UNIVERSITY OF SOUTHAMPTON
Division of Social Statistics and Demography
Faculty of Social and Human Sciences
and
Institute of Complex Systems Simulation

A thesis submitted in fulfilment of the requirements for a
PhD in Complex Systems Simulation

Supervisors:
Dr. Jakub Bijak
Prof. Seth Bullock

Managing Uncertainty in Agent-Based Demographic Models

by Jason Hilton

January 9, 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

DIVISION OF SOCIAL STATISTICS AND DEMOGRAPHY
FACULTY OF SOCIAL AND HUMAN SCIENCES
and
INSTITUTE OF COMPLEX SYSTEMS SIMULATION

A thesis submitted to meet the requirements for a PhD in Complex Systems Simulation

Managing Uncertainty in Agent-Based Demographic Models

by Jason Hilton

Population-level patterns are the object of formal demographic study, but result from thousands of decisions made by individuals who both act and are affected by the actions of others. These properties can lead to difficulties in determining the mechanisms underlying population change, not least because of the possibility that human population may display *complex* properties, where macro-level patterns are not reducible to the sum of their parts. Simulation methods may help overcome some of these analytical difficulties. Agent-Based Models are simulations which explicitly represent the behaviour of individuals and their interactions, and allow population patterns to emerge from such interactions.

Agent-Based Models are attractive to demographers because they allow the formalisation of theories about links between individual-level behaviour and interaction on one hand, and macro-level population patterns on the other. However, such simulations are notoriously difficult to analyse and calibrate; they tend to involve many free parameters, and include several sources of uncertainty. This thesis investigates how the application of techniques from the field of the design and analysis of computer experiments can be fruitfully applied to these problems. More specifically, three demographic simulations are used to demonstrate the utility of Gaussian process emulators for this purpose. Firstly, a replication of an existing demographic agent-based model is analysed using heteroskedastic emulators. Secondly, two emulator-based methods are trialled for their effectiveness in calibrating a microsimulation. Finally, Gaussian processes are used to analyse and calibrate a agent-based model of intergenerational fertility patterns against empirical observations.

These examples demonstrate the ability of Gaussian process emulators to flexibly capture non-linearities in the relationship between simulation inputs and outputs, and to coherently account for uncertainties. These properties mean that they are well suited to the problem of analysing and calibrating Agent-Based models. In the concluding chapter, thoughts are offered on strengths and limitations of the techniques in comparison to other methods, and directions for further work are suggested.

Declaration of Authorship

I, Jason Hilton, declare that the thesis entitled *Managing Uncertainty in Agent-Based Demographic Models* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.
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. Parts of this work have been published as:
   - Hilton, J., & Bijak, J. (2016). Design and Analysis of Demographic Simulations. In A. Grow & J. Van Bavel (Eds.), *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications* (pp. 211235). Basel: Springer.

Signed: ....

Date: ....

Contents

List of Figures

Notation

$f(.)$	— Uncertain function corresponding to simulator
N	— Number of Simulator Runs
k	— Input Dimension
x	— Point in the design space - vector length k
D	— Matrix of design points - dimension (N, k)
$m(.)$	— Mean function of Gaussian Process
$c(.,.)$	— Correlation function of Gaussian Process
$h(.)$	— Basis functions of input
q	— Number of elements in basis function per design point
H	— Matrix generated by $h(D)$
A	— Correlation Matrix generated by $c(D, D)$ - dimension (N, N)
m	— Number of Validation Samples
β	— Vector of basis function coefficient parameters
σ^2	— Gaussian Process variance parameter
ω_i	— Correlation function roughness parameter,i^{th} input
Ω	— Diagonal matrix of correlation inputs ω
τ^2	— Nugget parameter representing stochastic simulation variance
α	— $= \frac{\sigma^2}{\sigma^2+\tau^2}$ Fraction of Gaussian Process variance due to σ^2
$v(.,.)$	— $= \sigma^2 c(.,.) + \tau^2 I_{i=j}$ Variance-Covariance function for Gaussian Process

Acknowledgements

This thesis was funded from the Engineering and Physical Sciences Research Council. My doctoral studentship was funded through the Care Life Cycle project (EPSRC grant number EP/H021698/1), and my studies were undertaken as part of the Institute for Complex Systems Simulation doctoral programme (EPSRC Doctoral Training Centre grant EP/G03690X/1). I am deeply thankful for both of these groups for allowing me to develop a unique combination of skills which would have not otherwise been in a position to gain. In particular, I am thankful to Prof. Seth Bullock, whose desire for a truly multi-disciplinary institute allowed me to learn so much from fellow students at the ICSS, and who also kindly agreed to take over part of my supervision. Dr Jason Noble also supervised the earlier part of my thesis, as well as undertaking much of the teaching in the first year of my studies. His breadth of knowledge and dedication to students and colleagues was an inspiring example, and his departure is a sore loss to academia.

My thinking about the practical and philosophical aspects of Agent-Based Modelling developed through conversations with the Modelling Social Systems reading group, and in particular with Eric Silverman, Stuart Rossiter and Joe Viana. The Java code for the replication of the simulation model of Billari et al. (2007) analysed in Chapter 4 was developed by Viet Dung Cao and Eric Silverman. The subsequent analyses and the emulation code are my own.

This thesis was produced almost entirely using free and open-source software, from the analysis to the graphing to the typesetting, so I would particularly like to thank the countless contributors of 'R' and 'python' packages whose provision of time and energy provide so much utility to the rest of us. I hope that I am now in a position to be able to give something back to the community.

I also wish to thank the Max Planck Institute for Demographic Research in Rostock and in particular Prof. Frans Willekens for hosting and providing a studentship for a 3-month break in my studies. Discussions with colleagues working on agent-based modelling during this period were extremely beneficial in helping me develop my thinking on the topic, and I thank them greatly for their hospitality. I would also like to thank Dr. Sabine Zinn for providing files to enable the running of the MicCore microsimulation package used in Chapter 5.

Much of the computation undertaken as part of this thesis would have been impossible without access to the Iridis 4 High-Performance Computing Facility, and acknowledgement of the dedicated support staff is thus required. My deep thanks also to Hans Fangohr and Ian Bush from the University of Southampton and the National Algorithms Group respectively for their expert tuition on High-Performance Computing during the Advanced Computational Method II course.

I also wish to acknowledge my deep debt to my main supervisor, Dr. Jakub Bijak, who introduced me to both Agent-Based Modelling and Bayesian statistics. His constant encouragement and wise counsel, together with his willingness to make time for discussion despite his busy schedule, have been valuable beyond measure.

On a more personal level, my parents deserve endless thanks for bringing me up right and for always nurturing my curiosity. They provided me with the skills and opportunities to do whatever I liked with my life. My endless love and gratitude to them and to my sister Nikki, who can always make me smile, and is never afraid to tell me when I'm being an idiot. Finally, I would have been unable to complete this thesis without the support of my partner, Dafni Papoutsaki. Her encouragement, advice, and love have been constant sources of strength throughout my PhD, and her proof-reading of the final version of this document was also immensely valuable. She is the positive aspect!

1

Introduction

1.1 Overview

A considerable methodological challenge for contemporary Demography is the problem of reconciling individual behaviour with population level trends (Lee, 2001; Ní Bhrolcháin & Dyson, 2007; Courgeau, 2012; Billari, 2015). Agent-Based Modelling is a simulation methodology that explicitly models how macro-level patterns are generated by the behaviour and interaction of individuals at the micro-level (Epstein & Axtell, 1996). The method thus provides exciting opportunities for those seeking the micro-foundations of demographic changes (Billari et al., 2003), and as result is gaining traction as a modelling approach within the discipline. As Agent-Based Models gain more traction in demography, a more systemic approach to their analysis and calibration will be required, particularly as models may grow more complicated (Grazzini et al., 2013a)[1]. As demographers explore the possibilities of the methodology, they may attempt to make their agent-based simulations match reality more closely, to model more fields of social life, to pay attention to the effect of institutions and policy; and to enrich their model with more data to attempt to bring them in line with what is observed (Silverman et al., 2011; Squazzoni, 2012).

This progress towards greater sophistication in agent-based approaches will however introduce additional sources of uncertainty to the modelling process, which need to be accounted for (Kennedy & O'Hagan, 2001a). Furthermore, it will create a need for a more considered approach to the design and analysis of agent-based computer *experiments* within demography (Santner et al., 2003). This thesis discusses how existing techniques from the computational experiment literature might be fruitfully applied to overcome some of these difficulties in the specific realm of demographic Agent-Based Modelling.

1.2 Agent Based Modelling

All scientific modelling attempts to reduce some real-world phenomenon of interest to an abstracted idealised version, which can be easily manipulated in order to understand and analyse the underlying relationships and dynamics, and perhaps to predict future system states (Franck, 2002; Gilbert & Troitzsch, 2005; Werker & Brenner, 2004). Mathematical modelling, for instance, represents the real world in the form of systems of equations, the solutions of which it is hoped will provide insight into how the world works (an approach that is more effective for the natural world than it is for the social (Wigner, 1960; Moss & Edmonds, 2005)). With any such endeavour, the modeller is faced with a

[1] Throughout this document a distinction will be maintained between the complicated and complex, with the latter reserved as a technical term referring to properties of systems of many interacting parts exhibiting emergent behaviour (on which more throughout the thesis), while the former corresponds to a more everyday understanding of the term, for example referring to things that are difficult to understand and/or have large number of components (Mitchell, 2009).

number of decisions about what elements of the real world to include, how they should be represented, and what simplifications to make (Chwif et al., 2000; Rossiter et al., 2010).

