	49.8	16.8	42.5	113	7.55	12.7	30.9	6.79	13.4	32.8
Var	6	JEFF	3.47	19.4	171	5.51	14.6	39.3	7.11	13.0	31.9	9.69	19.9	49.6	16.7	42.4	113	7.55	12.7	30.9	6.66	13.2	32.7
Var	6	UNI	4.92	24.4	77.0	5.92	16.8	46.6	7.26	14.7	40.9	10.1	22.7	56.7	19.3	45.7	90.6	7.60	14.5	40.3	6.81	14.8	41.2
Var	24	FLAT	79.6	35171	9226951	11.9	58.8	343	12.0	52.8	319	15.9	66.7	375	25.2	111	576	12.2	52.0	330	12.0	53.1	326
Var	24	HCY	6.37	42.4	260	10.5	34.5	106	11.5	33.2	104	14.7	40.7	117	22.7	62.4	160	11.6	33.1	103	11.1	33.0	103
Var	24	HT3	6.67	43.0	218	10.7	35.2	105	11.6	33.7	104	14.8	41.3	117	23.1	63.2	159	11.7	33.6	103	11.2	33.3	103
Var	24	IG2	5.68	37.1	382	10.3	32.5	106	11.2	30.8	100	14.3	37.9	117	22.0	61.3	177	11.5	30.8	101	11.2	31.2	102
Var	24	IIN	6.51	39.1	396	10.6	33.3	107	11.3	31.0	101	14.4	38.0	118	22.0	61.5	178	11.6	30.8	102	11.4	31.5	103
Var	24	JEFF	5.76	37.5	411	10.3	32.7	107	11.3	30.9	101	14.3	38.1	118	22.0	61.5	179	11.5	30.8	102	11.2	31.3	103
Var	24	UNI	9.33	39.1	86.5	13.1	35.2	75.4	13.5	33.8	72.5	17.6	41.3	81.7	27.0	60.6	105	13.5	33.6	71.9	13.3	33.6	72.2

G.3 Median of CrI Limits and Posterior Median including hpdl CrIs for HT3 Prior

The median lower and upper limits of the hpd intervals of all the simulated datasets (together with a horizontal bar for the median of the posterior median) were summarised in Figures 8.2 to 8.6 for the main priors of interest. Figures G.1 to G.4 additionally include the hpdl CrIs for the HT3 prior (denoted HT3L) for comparison. It is seen that the median CrIs lie above those for the other priors.

FIGURE G.1: **Limits of hpdv CrIs (and posterior median) for σ_p^2 for selected priors (including hpdl CrIs for HT3) plotted on icc scale (constant=6). Dashed lines represent the true value of 6.**

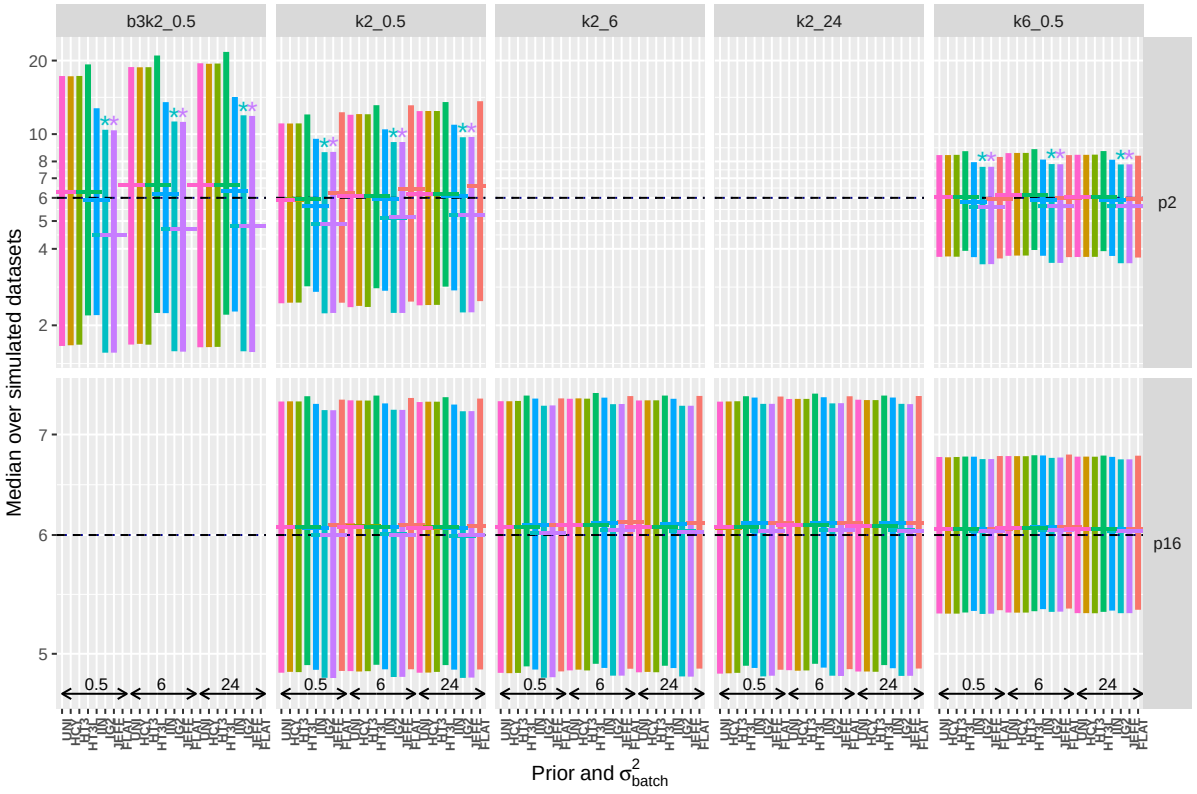


FIGURE G.2: **Limits of hpdv CrIs (and posterior median) for σ_b^2 for selected priors (including hpdl CrIs for HT3) plotted on icc scale (constant=6). Dashed lines represent the true values.**

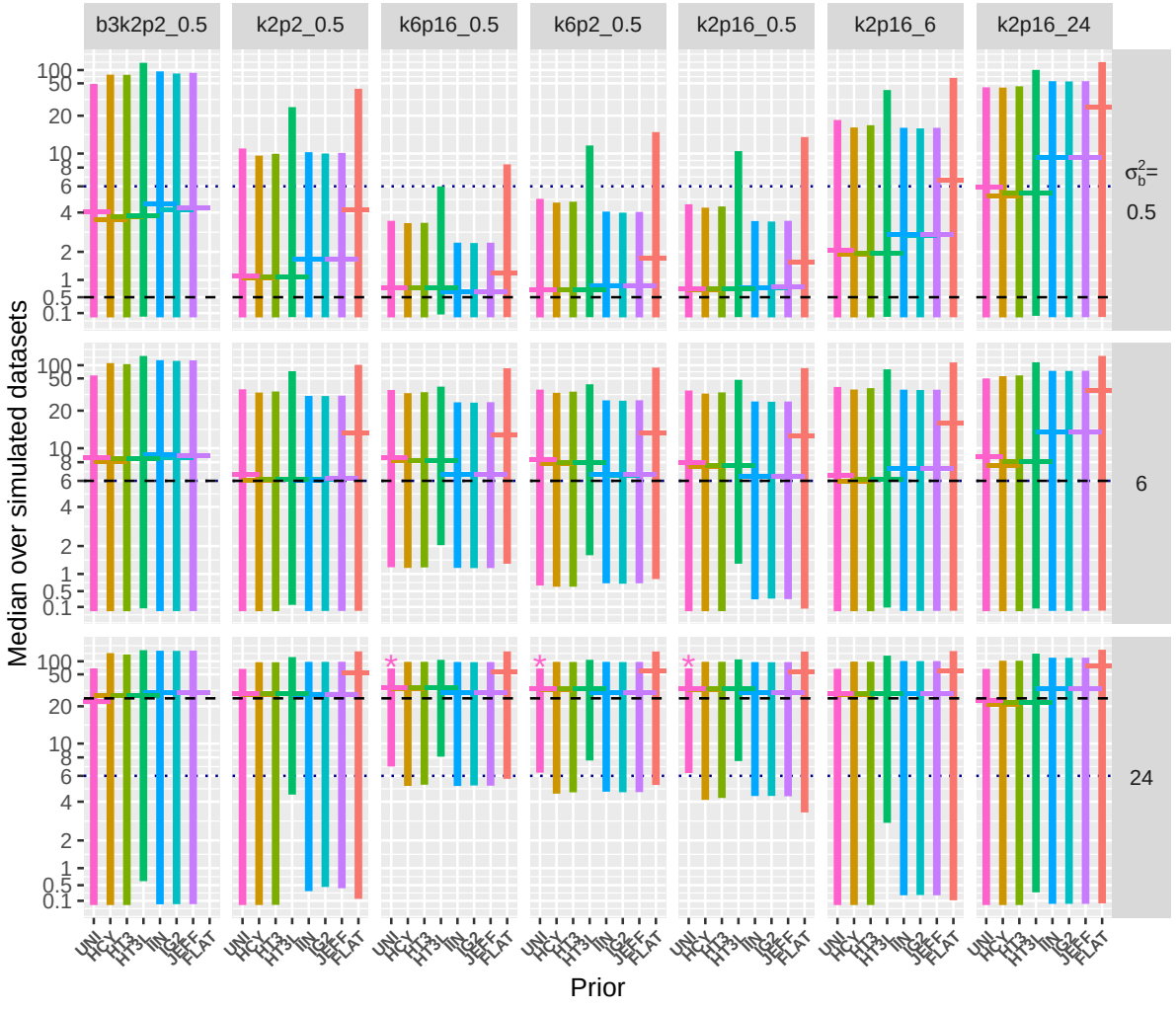


FIGURE G.3: **Limits of hpdv CrIs (and posterior median) for σ_k^2 for selected priors (including hpdl CrIs for HT3) plotted on icc scale (constant=6). Dashed lines represent the true values.**