Agent-Based Modelling is no different in this respect, except that as a form of simulation modelling, it represents the real world as a computer program (Gilbert, 2008). The ability to utilise computational power to approximate the behaviour of systems which can not be reduced to analytical forms means that more flexible types of model are possible (Squazzoni, 2012). Furthermore, two elements distinguish it from other forms of simulation modelling. Firstly, individuals are represented explicitly, in the sense that they have separate representations in the program code and are not treated merely in aggregate (Gilbert, 2008). Secondly, these individuals are endowed with agency, in that some element of their decision-making process is modelled, usually in the form of a set of simple behavioural rules, but sometimes involving more sophisticated Artificial Intelligence routines (Billari et al., 2006).

While these factors are considered here as *definitional* of Agent-Based Modelling, several other elements are crucial to ABM in practice. In particular, some mechanisms by which agents interact are almost always included, and random number generators are used to mimic the 'randomness' of the real world (Gilbert, 2008). This combination of many interacting units and stochasticity can lead to what is termed 'emergent' behaviour (Mitchell, 2009). This can be a difficult concept to pin down, but the essence is that repeated local interactions can cause the formation of positive feedback loops that come to dominate the behaviour of the whole system (ibid). When viewed in terms of the relationship between model parameters and outputs, this often results in highly non-linear mappings, path dependence, and even 'tipping point' and 'phase transition' behaviour, whereby the system moves quickly between two different regimes (Macy & Willer, 2002; Bonabeau, 2002; Mitchell, 2009; Squazzoni, 2012). As a result, it becomes difficult to anticipate the behaviour of the whole system from knowledge about the behaviour of individual elements (Axelrod, 2003). The field of complexity science devotes itself to understanding these sorts of systems, and Agent-Based Models are often used in this field.

These kind of emergent effects themselves form part of the justification for explicitly simulating individuals (Bonabeau, 2002); if interactions can give rise to unexpected and unpredictable system behaviour, then macro-level effects cannot be approximated by simple aggregation of the individual behaviour, as is common in the macroeconomic 'representative agent' approach (Potts, 2000; Grazzini et al., 2013a; Kirman, 2011). However, this also makes the analysis of simulation results more difficult, and therefore necessitates a more sophisticated statistical approach.

1.3 Analysing Agent-Based Models

This thesis advocates a framework within which the analysis of Agent-Based Models can be approached. To this end, it is useful to briefly outline the types of uses to which ABMs are typically put, and what might be the peculiar requirements of specifically demographic ABMs, in order to set out what is required.

1.3.1 "Opaque Thought Experiments" or Social Simulations?

As was discussed in the previous subsection, modelling necessarily involves simplification and abstraction from the target phenomena. ABM practitioners, however, often differ on the types of abstractions to be made. In some cases, ABMs are seen as ways of testing theories and abstract hypotheses about social systems through "opaque thought experiments" (Di Paolo et al., 2000) (opaque because of the complexity of the system), and thus the simulations built are often highly abstract. Silverman (2007) terms social research in this mode as 'Systems Sociology'. ABMs are also used to attempt to explain specific empirical phenomena, in which case they tend to be more concrete and are more readily comparable to data. At the far end of the spectrum, ABMs may be considered 'social simulations' which attempt to predict or represent reality very closely (Silverman, 2007), and thus tend to include more detail in their model specifications (e.g. Kniveton et al., 2011).

These positions are not irreconcilable, and may often be held at different times by the same people: it is possible to advocate simple models for exploratory modelling, and more complicated models if empirical fidelity is the aim (Squazzoni, 2012). Indeed, one can see them as representing respectively the starting point and end goal of an accretive modelling process (Cioffi-Revilla, 2010). However, it is clear that the requirements in terms of analysis of results are different. For models for which the aim is theory exploration, the focus is on understanding the model, the mapping of parameters to outputs and sensitivities of the model to its various inputs (Squazzoni, 2012). For more empirically focused models, the relationship between the simulation and reality must also be examined (Werker & Brenner, 2004; Boero & Squazzoni, 2005).

This thesis demonstrates how Gaussian process emulators (e.g. Kennedy & O'Hagan, 2001a) can be used to help analysis in both these cases within the context of demography. Gaussian process emulators are meta-models - higher-level models that aid in the analysis of the more complex underlying model. Meta-models are used extensively in simulation in engineering and the natural sciences (e.g. Forrester et al., 2008) and operations research (e.g. Kleijnen, 2008), but have not been used systematically in social science simulation (although see Silverman et al., 2013; Bijak et al., 2013; Grazzini et al., 2013a; Boukouvalas et al., 2014; Heard, 2014; Kamiński, 2015; Andrianakis et al., 2015).

1.3.2 Agent-Based Modelling in Demography

The edited volumes of Billari et al. (2003) and Billari et al. (2006) made eloquent cases for the use of Agent-Based Models in demography to enable the study of micro-macro linkages and to examine theories about individual behaviour in population studies and other social sciences. However, it is worth thinking about what a specifically *demographic* ABM entails, and how this might effect how we choose to analyse it. It is suggested that demographic ABMs are likely to differ from those in other fields in the simulation outputs of interest. Specifically, as with other methods, demographers are likely to be interested in *count* and *rate* data decomposed by *age*, *sex* and *time* - these elements constitute the bedrock of demography as a discipline (Keyfitz & Caswell, 2005). This requirement and some of the difficulties it imposes are discussed in future chapters.

1.4 Uncertainty

In demography as a whole, estimation, forecasting and projection techniques are gradually moving towards the increased use of probabilistic methods (e.g. Abel et al., 2013; Raymer et al., 2013). This is important as if projections are to be used for policy, then only those expressed in a probabilistic manner allow a coherent approach to decision making under uncertainty (Keilman et al., 2002; Bijak, 2011).

Using simulation methodology does not change this imperative, it merely adds additional factors about which we are unsure (Kennedy & O'Hagan, 2001a). A principled and consistent treatment of the uncertainties in Agent-Based models is required if we are to fully understand their behaviours, implications, and relationships to reality. This is particularly the case if we have designs on using such models to inform policy - obtaining predictive distributions of various quantities accounting for all sources of uncertainty would facilitate a decision analytic approach to such problems (Bijak, 2011). The lofty goal of producing policy relevant demographic ABM may not be within our grasp as yet, but similar endeavours are also taking place in other fields. In economics for example, large scale ABMs are currently being developed in order to help determine the likelihood of future financial crashes (Buchanan, 2009).

Unfortunately, often many sources of uncertainty are ignored in the analysis and calibration of Agent-Based models, with some notable recent exceptions (Boukouvalas et al., 2014; Andrianakis et al., 2015). There exist a number of methods by which to get a handle on model uncertainties within the context of simulation. This thesis focuses mostly on the use of Gaussian Process Emulators, because, it will be argued, these are flexible, efficient and aid with model interpretation. However, other approaches such as Bayesian Melding (Poole & Raftery, 2000; Ševčíková et al., 2007a) and Polynomial

Chaos (see O'Hagan, 2013) are also viable alternatives, and these are discussed briefly in chapter 3.

1.5 Contribution

The aim of this thesis is not to develop a new statistical technique for the analysis of demographic Agent-Based Modelling. Rather, it is to show how existing techniques in the design and analysis of numerical simulations can be fruitfully applied in this field. In particular, it shows how Gaussian Process Emulators can be helpful both in uncovering and helping to explain the internal behaviour of ABMs of population process, and for relating more empirically relevant simulations to observations of reality.

The work undertaken details a pragmatic methodological template for analysis of demographic agent-based models. Upon building an agent-based model based on empirical data and/or theoretical considerations, the simulation should be run on Latin-hypercube sample in order to maximise the information gained from each run (Santner et al., 2003). Repeated points should be obtained for at least some points, and the results should be used to fit heteroskedastic emulators (Boukouvalas, 2010) to better understand the relationships between simulation inputs and outputs. These emulators provide fast approximations to the simulator and also describe its variance, and therefore allow areas of the parameter space displaying interesting properties to be identified. Sensitivity analysis further allows the modeller to identify the parameters (and thus mechanisms) in their simulation which are key to driving demographic change.

The thesis also provides guidance on how to approach the calibration problem. Although the full Bayesian calibration approach of Kennedy & O'Hagan (2001a) is conceptually appealling, the more heuristic history matching method of Vernon et al. (2010) is favoured, because it provides more flexibility in capturing multiple outputs. Modellers should be careful not to confuse calibration with validation, however (Oreskes et al., 1994), as the former only identifies model specifications that are consistent with the observed data (and discrepancy function), of which there may be many.

Finally, as a demonstration of the efficacy of the methods set out, emulation techniques are used to calibrate a Agent-Based Model of the Easterlin Effect (Easterlin, 1987), showing that wave-like patterns in fertility can be generated by simple micro-level behavioural rules similar to those specified by Easterlin. As a by-product of this exercise, a relatively general demographic simulator was built that allows the introduction of arbitrary models for fertility, and thus is suitable for extension to investigate similar problems.

Parts of Chapters 3 and 4 have been published as a contribution to a peer-reviewed book on demographic Agent-Based Modelling (Hilton & Bijak, 2016), and develop on work in Silverman et al. (2013) and Bijak et al. (2013), to which the author contributed.