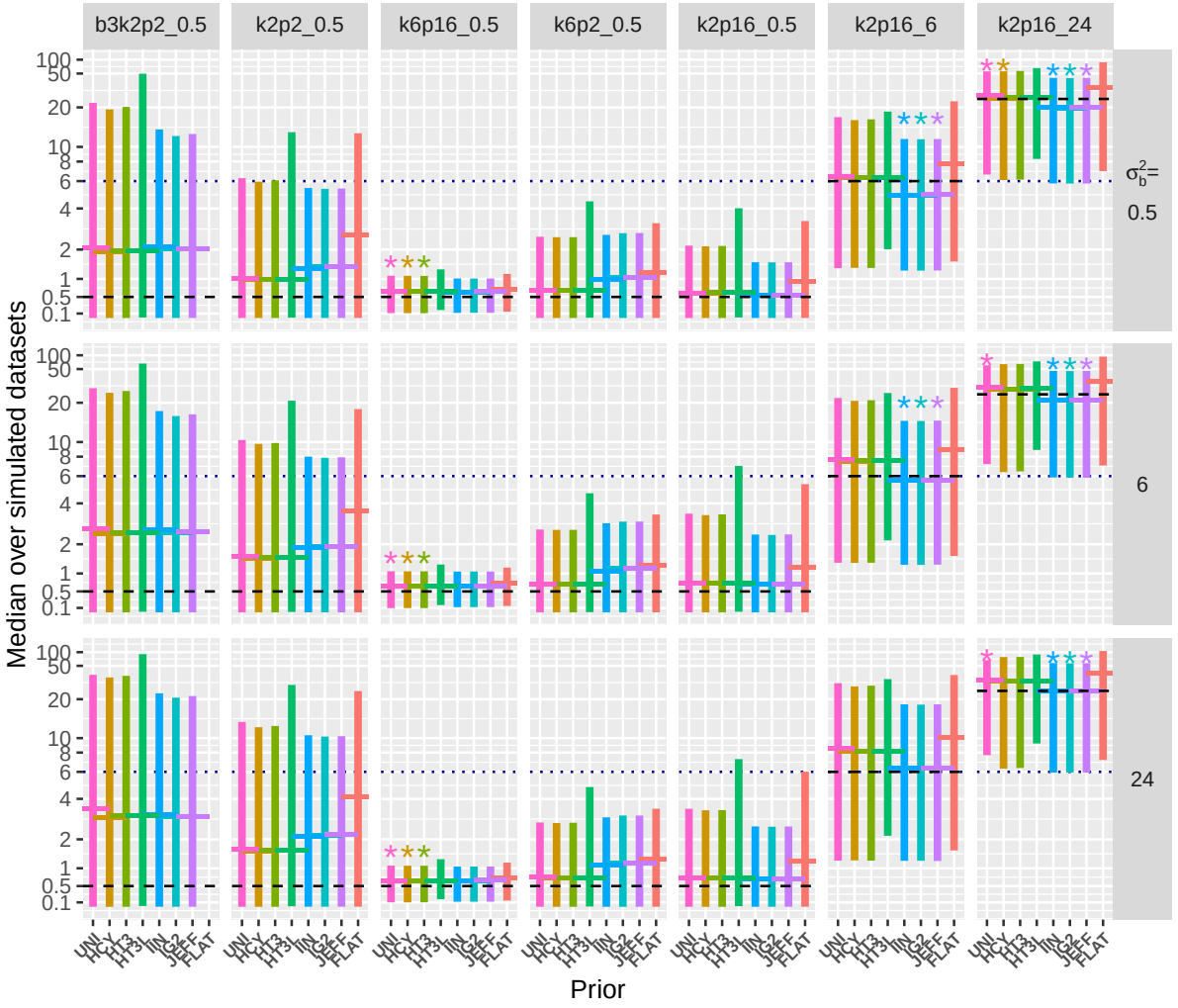
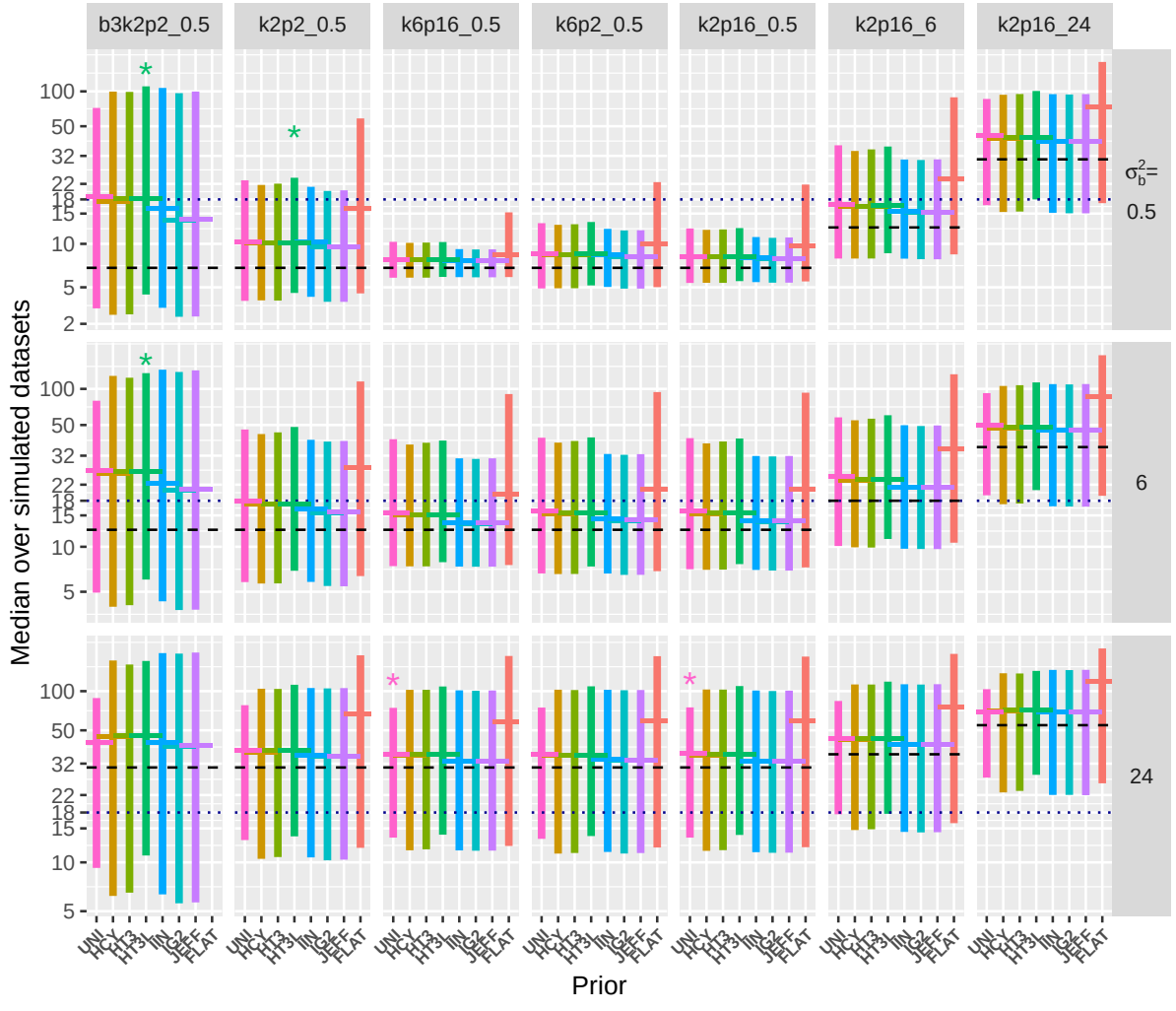


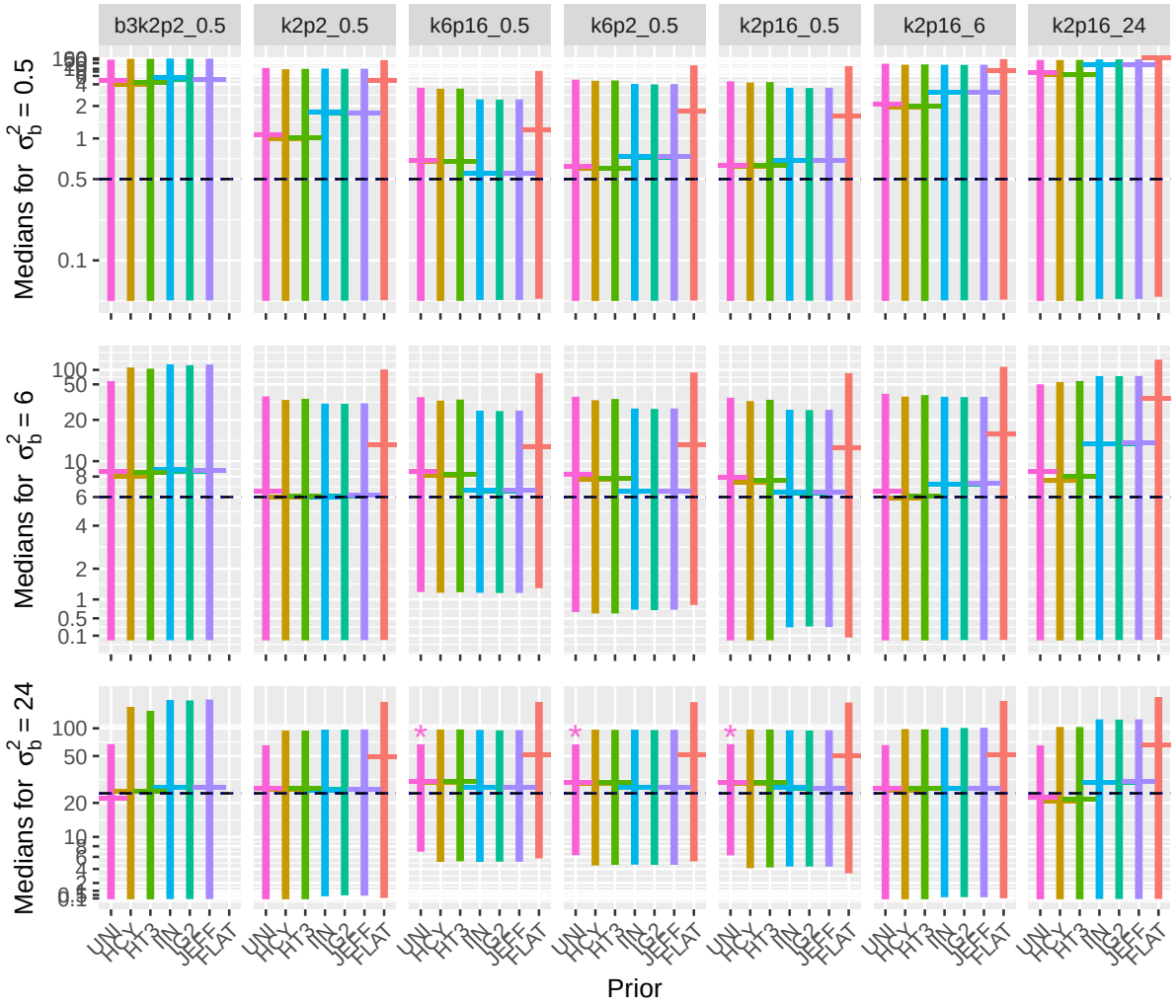
FIGURE G.4: **Limits of hpdv CrIs (and posterior median) for σ^2_{tot} for selected priors (including hpdl CrIs for HT3) plotted on icc scale (constant=18). Dashed lines represent the true values.**