1.6 Structure

An overview of the structure of this thesis is now given. Chapter 2 examines the scientific purposes of Agent-Based Models and existing approaches to their analysis and calibration, and additionally reviews existing work in Demographic Agent-Based Modelling. Chapter 3 focuses on a review of the design and analysis of computer experiments, and includes an introduction to Gaussian Processes and their use in this context. Chapter 4 describes the use of emulators to examine a simple demographic ABM, and examines how heteroskedastic emulators proposed by Boukouvalas (2010) may provide a better fit to the data given the non-linearities common in Agent-Based Models. Chapter 5 uses Gaussian Processes to analyse the microsimulation model of Zinn & Himmelspach (2009). The focus of this chapter is on the calibration problem, and how to adapt emulators to take into account distributions over age, a problem that is more specific to demographic ABMs. Two approaches are trialled: firstly, age is treated as continuous and the full Bayesian calibration methods of Kennedy & O'Hagan (2001a) are used to obtain a probability distribution for the unknown parameters. Secondly, separate emulators are fitted at different points along the age spectrum, and the 'history matching' techniques of Vernon et al. (2010) and Andrianakis et al. (2015) are applied. Chapter 6 describes a novel ABM of inter-generation cycles in fertility, adapting the work of Easterlin (1975) and Lee (1974) into an Agent-Based setting, and showcases the use of GPs to calibrate a demographic ABM. The final Chapter (7) offers conclusions, including some thoughts on the limitations and prospects of the approach in demography, with comparisons to other approaches.

1.7 The production of this thesis

This thesis was written in Rmarkdown, an R-based extension of the 'markdown' format, before being converted to `tex` and ultimately `pdf` format. The use of Rmarkdown allows R code for graphs and tables to be interleaved with text of the thesis, giving two practical benefits. Firstly, changes in the code for generating figures will be automatically reflected in the final document every time it is re-compiled, without the need for manually updating tables or graphs. Secondly, every graph and table in the document is traceable to the code that produced it through examination of the `Rmd` markdown file, meaning the analysis is transparent.

There are a number of software package available for the fitting of Gaussian Processes. The most comprehensive is probably gpml for Matlab, which includes a variety of likelihoods, link functions, and inference methods as standard. The `BACCO` (Bayesian Analysis of Computer Code Output) package for `R` provides much of the functionality detailed by the MUCM (2011) project, although mainly for deterministic simulators (Hankin, 2005). The DiceKriging (Roustant et al., 2012) package is an excellent set of tools specific to simulation analysis, including stochastic simulations. GEM-SA is a standalone Windows application written in `fortran` by Kennedy (2004) and is available from Professor O'Hagan's website. Finally, `tgp` is an `R` package for the fitting of treed Gaussian process. These use Bayesian model averaging and treed regression structures to fit different models across different dimensions and different areas of the input space, but it can also be used to fit the simpler (non-treed) Gaussian Processes used throughout this thesis (Gramacy, 2007).

Despite the availability of several software packages for the fitting of Gaussian Processes, such a package was not used to conduct the work in the following chapters. This is mainly because none of them provide exactly what is required, and for those with accessible source code, extending them was not really plausible. Instead, code to fit Gaussian Processes was written from scratch in R, although both `BACCO` and `gpml` where sources of inspiration and aids to understanding, and in particular the fast algorithm for computing covariance matrices used by Hankin (2005) was crucial. The code for all analyses conducted is available on the author's bitbucket page, together with the rmarkdown code and the python libraries relating to the simulation in Chapter 6.

2

Agent-Based Modelling: Literature Review

2.1 Context of the Analysis of Agent-Based Modelling in Demography

To determine how best to go about analysing and calibrating an Agent-Based Model, it is important to have a clear idea of what scientific purpose this model serves (Di Paolo et al., 2000; Rossiter et al., 2010). In various guises, ABMs have been employed to *explore* consequences of abstract assumptions, to *explain* empirical regularities, and to *predict* possible futures (see Macy & Willer, 2002; Squazzoni, 2012; Gilbert & Troitzsch, 2005; Gilbert, 2008; Kirman, 2011; Aparicio Diaz et al., 2011). Furthermore, Agent-Based Modelling has been applied across disciplines as diverse as ecology (Grimm et al., 2005), economics (Kirman & Vriend, 2001), and archaeology (Axtell et al., 2002), and individual simulations vary considerably in the degree of complicatedness; theoretical grounding; empirical relevance; and use of data, to name just some dimensions of difference (Rossiter et al., 2010; Squazzoni, 2012).

There is no reason to expect that usages in demography will be any more uniform; indeed, the overview of existing agent-based work in demography in Section 2.4 pays testament to the diversity of existing applications relating to population. Thus, any recommendations made about analysis and calibration methods are necessarily conditional on the types of model to be analysed. The next Section (2.2) provides a discussion of scientific justification for the use of ABM, and the implications for analytic approaches. In later chapters, it is argued with the aid of examples that Gaussian Process Emulators may be helpful in demographic applications both in theoretical and more applied contexts, although not to the exclusion of other approaches. It is also necessary to provide a discussion of the existing approaches to the calibration and analysis of agent-based simulation; this follows in Section 2.3.

2.2 Scientific Value of Agent-Based Models

Agent-Based Simulation is often seen as a fundamentally different way of approaching social science (Epstein & Axtell, 1996; Bonabeau, 2002; Werker & Brenner, 2004). Simulation models attempt to recreate simplified versions of some elements of reality within a computer program (Gilbert & Troitzsch, 2005). The question then becomes: how can these programs create new knowledge about the real world (Di Paolo et al., 2000; Rossiter et al., 2010)?

In the natural sciences, and in many areas of empirical social science, scientific progress occurs through deduction (Popper, 1959). A hypothesis is generated, and the logical implications of this hypothesis are subjected to empirical tests, that is, through experiment or other targeted collection of empirical data (ibid). The alternative classical inference method is induction, which is characterised by generalisation from empirical data to theories (Werker & Brenner, 2004). Squazzoni (2012) suggests that Agent-Based Models often include elements of both these approaches: as with deduction, the starting point is usually some theory or set of assumptions which form the logical basis for the simulation model. However, the model is then used to generate *synthetic data*, which is then analysed inductively in order to understand how the macro-behaviour of the model follows from the assumptions (Grazzini et al., 2013b).

Within this framework, the *model*, rather than reality, becomes the immediate object of study (Franck, 2002). Modelling is required because of the difficulty of studying the phenomena directly. Complex phenomena, which are generally the focus of Agent-Based Models, have particular characteristics that often necessitate the use of *simulation* modelling (Bonabeau, 2002). By definition, complex systems are those which display *emergent* behaviour at the macro-level, which is generated from the interaction of many units at the micro-level, but is not a trivial consequence of these interactions and can not easily be inferred from knowledge of the rules that govern them (Mitchell, 2009). In particular, analytical solutions to such systems are difficult; the assumptions that allow for tractable solutions in other models (such as agent homogeneity and independent action), are not available here (Leombruni & Richiardi, 2005). As well as suggesting the use of simulation, this also means that the behaviour of any constructed simulation will itself require careful examination (Di Paolo et al., 2000).

2.2.1 Agent-Based Models for Exploring Theories

In some cases, the aim of model-building is merely to understand the consequences of some assumptions or theory, and to examine its coherence and internal consistency through the process of formalisation that constructing a simulation entails (Squazzoni, 2012). For example, it might be theorised that X acting according to some set of rules A is likely to result in some outcome Y; building a simulation that formalises this notion allows testing of this posited relationship and the conditions under which it holds (Epstein & Axtell, 1996; Macy & Willer, 2002). Rossiter et al. (2010) refer to this process as examining the *analytical adequacy* of the model. In this case, there may be no specific real-world phenomena or class of phenomena which the simulation aims to represent exactly; instead, the model can help re-organise and develop the theoretical and conceptual machinery of its author (Di Paolo et al., 2000). Di Paolo et al. (2000) suggest that investigations in this mode can be considered as “opaque thought experiments”. Simulations of complex behaviour, they argue, perform the same role as traditional thought experiments, in that they use the logical consequences of a set of assumptions to show their coherence or incoherence, but unlike thought experiments, they are opaque: the complexity of the system involved means that these consequences do not follow trivially from the premises, and thus simulations are necessary to uncover these relationships.

It is also important to realise that the investigation of simulation results can result in the development of new theory; if the simulation results reveal the inadequacy of the original theory, it may also imply new explanations or consequences of the subject of study (Di Paolo et al., 2000). This suggests that the process is often iterative; theory suggests a model, which implies adjustment of the theory, which in turn highlights new areas for modelling (Cioffi-Revilla, 2010; Squazzoni, 2012).

Agent-Based Models constructed with these purposes in mind tend to be as simple as possible while still encoding the theory they aim to examine (Epstein & Axtell, 1996; Chwif et al., 2000). This increases the ease with which simulation results can be interpreted; the complex mechanisms involved may make this process of interpretation and analysis difficult in any case, so needless complication should be avoided (Macy & Willer, 2002; Carley, 2002; Chwif et al., 2000). This is often referred to as the KISS principle: "Keep it Simple, Stupid", and Axelrod sums up the approach through the suggestion that "The complexity of agent-based modelling should be in the simulated results, not in the assumptions of the model" (Axelrod, 2003, p6).