G.4 Additional Summary Plot for Batch Variance

The medians (over all simulated datasets) of the hpdv CrI limits and the posterior median are shown in Figure G.5 for the batch variance, categorised by scenario and true batch variance. Instead of using the true portion variance (considered to be the intrinsic variance) as the constant (as used in Figure 8.3) the true batch variance was used as the constant for the icc scale and thus each row of subplots has a different constant. Given that the size of the batch variance relative to the portion variance is of interest, the graph is considered less useful as a summary than Figure 8.3. For example, for true σ_b^2=0.5 it is hard to see how the upper limits compare with the true portion variance of 6. However in circumstances where the comparison with the portion variance is not of interest then this would be a better alternative where for example, for true σ_b^2=0.5 the lower CrI limits are more easily compared.

FIGURE G.5: **Limits of hpdv CrIs (and posterior median) for σ_b^2 for selected priors plotted on icc scale (constant=0.5, 6,24). Dashed lines represent the true values.**



Bibliography

Adamec, E. and R. K. Burdick (2003). Confidence intervals for a discrimination ratio in a gauge R&R study with three random factors. *Quality Engineering 15*, 383–389.

Baldwin, S. A. and G. W. Fellingham (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods 18*, 151–164.

Barnett, K., P. McGregor, G. Martin, D. Le Blond, J. Weitzel, J. Ermer, S. Walfish, P. Nethercote, G. Gratzl, E. Kovacs, and H. Pappa (2016). Analytical target profile: Structure and application throughout the analytical lifecycle. *Pharmacopeial Forum 42*.

Bates, D. (2006). Rhelp lmer, p-values and all that. https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html.

Bates, D. (2010). lme4: Mixed-effects modeling with R. http://lme4.r-forge.r-project.org/book/.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles 67*(1), 1–48.

Bayarri, M. J. and J. O. Berger (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science 1*, 58–80.

Berger, J. O. and J. M. Bernardo (1992). Reference priors in a variance components problem. In P. K. Goel and N. S. Iyengar (Eds.), *Bayesian Analysis in Statistics and Econometrics* (1st ed.), Volume 75, Chapter 10, pp. 177–194. Springer, New York.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological) 41*(2), 113–147.

Besag, J. and P. J. Green (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society. Series B (Methodological) 55*(1), 25–37.

Borman, P., P. Nethercote, M. Chatfield, D. Thompson, and K. Truman (2007). The application of quality by design to analytical methods. *Pharmaceutical Technology 31*(12), 142–152.

Borman, P. J., M. J. Chatfield, I. Damjanov, and P. Jackson (2009). Design and analysis of method equivalence studies. *Analytical Chemistry 81*(24), 9849–9857.