2.2.2 Agent-Based Models for Identifying Plausible Explanatory Mechanisms

Early work in Agent-Based Modelling focused on this type of abstract, exploratory investigation (e.g. Epstein & Axtell, 1996). More recently, much Agent-Based work instead seeks the particular micro-foundations of specific empirical phenomena or classes of phenomena (Boero & Squazzoni, 2005). In this case, we need not only to explain the workings of the model itself, but also to define its relationship to reality. Thus, the aim is to attempt to identify the micro-rules A that generated some *observed* pattern or class of patterns Z. Werker & Brenner (2004) describe this mode of inference as *Abductive*, following the work of American Logician Charles Sanders Peirce. The key to this approach is that a simulation that generates synthetic data Y which are close to the observed data Z (by some definition) operates according to a set of micro-rules which is a plausible candidate for acceptance as the true data generating mechanism (Squazzoni, 2012). A similar approach is proposed by Silverman et al. (2014) : in their proposed functional-mechanistic approach, observation of reality allows inference about the 'formal structure' of the system under study [1], and Agent-Based Modelling allows the testing of candidates for the social mechanisms that perform this function (see also Courgeau, 2012).

The plausibility of an explanation of a given process, then, is determined by comparison of model outputs against that which is observed in reality, a process termed of

[1] A 'formal structure' in this context is taken to mean a representation of the system as a functional transformation of inputs or starting conditions to outputs or termination points; this corresponds with a description of the nature of mechanisms in Casini et al. (2010).

empirical validation (Werker & Brenner, 2004). Agent-Based Models can thus provide 'existence proofs' about what conditions are necessary to generate a particular emergent pattern (Marks, 2012). However, a major barrier to the production of knowledge using Agent-Based Models is the problem of equifinality or under-determination; that is, the flexibility in specification of the model leaves a large number of degrees of freedom, and any one empirical pattern could have been generated by a possibly infinite number of potential model specifications (Squazzoni, 2012). It is therefore easy to prove that some conditions are sufficient to generate a particular structure, but all but impossible to prove that they are necessary (Marks, 2012). However, as Leombruni & Richiardi (2005) claim, it is preferable to provide an under-determined model that is a reasonable description of the phenomena in question than an identified one based on unreasonable simplifying assumptions. Further discussion of strategies for increasing confidence that the modelled process is the one that holds in reality are discussed in Section 2.3.

The process of comparison may be aided by the calibration of the model. An Agent-Based simulation will generally have collections of parameters that determine the model's behaviour. Some of these may be fixed or constrained to small ranges with reference to empirical data (Brenner & Werker, 2007). Others, however, may be free to vary, and the calibration problem is to find the values of these parameters that provide the closest fit to the observed pattern (Boero & Squazzoni, 2005). However, the discussion of the previous paragraph should be borne in mind here: just because a model has been calibrated, this does not mean that it represents *the* data generating mechanism that generated the observed patterns; rather it allows a better judgement of whether it is indeed a plausible candidate or not.

With models focused on specific empirical entities, we begin to see a tension between the desire of modellers to keep their models simple, and the imperative to match reality as closely as possible (Carley, 2002). Simplicity is desired for two reasons. Firstly, it aids interpretation; part of the reason for modelling is to abstract away from the incomprehensible complicatedness of the real world phenomena, and so including numerous elements and moving parts is somewhat self-defeating (Chwif et al., 2000). Secondly, it limits the potential sources of error. Specifying individual sub-processes within a simulation is difficult and painstaking, and inevitably involves simplifications and approximations. These mis-specifications may or may not be ruinous to developing an explanation of the phenomena in question, but identifying whether this is the case is more difficult if one does not know which element out of many is 'wrong' (that is, causing a model to fail empirical validation tests) (c.f. Galán & Izquierdo, 2009).

However, not all modellers share this viewpoint. Edmonds & Moss (2004) claim that the burden lies with those who wish to make their model simpler than a specific observed target to show why this simplification is justifiable or necessary for practical purposes (computational cost, coding time, etc.). While they raise an important point, one counter-argument is that by including every element that could effect the process of interest we lose the ability to generalise from our model; any information gained is specific to the present context. This might be acceptable, however, if we wish to use the model for prediction.

2.2.3 Agent-Based Models for Prediction

Agent-Based Simulation has had limited success in attempting to predict future states of target social systems. This is clearly a difficult problem; unlike in natural systems, in many cases we cannot assume that the parameters governing a social system are consistent across time (Moss & Edmonds, 2005). Without this continuity, prediction becomes difficult. In social science, of course, even explanation of a given phenomenon is a tough prospect, and theories about the underlying processes governing social systems abound (Edmonds & Moss, 2004; Squazzoni, 2012). For this reason, many social simulators are happy to restrict themselves to explanatory goals (Epstein, 2008). However, it seems reasonable to suggest that if we are able to describe the underlying micro-foundations of social processes, this may increase our ability to predict them.

Part of this argument depends on our definition of 'predict'. Troitzsch (2009), in response to the Epstein (2008) article defending non-predictive modelling, identifies at least three different types of prediction:

1. Predictions of kinds of possible behaviour of a given system, for arbitrary parameters and initial conditions.
2. Predictions of what *kind* of behaviour a given system will display in the near future
3. Predictions of the *state* of a target system in the near future.

The latter category, Troitzsch (2009) states, includes both point and probabilistic predictions, although the former would seem to be out of reach for all but the most trivial social systems, and provides particular problems because of the difficulty in specifying initial conditions when these involve agents' states of mind at a particular point in time (ibid). Agent-based models, assuming they can be considered 'structurally valid' (that is, that they are a true reflection of the reality of the underlying system), provide adequate tools for the first two types of prediction (ibid). It is to be hoped that in demography they might be able to contribute to making probabilistic projections of population, given demography's position as one of the more predictable social sciences (Bijak, 2011; Silverman et al., 2011). However, the problem that has plagued all 'structural' prediction models in demography still remains, even if the underlying micro-foundations can be specified: individual behaviour will still depend on other factors (e.g. the economy) which are even more difficult to predict (Booth, 2006).

2.2.4 Analytical requirements for Agent-Based Models

From this discussion, however, we can differentiate two specific types of problem relating to drawing inferences from Agent-Based Models. Firstly, there is the challenge of uncovering the specific relationships in the model itself, the process of *explaining* the behaviour of the model through analysis. Secondly, the relationship between the model and reality must be made clear; this involves both calibration and model validation. Any given approach to the analysis and calibration of ABM is unlikely to cover all of these diverse purposes to which agent-based models are put. A number of frameworks have been suggested, with many of the differences between these being related to the nature of the disciplines from which they emerged.

2.3 Approaches to analysis and calibration of Agent-Based Models

Historically, the fragmented and nascent nature of the ABM approach had led to a rather ad-hoc approach to the analysis of simulations outputs (Grazzini & Richiardi, 2013). However, in certain sub-spheres of the community there are moves towards a more unified and consolidated approach (e.g. Grimm et al., 2005; Werker & Brenner, 2004; Grazzini et al., 2013b).

2.3.1 Pattern-Oriented Modelling in Ecology

Within Ecology, a consensus has arisen around a general methodology named Pattern-Orientated Modelling (Grimm et al., 2005). The term ‘pattern’ is invoked because often ecological Individual-Based Models (IBMs, the equivalent term for an ABM in the discipline) include spatial elements, and hence the outcome involves patterns over geographical locations. Pattern Oriented Modelling helps to increase confidence in model validity because it advocates matching against multiple patterns at different hierarchical scales (that is, at both the individual and the population level), thus increasing the number of constraints on the system (ibid). Furthermore, it also encourages the testing of multiple theories of behaviour, including ‘null’ models and nonsensical models, in order to understand the extent to which model success in matching reality is down to a specific behavioural theory, and how much is inevitable given other arbitrary or environmental factors (Grimm & Railsback, 2005). Pattern Oriented Modelling also advocates matching against patterns during construction of the model, in order to help decide upon its structure. In this context, model structure involves specification of which elements of the real system to include, and which to exclude. Grimm et al. (2005) also advocate the use of pattern matching to *calibrate* the model. A calibrated model involves finding the set of parameter values such that outputs for these values match all of the observed patterns (where such a set exists).

However, overall, Pattern Oriented Modelling represents more of a set of recommendations about the processes involved in producing an Individual-Based Model than a strict methodology. This is perhaps sensible, given the diversity of model types it intends to cover. The methodology section of the textbook by Grimm & Railsback (2005, Chapter 9) is ambivalent about what should constitute a match between patterns; to some extent this is left to the judgement of the modeller. In most cases, qualitative or visual matches are suggested to be sufficient. For larger and more empirically focused simulations, Grimm & Railsback (2005) do suggest the use of more sophisticated experimental design and statistical analyses similar to those discussed in the later chapters of this thesis, but they do so with some reticence. For example, with respect to the Uncertainty and Sensitivity Analysis techniques of Saltelli et al. (2004), they (Grimm & Railsback, 2005) state:

> Although SA [Sensitivity Analysis] and UA [Uncertainty Analysis] can help us understand an IBM, they are a reversion to “black box” modelling: we look at what goes in and what comes out without trying to understand what is going inside the IBM.