Borman, P. J., M. J. Chatfield, I. Damjanov, and P. Jackson (2011). Method ruggedness studies incorporating a risk based approach: A tutorial. *Analytica Chimica Acta 703*(2), 101–113.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Addison Wesley Publishing Co.

Brooks, S. P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics 7*(4), 434–455.

Browne, W. J. and D. Draper (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics 15*(3), 391–420.

Browne, W. J. and D. Draper (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis 1*(3), 473–514.

Burdick, R. K., C. M. Borror, and D. C. Montgomery (2003). A review of methods for measurement systems capability analysis. *Journal of Quality Technology 35*(4), 342–354.

Burdick, R. K., C. M. Borror, and D. C. Montgomery (2005). *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models.* Society for Industrial Applied Mathematics (SIAM), Philadelphia, PA, USA.

Burdick, R. K. and F. A. Graybill (1992). *Confidence Intervals on Variance Components.* Marcel Dekkker.

Burdick, R. K., J. Quiroz, and H. K. Iyer (2006). The present status of confidence interval estimation for one-factor random models. *Journal of Statistical Planning and Inference 136*(12), 4307–4325.

Chang, I. H., B. H. Kim, and R. Mukerjee (2003). Probability matching priors for predicting unobservable random effects with application to ANOVA models. *Statistics & probability letters 62*(3), 223–228.

Chung, Y., A. Gelman, S. Rabe-Hesketh, J. Liu, and V. Dorie (2015). Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics 40*(2), 136–157.

Chung, Y., S. Rabe-Hesketh, and I. Choi (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in medicine 32*(23), 4071–4089.

Civit, S., M. Vilardell, A. Hess, Z. Matthew, Y. Ge, and A. Caballe (2011). *varcompci: VARCOMPCI: A Package for Computation of Confidence Intervals for Variance Components of Mixed Models in R.* R package version 1.0.

Cox, D. R. and P. J. Solomon (2003). *Components of Variance*. Chapman & Hall.

Crainiceanu, C. M. and D. Ruppert (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis 91*(1), 35–52.

Daniels, M. J. (1999). A prior for the variance in hierarchical models. *Canadian Journal of Statistics 27*(3), 567–578.

Datta, G. S. and D. D. Smith (2003). On propriety of posterior distributions of variance components in small area estimation. *Journal of Statistical Planning and Inference 112*, 175–183.

Datta, G. S. and T. J. Sweeting (2005). Probability matching priors. In D. Dey and C. Rao (Eds.), *Handbook of Statistics: Bayesian Thinking Modeling and Computation*, Chapter 3, pp. 91–114. Elsevier B.V. North Holland.

Demirhan, H. and Z. Kalaylioglu (2015). Joint prior distributions for variance parameters in bayesian analysis of normal hierarchical models. *Journal of Multivariate Analysis 135*, 163–174.

Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo method. *Los Alamos Science 15*, 131–137.

Ermer, J. and J. H. M. Miller (Eds.) (2005). *Method Validation in Pharmaceutical Analysis: A Guide to Best Practice* (first ed.). Wiley-556 VCH, Weinheim, Germany.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith (1990). Illustration of Bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association 85*(412), 972–985.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515–533.

Gelman, A. (2009). Bayes, Jeffreys, prior distributions and the philosophy of statistics. *Statistical Science 24*(2), 176–178.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science 7*, 457–511.

Ghosh, M. (2011). Objective priors: An introduction for frequentists. *Statistical Science 26*(2), 187–202.

Giesbrecht, F. G. (1989). A general structure for the class of mixed linear models. *Applications of Mixed Models in Agriculture and Related Disciplines, Southern Cooperative*

Series Bulletin, Louisiana Agricultural Experiment Station, Baton Rouge (343), 183–201.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (Eds.) (1996). *Markov Chain Monte Carlo in Practice* (first ed.). Chapman & Hall/CRC.

Gong, L., R. K. Burdick, and J. Quiroz (2005). Confidence intervals for unbalanced two-factor gauge R&R studies. *Quality and Reliability Engineering International 21*(8), 727–741.

Gustafson, P., S. Hossain, and Y. C. MacNab (2006). Conservative prior distributions for variance parameters in hierarchical models. *The Canadian Journal of Statistics 34*(3), 377–390.

Hammersley, J. M. and D. Handscomb (1964). *Monte Carlo Methods* (1st ed.). Methuen, London.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Hess, A. and H. Iyer (2002). A SAS/IML® Macro for computation of confidence intervals for variance components of mixed models. pp. Paper 252–27. SUGI (SAS Users Group International).

Hofer, J. D. and A. P. Rauk (2017). Use of Bayesian methods to analyze and visualize content uniformity capability versus United States Pharmacopeia and ASTM standards. *Journal of Pharmaceutical Sciences 106*(2), 554–559.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications* (1st ed.). Psychology Press.

Huang, A. and M. P. Wand (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis 8*(2), 439–452.

ICH (1994). ICH-Topic Q2(R1): Validation of analytical procedures.

Ionan, A. C., M. C. Polley, L. M. McShane, and K. K. Dobbin (2014). Comparison of confidence interval methods for an intra-class correlation coefficient (ICC). *BMC Medical Research Methodology 14*(1), 121.

Iyer, H. and T. Mathew (2002). Comments on Chiang (2001). *Technometrics 44*(3), 284–285.

Iyer, H. K. and P. D. Patterson (2002). A recipe for constructing generalized pivotal quantities and generalized confidence intervals. http://www.stat.colostate.edu/research/2002_10.pdf.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A 186*(1007), 453–461.

Johnson, L. and M. Deaner (2014). Necessary measures; expanded gage R&R to detect and control measurement system variation. *Quality Progress 47*(7), 34–38.

Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association 91*, 1343–1370.

K.M. Kurian, T. Mathew, G. S. (2008). Generalized confidence intervals for process capability indices in the one-way random model. *Metrika 67*, 83–92.

Lambert, P. C. (2006). Comment on article by Browne and Draper. *Bayesian Analysis 1*(3), 543–546.

Lambert, P. C., A. J. Sutton, P. R. Burton, K. R. Abrams, and D. R. Jones (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine 24*, 2401–2428.

Lesaffre, E. and A. B. Lawson (2012). *Bayesian Biostatistics.* (1st ed.), Volume 75. John Wiley & Sons.

Li, H. and H. S. Stern (1997). Bayesian inference for nested designs based on jeffreys's prior. *The American Statistician 51*(3), 219–224.

Lira, I. (2016). The GUM revision: the Bayesian view toward the expression of measurement uncertainty. *European Journal of Physics 37*(2), 025803.

Littell, R. C., G. A. Milliken, W. W.Stroup, R. D. Wolfinger, and O. Schabenberger (2006). *SAS for Mixed Models* (2nd ed.). SAS Publishing.

Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing 6*(2), 113–119.

Liu, Y., A. Gelman, and T. Zheng (2015). Simulation-efficient shortest probability intervals. *Statistics and Computing 25*, 809–819.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis.* CRC Press.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing 10*, 325–337.

Lynn, H. S., Z. Dong, and Z. Mu (2010). Comparison of software algorithms for calculating REML Wald type confidence limits for the between-group variance component in a small sample one-way random effects model example. *American Statistician 64*, 83–87.

Maechler, M. (2013). *Bessel: Bessel – Bessel Functions Computations and Approximations.* R package version 0.5-5.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (2008). *Generalized, Linear, and Mixed Models* (2nd ed.). John Wiley & Sons.

McNeish, D. M. and L. M. Stapleton (2016a). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review 28*(2), 295–314.

McNeish, D. M. and L. M. Stapleton (2016b). Modeling clustered data with very few clusters. *Multivariate Behavioral Research 51*(4), 495–518.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics 21*(6), 1087–1091.

Minitab 17 (2018). Gage R&R study (crossed). http://support.minitab.com/en-us/minitab/17/Assistant_Gage_R_and_R.pdf. Accessed 7 Apr. 2018.

Ni, S. and D. Sun (2003). Noninformative priors and frequentist risks of Bayesian estimators of vector-autoregressive models. *Journal of Econometrics 115*, 159–197.

Nijhuis, M. B. and E. R. Van den Heuvel (2007). Closed-form confidence intervals on measures of precision for an interlaboratory study. *Journal of Bipharmaceutical Statistics 17*, 123–142.

Perez, M.-E., L. R. Pericchi, and I. C. Ramirez (2017). The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis 12*, 615–637.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News 6*(1), 7–11.

Polson, N. G. and J. G. Scott (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis 7*, 887–902.

Prada-Sáichez, J. M. and T. Cotos-Yánez (1997). A simulation study of iterated and non-iterated bootstrap methods for bias reduction and confidence interval estimation. *Communications in Statistics-Simulation and Computation 26*, 927–946.

Quiroz, J. and R. Baumgartner (2017). Interval estimations for variance components: A review and implementations. https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/NCB_Conference/Presentations/2017/03_Jorge%20Quiroz_2017-05-11-NCB2017.pdf. Presentation at 2017 Nonclinical Biostatistics Conference, accessed 17 Apr. 2018.

Raftery, Adrian E.; Lewis, S. M. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science 7*, 493–497.

Rao, C. R. (1972). Estimation of variance and covariance components in a linear model. *Journal of the American Statistical Association 67*, 112–115.

Rasbash, J. and H. Goldstein (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral statistics 19*, 337–350.

Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley & Sons, New York.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*, 319–392.

Sahai, H. and M. M. Ojeda (2005). *Analysis of Variance for Random Models: Theory, Methods, Applications, and Data Analysis, Volume 2: Unbalanced Data*. Birkhauser, Boston.