This is reasonable criticism, but using statistical techniques to understand the relationships in a model does not preclude deeper investigations to understand its mechanistic workings, and indeed it may make the process easier and less time consuming by giving indications of what the modeller should be looking for. Statistical techniques for dealing with model uncertainty are applied in ecological Individual-Based Modelling, however, although not widely (Grimm & Railsback, 2005). Wiegand et al. (2004), for instance, conduct a sensitivity analysis on their model of bear populations in order to identify the most important inputs by comparing average standardised linear regression coefficients at randomly sampled points across the parameter space. They also calibrate the model using Monte-Carlo filtering (Saltelli et al., 2004). This consists of defining acceptable criteria for the acceptance of a simulation run, and accepting only those sampled parameters for which outputs meet the criteria (ibid).

2.3.2 Analysis of Agent Based Modelling in Economics

In Agent-Based Computational Economics, a slightly different approach is taken. Given the analytical rigour and methodological prescriptiveness of the discipline, Agent-Based Modellers have to work harder to gain acceptance amongst their peers (Leombruni & Richiardi, 2005). This has led to an interesting approach to the justification of the use of Agent-Based Models from Leombruni & Richiardi (2005). Here, ABMs are considered as a specific type of recursive difference equation; they are to be understood in exactly the same way as other (economic, in this context) mathematical models, except that their complexity is such that the shape of the relevant functions are unknown and they must be solved numerically. There is thus no need to consider them in any different manner to their more tractable cousins.

In this spirit, Grazzini & Richiardi (2013) propose that Agent-Based Economic modellers borrow estimation tools used for Macro-Economic models. They suggest that Agent-Based Models differ from Dynamic Stochastic General Equilibrium (DSGE)[2] models only in the fact that with DSGE models, "(some) properties of the steady state can be studied analytically" (Grazzini & Richiardi, 2013, pp19), whereas with Agent-Based models these must be inferred from simulated outputs. Thus, they advocate methods used to calibrate DSGEs can serve the same purpose for ABMs, and in particular they suggest Minimum Simulated Distance methods. As the name suggests, these compute summaries of the time series of both the simulated and the real data, and minimise the distance between these by optimising with respect to the unknown parameters. The summaries can be the relevant moments of the data (in which case the technique is referred to as the Method of Simulated Moments), or alternatively, an arbitrary meta-model (a statistical model of the model) can be fitted to both sets of data, and a relevant coefficient chosen as the summary (termed 'indirect inference'). Because the same meta-model is fitted to the simulation as to reality, this approach is robust to mis-specification (ibid). A necessary precondition to these approaches is to ensure that the chosen output is stationary (in that its properties are constant in time) and ergodic (in that different initial condition lead to the same equilibrium) (Grazzini, 2012).

Other approaches are also based around minimisation of some metric measuring the distance between the simulated and observed distributions. Marks (2011), for example, examines three potential distance metrics in order to determine the similarity between two time series of brand price data, one simulated, one obtained from real measurements.

In a more heterodox economic approach, Werker & Brenner (2004) advocate using (different) empirical data in two parts of the modelling process; firstly, to parametrise and inform some of the micro-level behaviour of the model; and secondly to identify the areas of the remaining parameter space which are consistent with the observed data through a process of abduction (as was defined above). In Werker & Brenner (2004), they advocate Bayesian methods to accomplish this second goal. However, they are not specific about how exactly this should be done, and they explicitly reject the possibility of assigning probabilities to different models, so there is some ambiguity in their suggestions. In the later paper of Brenner & Werker (2007), they instead seem to advocate a process of Monte Carlo filtering as used by Wiegand et al. (2004), although they do not refer to it as such.

[2]Dynamic Stochastic General Equilibrium models are macro-economic models that use sets of differential equations to forecast key economic variables (e.g. Smets & Wouters, 2002).

2.3.3 Other approaches

Other approaches to calibration and analysis exist outside ecology and economics. One particularly interesting general approach focuses on the analysis of simulation models with very varied local behaviour across the parameter space (Luke, 2007). Using an artificial test function Luke uses a genetic algorithm to find those parts of the space where the output gradient is highest; for the Agent Based modeller, this provides a useful way of identifying phase transitions, which in many cases is where interesting model behaviour lies (Bijak et al., 2013). Genetic algorithms are also used by Marks (2011) to attempt to find the optimal strategy for agents playing a lottery.

In contrast, Brown et al. (2005) describes an attempt to validate a model of land use using a method similar to a pattern oriented approach advocated by Grimm et al. (2005). He divides his spatial model into areas which are invariant to replications, and those for which the outcome is path dependent, thus paying some attention to the uncertainties in the model during the analysis process.

Meta-models often provide a way to deal with model uncertainty. Kamiński (2015) utilises meta-models to analyse the classic Schelling segregation ABM. Specifically, he utilises 'interval meta-models', which assume a simple regression model but return intervals rather than point predictions in order to ameliorate bias induced by the simplicity of the model. He then uses Bayesian techniques to investigate the probability that the intervals include the simulation outputs, thus accounting for uncertainty in the meta-model. This anticipates the emulator/meta-model approach explored in the remainder of this thesis. A small number of authors have already explored this direction by applying Gaussian process emulation techniques to Agent-Based Simulations (Boukouvalas et al., 2014; Heard, 2014; Andrianakis et al., 2015). These approaches are discussed in the next chapter (3), which deals specifically with emulator-based approaches.

2.4 Demographic Agent Based Models

2.4.1 Microsimulation in Demography

Agent-Based Modelling in demography has long antecedents in the form of microsimulation; as with ABM, microsimulation simulate individuals, but does so through the use of transition rates rather than behavioural rules (Gilbert & Troitzsch, 2005). A review of demographic microsimulation methods is offered in Van Imhoff & Post (1998). The main purpose of such methods is generally to estimate current or predict future states for policy purposes, so the approach is much more geared towards accuracy rather than explanatory power, and is rather more data driven. The collection edited by Zaidi et al. (2009) provide many examples of the use of microsimulation for these purposes. Despite this, at the margins of the methodology, some theoretical investigations have been conducted that are quite close to Agent-Based Modelling in spirit. Hammel et al. (1979) used a microsimulation with behavioural augmentations to investigate what rules regarding the prohibition of incest would have enabled continued survival of isolated small breeding groups in human prehistory, while avoiding debilitating mutations. Similarly, Murphy (2006) examined the inclusion of the effects of correlations in fertility between parents and children on estimates of population size. Both use microsimulations as pseudo-experiments to attempt to investigate theoretical questions. Zinn (2011) also included behavioural elements in order to choose partners for simulated individuals, a problem not easily decidable by reference to data.

Microsimulations do not generally, however, display the kinds of emergent behaviour we associate with Agent-Based Models, and as a result, the problem of explaining their behaviour is not really an issue. It is an obvious consequence of the model setup. In terms of calibration, the main technique employed is a process of alignment whereby outputs are constrained to macro-level forecasts (Spielauer, 2007). This top-down calibration does not reveal anything about the micro-dynamics of the system however, and shows a mistrust of the micro-level which goes against the grain of the generative ABM approach (Epstein & Axtell, 1996).

2.4.2 Demographic Agent-Based Modelling

The sugarscape models of Epstein & Axtell (1996), a series of experiments on small scale 'artificial societies' that kick-started Agent-Based modelling in social science, include many demographic elements. Agents in their simulations gave birth, moved from one area of the world to another, and died. However, these investigations were not aimed at demographic questions as such, and Agent-Based Modelling in demography proper can be traced to the edited volume of Billari et al. (2003). This collection, and that which followed soon after (Billari et al., 2006), advocated Agent-Based Modelling as a way of reconciling traditional macro-demography and the micro-demography that has flourished in recent decades.

Broadly speaking, agent-based contributions to demography can be divided into two groups. The first consists mainly of investigations of fertility and nuptiality and focuses on the effects of social pressure and social influence amongst peers (see Bernardi, 2003, for a discussion of these concepts in demography). Those in the second group focus on migration, and mainly use agent-based approaches in order to include the effect of information transfer across networks (see Massey, 1990; Munshi, 2003).

2.4.3 Social Influence and Social Pressure

Agent-based models of social interactions in demography include the Wedding Ring model of Billari et al. (2007), which is analysed in greater detail in Chapter 4. This investigates the effect of social pressure on agents' partner search intensity, and concludes that this mechanism can reproduce the shape of the observed marriage hazard curve. Billari et al. (2007) specify a benchmark parameter set for which this finding holds, but do not elucidate how this parameter set was found. They also attempt a simple scenario-based sensitivity analysis by examining results for a small number of alternative model specifications.

Similarly, Aparicio Diaz et al. (2011) examine the effect of social network pressure on transition to first birth in Austria. They are able to replicate changes in the timing of first birth seen in Austria in the 2000s. A metric measuring the distance between simulated and real Age-Specific Fertility Rates is used to assess model performance. A more systematic sensitivity analysis is attempted here, with combinations of parameters evaluated over a grid, and a 'null' model in which the network effects are turned off is also investigated. Presumably, the stated default parameters are those on the grid for which this distance is minimised. However, because of the practice of varying two parameters while keeping the others fixed, large areas of the parameter space were left unexplored. This study is notable, however, for its attempt to forecast future values of the system in question.

Fent et al. (2013) built on this work by examining the effect of social networks on the success of family policies. They modelled preferred family size as dependent on opinions of social network members. This time, social network growth is endogenous and dependent on agent similarity, degree of relatedness, and number of shared network contacts. In Fent et al. (2013), a grid search over 6 parameters leads to a total of 741,312 simulations. With efficient simulation designs such as those discussed in the next chapter, a much smaller number of simulations would have resulted a greater understanding of the dynamics of the simulation. More recently, Kashyap & Villavicencio (2016) examine sex-selective abortion within the analytical framework of 'ready, willing, and able' (Coale, 1973), showing the dynamic relationship between son preference, social pressure relating to family size, and diffusion of abortion technologies. Most interestingly, they showed that increasing sex ratios at birth might occur in the context of declining son preference, because of the additional pressures derived from smaller family size norms. Kashyap & Villavicencio (2016) calibrate their model by fitting a regression meta-model to a Latin-hypercube sample of points (defined in the next section), and then use these predictions to minimise the predicted root-mean-squared error of their model relative to observations.

2.4.4 Migrant networks

A number of authors have used Agent-Based Models to investigate the effects of network contacts on migration. Rehm (2012) simulated migration from Ecuador to the USA and investigated the effect of networks in encouraging contacts of recent migrants to also migrate. Rehm (2012) fixed parameters only on the criteria that they should result in a 'non-corner' solution, where not everyone migrates or stays at home. A stability analysis was conducted on a simplified, differential equation version of her simulation model; linearisations of these equations around many parameter values were examined to understand the local behaviour of the model. Rehm (2012) also used her model to test the plausibility of a number of different theories as to the causes of migrant return migration and remittance behaviour, and concludes that no one theory fits the data well, but a combination might form a plausible explanation for observed behaviour.

Klabunde (2013) examined circular migration between Mexico and the USA, modelling agents as deciding whether or not to migrate using a discrete choice framework. Following Werker & Brenner (2004), she fixed some parameters of the model using the Mexican Migration project data, whilst others were based on behaviour rules, the parameters for which were found through a grid search of possible combinations, with a goodness of fit metric forming the criterion of choice. Willekens (2016) also investigated the migrant decision within an agent-based framework, recreating observed features of the migratory flow between Africa and Europe using the Theory of Planned Behaviour (Ajzen, 1991) as a framework for a multistage agent decision process. Inputs to the model were largely estimated from observed data, with other values given 'best guess' values without formal calibration procedures. Kniveton et al. (2011) investigated climate related migration from Burkina Faso, and produced forecasts of the predicted migratory response to climate change under various possible climate change scenarios. This model was also based on theory of planned behaviour, and migration rates were based on those empirically estimated, with deviations from these induced by various behavioural inputs, such as knowledge of others living abroad.

2.4.5 Other Demographic Agent-Based Models

Outside these two main groupings of demographic agent-based modelling, a number of other models have been constructed. Geard & McCaw (2013) constructed an empirically focused individual-based simulation, which could be said to be closer in spirit to a microsimulation, with the aim of investigating future household dynamics. In contrast, Hills & Todd (2008) modelled partnership formation as a search and match process where agents aim to find agents similar to themselves, but relaxed their criteria as to what constitute a good match over time. They examined the theoretical consequences of greater population heterogeneity and cultural diversity on the matching process, in a typically ABM-like theoretical investigation, but also compared simulation results to empirical data. They were not only able to match US marriage curves, but also recreated the divorce rate using the same parameter settings as an attempt at model variation. Also looking at marriage, but focusing on phenomena of assortative mating, Grow & Van Bavel (2014) develop an agent-based model examining how increased educational attainment amongst women has led to changes in the marriage market. The simulation is calibrated against empirical data using regression metamodels, and validated against a hold-back set of data from other countries.

2.4.6 Summary of Demographic literature

Overall the demographic literature on Agent-Based Models is growing quickly and more contributions are forthcoming. At present, approaches to calibration and analysis are ad-hoc, with grid searches being the most prevalent, and although more recent studies have adopted more principled approaches to the calibration problem (e.g. Grow & Van Bavel, 2014; Grow, 2016), there remain areas for improvement, particularly with respect to uncertainty quantification. The next chapter of this thesis (3) examines one possible alternative approach to the analysis of demographic Agent-Based Models.

3

Gaussian Process Emulators

3.1 Design of Computer Experiments

The previous chapter (2) provided a brief overview of some of the analysis and calibration techniques used in the field of Agent-Based Modelling. Although some authors, particularly in Ecology (Grimm et al., 2005) and Economics (Grazzini et al., 2013a), have attempted to apply systematic approaches, many ABMs are not analysed adequately (Boero & Squazzoni, 2005; Grazzini et al., 2013a), and the use of more formal tools that account for the uncertainty associated with these models could be beneficial. This chapter outlines some of the design, analysis and calibration techniques that could be useful in this respect.

Epstein and Axtell's seminal book on 'Growing Artificial Societies' (1996) famously considered Agent-Based Models as an analogue of physical experiments for social scientists. Modellers, they state, could grow experimental scenarios *in silico*, enabling them to examine the effects of manipulating various inputs. To take this claim seriously, and to maximise what we learn from our quasi-experiments, borrowing from the literature on the design of experiments can be instructive, as at present most approaches to design of ABM are ad-hoc, use grids or employ optimisation-based methods (Chapter 2).

Computational experiments differ from physical experiments in several important respects (Santner et al., 2003). Firstly, computational experiments tend to be cheaper, and so can be run more times and at more points. Secondly, the modeller has complete control of the experimental conditions, and results are therefore not subject to unobserved nuisance factors that may cloud inference. Thirdly, greater freedom is possible in the specification of the experiments to be run.

One consequence of this freedom, however, is that simulations tend to have a greater number of free parameters - that is, inputs to simulation which are not fixed by theoretical considerations or empirical measurements - than their physical equivalents (Santner et al., 2003). This often means that a large number of runs are required to get a handle on the behaviour of the simulation over the entire parameter space (Grimm et al., 2005). This is even more the case if the simulation is also stochastic, in the sense that repetitions of a simulation run at the same parameter values will give a different outcome.

3.1.1 Factorial Designs

At this point it is helpful to introduce some notation. Generally, we consider an experiment, whether simulated or physical, as a mapping of some input x to output y. We denote this mapping as a function $f(.)$. Generally x will be multidimensional, in that there are k parameters to the model $x = \{x_1, x_2, \ldots, x_k\}$ and we restrict attention here to real inputs, so that $X \in \mathbb{R}^k$. The design problem is to choose a vector of input points D so that we can learn as much as possible about how $f(.)$ responds to x.

The larger the number of dimensions, the harder this problem becomes. The instinctive response is simply to pick a set of equally spaced points (levels) for each parameter, and run at all combinations of these parameters. This is known as a full factorial design (Santner et al., 2003). However, even for relatively small numbers of levels and parameters, this can quickly become prohibitively time consuming, unless the simulation is extremely fast.

Consider as an example a simulation with 5 input parameters ($k = 5$), each of which we want to run at 5 levels. We then need a total 5^5, or 3125 runs. If our simulation takes only one minute to run, this would require a total of 52 hours of runtime, or some multiple of this number if we wish to repeat observations at each point. Of course, computing power is relatively cheap at present, and many will have access to multi-processor cluster computing resources that can easily reduce this time to a few hours. However, as ABMs begin to simulate more agents and involve more complicated decision making, run-times are likely to increase. As a result, it makes sense to consider more efficient experimental designs.

3.1.2 Latin Hyper-cube Sample Designs

The key limitation with grid designs is that when projected or 'collapsed' onto one dimension, many design points are replicated, and thus wasted (Urban & Fricker, 2010). Latin Hyper-cube Samples (LHS) avoid these problems by ensuring no two design points share values of any parameter (idem). The principle is simple - to create a LHS, divide the input space equally into g bins along each axis, so that there are g^k cells in total. Then, choose g of these cells such that there is only one cell in each chunk (or column) for every axis. To complete the process pick a point randomly within each chosen cell, resulting in a sample of size g (Santner et al., 2003). Figure 3.1 contrasts grid and LHS designs in two dimensions. Note that the LHS design adequately fills the space to be explored, but uses only two-thirds of the number of points. Furthermore, many more values of each parameter are evaluated by the LHS design, as can be seen by the tick-marks across the axes.

A desired property of simulation design is that they be space-filling; that is, that they provide close to optimal coverage of all areas of the parameter space, so that no area is left without a design point (O'Hagan, 2006). Latin Hyper-Cube Samples are not guaranteed to be space-filling, so some further criteria are needed to ensure that all parameter combinations are explored adequately. Generating several candidate samples, and picking the one with the highest minimum distance between points will in general suffice (O'Hagan, 2006). The r package `lhs` has a number of functions for producing such samples for arbitrary dimensions very easily. These will then need to be scaled up from the existing $[0, 1]$ range to reflect the input ranges required by any simulation.