SAS Institute Inc.: B80011 (2011). SAS/STAT 9.22 hot fix downloads, B80011, SAS Institute Inc. http://ftp.sas.com/techsup/download/hotfix/HF2/B80.html. Accessed 29 Dec. 2017.

SAS Institute Inc.: PN41931 (2010). Problem Note 41931: LOGG values computed in the OUT= data set of the PRIOR statement in PROC MIXED are incorrect. http://support.sas.com/kb/41/931.html. Accessed 29 Dec. 2017.

SAS Institute Inc.: PN48940 (2013). Problem Note 48940: Conjugacy Sampler in PROC MCMC produces incorrect results with MISSING=CC. http://support.sas.com/kb/48/940.html. Accessed 17 Apr. 2018.

SAS Institute Inc.: SAS/STAT (2011). SAS/STAT software. https://www.sas.com/en_gb/software/stat.html. Accessed 29 Dec. 2017.

SAS Institute Inc.: The MIXED Procedure (2013). Chapter: The MIXED Procedure from SAS/STAT® 13.1 user's guide. Copyright © 2013 SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/131/mixed.pdf. Accessed 29 Dec. 2017.

Searle, S. R. (1970). Large sample variances of maximum likelihood estimators of variance components using unbalanced data. *Biometrics 26*, 505–524.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science 32*, 1–28.

Smith, A. F. M. and A. E. Gelfand (1992). Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician 46*, 84–88.

Stan (2017). Wiki: Prior choice recommendations. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations. Accessed 11 Apr. 2018.

Stegmueller, D. (2013). How many countries for multilevel modeling? a comparison of frequentist and bayesian approaches. *American Journal of Political Science 57*, 748–761.

Stroup, W. and R. Littell (2002). Impact of variance components estimates on fixed effect inference in unbalanced linear mixed models. pp. 32–48. Proceedings of the 14th Annual Kansas state university conf in Applied Statistics in Agriculture.

Sturtz, S., U. Ligges, and A. Gelman (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software 12*(3), 1–16.

Sweeting, T. J. (2008). On predictive probability matching priors. *IMS Collections Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh 3*, 46–59.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 22*(4), 1701–1728.

Tong, Y. and R. L. Brennan (2007). Bootstrap estimates of standard errors in generalizability theory. *Educational and Psychological Measurement 67*(5), 804–817.

Tsai, M. Y. and C. K. Hsiao (2008). Computation of reference Bayesian inference for variance components in longitudinal studies. *Computational Statistics 23*(4), 587–604.

van de Schoot, R., J. J. Broere, K. H. Perryck, M. Zondervan-Zwijnenburg, and N. E. van Loey (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology 6*(1), 25216.

van der Linde, A. (2000). Reference priors for shrinkage and smoothing parameters. *Journal of Statistical Planning and Inference 90*(2), 245–274.

Vardeman, S. B. and E. S. VanValkenburg (1999). Two-way random-effects analyses and gauge R&R studies. *Technometrics 41*, 202–211.

Walfish, S. (2006). Analytical methods: A statistical perspective on the ICH Q2A and Q2B guidelines for validation of analytical methods. *BioPharm International 19*(12), 28–36.

Wang, F. K. and E. Li (2003). Confidence intervals in repeatability and reproducibility using the Bootstrap method. *Total Quality Management & Business Excellence 14*(3), 341–354.

Wang, Y. and H. Chen (2012). On testing an unspecified function through a linear mixed effects model with multiple variance components. *Biometrics 68*(4), 1113–1125.

Weaver, B. P., M. S. Hamada, S. B. Vardeman, and A. G. Wilson (2012). A Bayesian approach to the analysis of gauge R&R data. *Quality Engineering 24*(4), 486–500.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association 88*(423), 899–905.

Weitzel, J., J. Meija, D. LeBlond, and S. Walfish (2018). Measurement uncertainty for the pharmaceutical industry. *Pharmacopeial Forum 4*.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2016). *scales: Scale Functions for Visualization*. R package version 0.4.0.

Witkovsky, V. (2001). Computing the distribution of a linear combination of inverted gamma variables. *Kybernetika 37*, 79–90.

Wolfinger, R. D. and R. E. Kass (2000). Nonconjugate Bayesian analysis of variance component models. *Biometrics 56*, 768–774.

Wolfram|Alpha (2015). Wolfram Alpha LLC. http://www.wolframalpha.com/calculators/integral-calculator/. Accessed during 2015.

Wu, C.-W. (2008). Assessing process capability based on bayesian approach with subsamples. *European Journal of Operational Research 184*(1), 207–228.

Ye, K. (1994). Bayesian reference prior analysis on the ratio of variances for the balanced one-way random effect model. *Journal of Statistical Planning and Inference 41*, 267–280.

Ye, R. and S. Wang (2007). Generalized p-values and generalized confidence intervals for variance components in general random effect model with balanced data. *Jrl Syst Sci & Complexity 20*, 572–584.