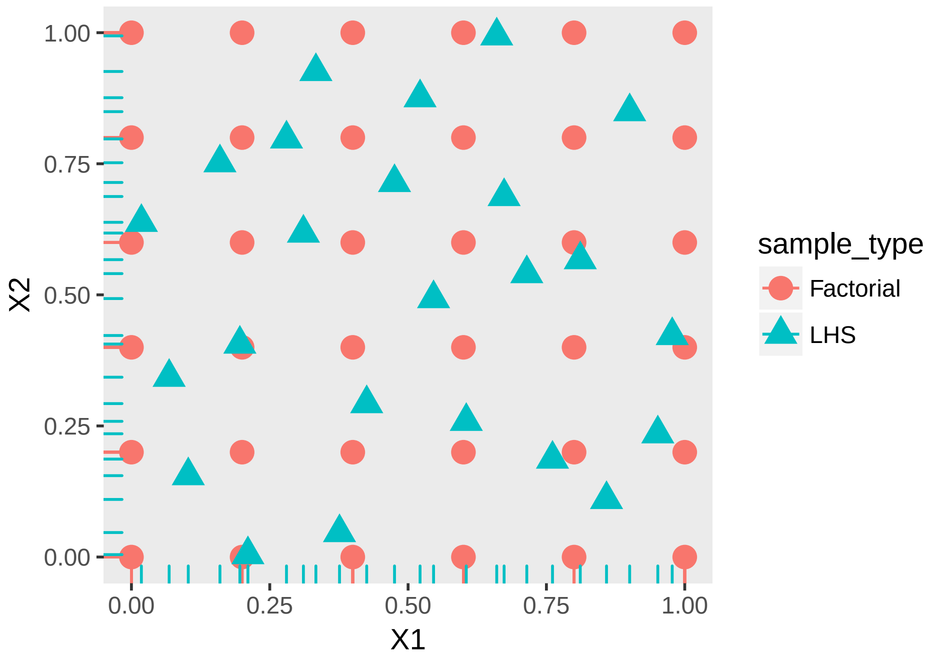

FIGURE 3.1: Comparison of full factorial and latin hyper-cube sample design

The advantage of LHS designs, however, is the scaling in high dimensions. For the example above, rather than requiring thousands of simulations to explore a 5-dimensional space, samples of the order of a few hundred may be sufficient. Other advantages of the LHS related to their optimality for analysis are discussed in Santner et al. (2003). Repetition of simulations at individual design points may also be desirable in order to account for stochasticity in simulation outputs. Boukouvalas (2010) provides a good summary of these issues.

3.2 Gaussian Process Emulators for Agent Based Models

Once a design has been settled on, and the simulation has been run at these points, the difficult job of analysing the simulation results must begin. In high dimensions, understanding the relationships between inputs and outputs simply from the raw results is often difficult, particular as they are typically non-linear and subject to a variety of feedback effects. This is particularly the case with Agent-Based Models; because such simulations are by definition caused by the interaction of many autonomous units, the system as a whole can be defined as *complex*. As discussed in Chapter 2, in this context the word has a technical meaning, and complex system tend to be characterised by tipping points, non-linearities, and other such features (Epstein & Axtell, 1996; Mitchell, 2009).

From a statistical point of view, modellers must also ensure that they account for the various sources of uncertainty inherent in the analysis of simulations (Kennedy &

O'Hagan, 2001a). Following Kennedy & O'Hagan (2001a), such sources of such uncertainty include:

- **Uncertainty due to simulation stochasticity**. This occurs when running the code twice at the same parameters gives different results, because of the use of random number generation in the simulator itself. This means even if we have the result of one trial at a point, we can't predict with certainty the value of another such trial. Random number generators are used to represent *aleatory* uncertainty in the real world phenomena; that is, uncertainty due to inherent randomness (O'Hagan, 2006).
- **Uncertainty about the output at points not yet run**. In a continuous parameter space, we can never run the simulation at every conceivable point. Instead, we must estimate at points we have yet to run, which we do with some error.
- **Input Uncertainty**. In many cases, we do not know what the 'correct' value of any given parameter is. We may, however, have a reasonable range or probability distribution that characterises our beliefs as to where the true value lies. This uncertainty about inputs clouds our knowledge about outputs.
- **Model discrepancy**. The model is unlikely to be a perfect representation of reality. The mechanisms simulated will differ from what takes place in the world in appreciable but uncertain ways. Thus, our lack of knowledge about the ways and extent to which our model is wrong is another source of uncertainty.
- **Measurement error**. Comparing simulated results to real results may add an additional source of uncertainty, as real world quantities are subject to errors in measurements (Kennedy & O'Hagan, 2001a).

The last four sources of uncertainty are largely *epistemic*, in that the uncertainty is the result of our lack of knowledge about the quantities of interest (although there is an argument to be made that the latter is partly also aleatory) (O'Hagan & Oakley, 2004). Failing to take these sources of uncertainty, whether aleatory or epistemic, into account can lead to faulty inferences (O'Hagan, 2006). This can be particularly problematic if policy advice is the goal of the simulation; representing uncertainty about the phenomena in question is vital if potential risks are to be mitigated (Bijak, 2011).

3.2.1 Alternatives to Gaussian Process Emulators

Although this thesis will focus on the use of Gaussian process emulators, there are a number of ways in which the statistical analysis and calibration of simulation outputs can be approached. Firstly, a 'brute-force' Monte-Carlo approach can be considered. Sampling from distributions representing the above sources of uncertainty many times and obtaining simulation results for each sample would allow for a coherent accounting. However, this requires *many* replications, and so has computation time implications.

Several other approaches to the problem have been proposed. Bayesian melding (Poole & Raftery, 2000) involves reconciling prior knowledge about outputs and inputs with observations and simulation outputs. Simulation inputs are generated from prior distributions, and outputs are obtained for these inputs by running the simulation; this induces a prior on the output $f(\theta)$. This prior is combined with an independent existing prior on that output variable Y using logarithmic pooling, and then importance weights for sampling from the posterior of θ are created by combining this induced prior with any likelihood data for θ or Y using Bayes' rule. This technique is robust in its accounting for most sources of uncertainty, and incorporating prior knowledge, but also requires many replications to build up posterior distributions. Ševčíková et al. (2007a) used this approach successfully in an stochastic urban simulation that modelled household behaviour, and thus was in essence an Agent-Based Model.

Approximate Bayesian Computation (ABC) can also be considered as an approach to the calibration problem. ABC is a 'likelihood-free' method that, when applied to computer experiments in its simplest form, runs simulations at inputs drawn from a prior distribution, and then approximates the posterior distribution by rejecting points that fall outside some distance from the observed data (Sunnåker et al., 2013). As with Bayesian melding, a large number of simulation runs are likely to be needed to obtain good approximations, and although techniques exist to improve the rejection rates of ABC to reduce the number of iterations needed (Heard, 2014), the computational burden is still likely to be high for simulations that take a moderate time to run. Furthermore, neither ABC nor Bayesian Melding provide approximations to the underlying simulation at new points, meaning they are of limited use for analysing the structure of the parameter space and helping the modeller understand simulation behaviour.

The preferred approach in this thesis, then, is the use of statistical emulators to approximate the simulation - so called because they emulate the behaviour of the simulation: they are *meta-models* of the underlying models. Emulators introduce another layer of error and provide only approximate solutions, but are more flexible and less computationally expensive to run than other approaches. Many forms of meta-models are available, including neural nets and polynomial regression forms (Kleijnen, 2008; Forrester et al., 2008; Kamiński, 2015). However, in the case of highly non-linear simulations, polynomial regression-based functions may be insufficiently flexible to capture the underlying shape of the simulation response, while neural nets by default provide only point estimates of outputs.

Other potential forms of meta-model might include Generalised Additive Models (GAMs), as described by Wood (2006). This approach approximates the function of interest using sums of smooth functions of inputs. Generally these functions are modelled as penalised splines, with the smoothing parameters estimated using cross-validation (ibid). An approach using GAMs may be plausible, but the underlying assumption of smoothness, enforced by overlapping spline basis functions, may cause problems in the case of ABMS (although a similar criticism might also be levelled at Gaussian Process Emulators).

Polynomial Chaos expansions (PCE) are also proposed as a generic method for dealing with uncertain simulations by some in the applied mathematics community (e.g. Knio & Le Maître, 2006, for Computation Fluid Dynamics, Chalabi (2014) for Agent-Based Models), but these also have critics (O'Hagan, 2013). In particular, O'Hagan claims that although so-called 'intrusive' PCEs (which are tailored to a specific simulator) provide efficient meta-models, these are time-consuming to construct and most often spurned in favour of non-intrusive methods which treat simulators as black-boxes. Furthermore, the use of polynomial bases rather than the radial basis functions used in Gaussian processes (GPs) may lead to instabilities and less success in capturing local features of the output surface, and the need for approximations in the polynomial expansions in PCE means that not all uncertainty about the output predictions is accounted for (ibid). This stands in contrast to GPs, which explicitly account for uncertainty about unobserved parameter points.

All of these approaches are plausible, but Gaussian Process emulators have a number of appealing features which mean they can be conveniently applied to the problems of analysing and calibrating an Agent-Based Model. As a result, this combination of flexibility and desirable statistical properties mean Gaussian Process emulators are preferred throughout the thesis. Emulation will also be conducted within the Bayesian framework, which ensures that all the uncertainties can be described coherently (Kennedy & O'Hagan, 2001a, Oakley & O'Hagan (2002)). An introduction to Gaussian process emulators is provided below, and detailed information can be found on the website of the research community Managing Uncertainty in Complex Models (MUCM, 2011).

3.2.2 General premises

Gaussian processes are flexible statistical tools as they make few assumptions about the form of the function they are used to represent. Given what has been said about the complex and non-linear nature of Agent-Based simulations, this makes GPs well suited to the purpose described. The underlying premise is that outputs near to each other in the parameter space are also nearby in the output space. A Gaussian process represents this idea by insisting that outputs at any collection of inputs points are *joint multivariate normal*. A more formal treatment follows.

Various formal descriptions of Gaussian Processes are available. This follows most closely that from Kennedy & O'Hagan (2001a), MUCM (2011), and Andrianakis & Challenor (2011). Assume $f(.)$ is the uncertain function about which inference is required. For our purposes, it refers to a computer model that we want to analyse. One input to function can be denoted x, which is a vector of dimension k so that $x \in X \subset \mathbb{R}^k$, while an output is a scalar $y \in Y \subset \mathbb{R}$, such that $y = f(x)$. A Gaussian process is described conditionally on the parameters as a multivariate normal distribution for n realisations of f, $y_1 = f(x_1), \ldots, y_n = f(x_n)$, and is defined in terms of its mean and covariance functions (Andrianakis & Challenor, 2012, p4216):

$$f(.) \mid \beta, \sigma, \Omega \sim N[\,m(.),\ \sigma^2 c(\,.\,,\,.)\,] \tag{3.1}$$

The mean of the process, m, is generally modelled as a function of a vector of basis functions of x, $h(x)$, with coefficients β, such that for every output $f(x)$, $m(.) = h(.)^T \beta$ (MUCM, 2011). In this thesis, $h(x)$ is almost always modelled as a simple linear function of the inputs, so that $h(x)^T = (1, x)$ and $m(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, with k the number of input dimensions. This gives $q = k + 1$ as the number of basis functions (ibid).

The covariance matrix is determined by the correlation function $c(.,.)$, which determines how quickly nearby points become uncorrelated and is dependent on the parameters Ω (discussed below), and the variance parameter σ, which determines the extent of deviation from the mean function (Andrianakis & Challenor, 2011). Several forms are possible for the function $c(.,.)$, the most common of which is the squared exponential or Gaussian function (Rasmussen & Williams, 2006, p83; Kennedy & O'Hagan, 2001a, p433):

$$c(x_i, x_j) = exp\{-(x_i - x_j)^T \Omega (x_i - x_j)\} \tag{3.2}$$

The diagonal matrix $\Omega = diag(\omega_1, \ldots, \omega_k)$ is composed of roughness parameters $\omega_1, \ldots, \omega_k$, which indicate how strongly the emulator responds to particular inputs (Kennedy & O'Hagan, 2001a, p432-433; O'Hagan, 2006). Alternative parametrisation of these quantities as correlation lengths $\delta_i = 1/\sqrt{\omega_i}$ (e.g. MUCM, 2011) or $\psi_i = 1/\omega_i$ (e.g. Pepelyshev & Oakley, 2009; Gramacy & Lee, 2012) may also be seen.

Almost all ABMs, are stochastic in the sense that different results are obtained for repetitions of the same input point. In order to incorporate this inherent simulator stochasticity into the emulator, an additional variance term τ^2 (referred to as a *nugget*) can be added to the diagonal of the covariance matrix of the multivariate normal distribution in Equation 3.1. This formulation makes the assumption that this stochasticity is homoskedastic over the parameter space; this may not always be appropriate, and alternative methods allow relaxation of this assumption (Boukouvalas, 2010).

3.2.3 Estimation

In order to estimate the parameters of the emulator, a set of simulation data $f(D) = [f(x_1), \ \ldots \ , f(x_n)]$ (the *training set*) is required for N experimental points $D = x_1, \ldots, x_N$. Conditional on this training data and the values of the Gaussian Process parameters β, σ^2, and Ω in Equation 3.1, the distribution of simulator output at new points x is also joint multivariate normal (Oakley, 1999). Taking non-informative priors on β and σ^2 such that $p(\beta, \sigma^2) \propto \sigma^{-2}$, it is possible to marginalise β and σ^2, obtaining a multivariate-t posterior for outputs at new points, and the following distribution for the training data given roughness parameters (Andrianakis & Challenor, 2011, p4):

$$p(f(D)|\Omega) \propto |A|^{-1/2}|H^T A^{-1} H|^{-1}(\hat{\sigma}^2)^{\frac{n-q+1}{2}} \tag{3.3}$$

where H is the matrix of basis function generated by $h(D)$ and A is the training set correlation matrix defined by $c(D, D)$. The values that maximise this distribution can then be found and can be used as 'plug-in' posterior mode estimates of the values for Ω (Kennedy & O'Hagan, 2001a; Oakley, 1999).

Given Ω, expressions for estimates of β and σ^2 are (Andrianakis & Challenor, 2011, p4.):

$$\begin{aligned} \hat{\beta} &= (H^T A^{-1} H)^{-1} H^T A^{-1} f(D) \\ \hat{\sigma^2} &= \frac{1}{n-q-2}(f(D) - H\hat{\beta})^T A^{-1}(f(D) - H\hat{\beta}) \end{aligned} \tag{3.4}$$

Although this approach neglects the uncertainty around values of Ω, it is suggested that this uncertainty is not significant compared to that for other quantities (ibid). Full details and examples can be found in Andrianakis & Challenor (2011) and on the MUCM website (MUCM, 2011).

Alternatively, full MCMC sampling approaches can be used to estimate the posterior distributions of the unknown hyper-parameters (ibid). Direct maximisation of the multivariate normal likelihood for all parameters is also often used, particularly in machine learning contexts (Rasmussen & Williams, 2006; Boukouvalas, 2010).

To estimate the nugget parameter, following Roustant et al. (2012, Appendix A2) we introduce an additional parameter α which determines the proportion of total variance $\nu^2 = \sigma^2 + \tau^2$ explained by the inputs. The covariance function thus becomes (ibid, p49):

$$v(x_i, x_j) = \hat{\nu^2} * \{\alpha * exp\{-(x_i - x_j)^T \Omega(x_i - x_j)\} + (1 - \alpha) * I_{i=j}\} \qquad (3.5)$$

where $I_{i=j}$ is an indicator variable that equals 1 if $i = j$ and 0 otherwise. Note that now $\alpha * \nu^2 = \sigma^2$, while $(1 - \alpha) * \nu^2 = \tau^2$. The α can be estimated by including it in the set of parameters estimated by maximising the distribution in Equation 3.3 (ibid). Care must be taken if this approach is used to ensure that this variance matrix scaling is taken into account when predicting new points (see Appendix B to Gramacy, 2005).

3.2.4 Predicting new quantities

One immediate advantage of the Gaussian process approach is that once the parameters are estimated and the posterior distribution of the function f is obtained, new estimates of simulator outputs are very efficient to obtain, a particular advantage if the simulation is slow to run. Conditional on the training sample and the hyper-parameter estimates, the posterior mean prediction for simulator outputs at the new point x is just the result of matrix multiplication (Oakley, 1999, p13):

$$m^\star(x) = h(x)\beta + t(x)^T A^{-1} e \qquad (3.6)$$

where $m^\star(x)$ denotes the posterior mean function; $t(x)$ the correlation between the new point x and the elements of the training set D; and e is the difference between simulator outputs $f(D)$ and the prior mean prediction $h(x)\beta$. This allows the analyst or modeller to get a complete picture of the parameter space very easily. Furthermore, the uncertainty induced by the using the emulator as an estimate of the simulator is readily evaluated as well, so this source of uncertainty is not lost (Oakley, 1999, p13-14):

$$\begin{aligned} v^\star(x_i, x_j) =& \hat{\sigma^2}\{c(x_i, x_j) - t(x_i)^T A^{-1} t(x_j) \\ &+ (h(x_i)^T - t(x_i)^T A^{-1} H)[H^T A^{-1} H]^{-1} (h(x_j)^T - t(x_j)^T A^{-1} H)^T\} \end{aligned} \qquad (3.7)$$

where $v^\star(x_i, x_j)$ is the posterior covariance between points x_i and x_j and H is the matrix of basis function generated by $h(D)$. As discussed, the marginalisation of the σ^2 and β parameters mean that the variance at any collection of points is a multivariate T distribution with $n - q$ degrees of freedom. Whilst the variance function above is a complicated-looking function, it is simple to evaluate as again it only requires linear algebra.

3.2.5 Uncertainty Analysis

The emulator, once built, can also be used for an uncertainty analysis, which looks at how much uncertainty in the output is being induced by the set of inputs X (Oakley, 1999). This is particularly important in predictive, real world applications of Agent-Based Models where we might wish our simulation to inform decision making. Some inputs to our agent-based models might be based on noisy estimates from real world data, while others may be given priors that reflect our subjective assessment of their probable values (Werker & Brenner, 2004). In either case, we would like to quantify this lack of knowledge by treating these inputs as random variables with some assumed probability distributions. The uncertainty analysis propagates this uncertainty through the emulator, and takes it into account in providing estimates of the simulator's mean and variance (Oakley, 1999).

An orthodox Monte Carlo approach to this problem would be to repeatedly sample from the input distributions, run the simulator at each point, and examine the resulting distribution on the output (Saltelli et al., 2004). However, this is computationally expensive, as many simulation runs are required to get a good approximation of the output distribution. An alternative approach is to use the fitted emulator to conduct the Monte Carlo analysis instead, as it is many orders of magnitude faster in generating predictions. Even better, however, is that for inputs with normally distributed priors and squared exponential covariance functions, the work of Haylock (1997) provides analytical expressions for the relevant integrals of the emulator output over the input uncertainty. This allows easy computation of the expectation of the simulation output, the variance of this estimate, and the expectation of the simulator variance.

To summarise, assuming that $G(.)$ is the distribution function of the random input variables X, then the mean $E[f(X)]$ and variance $V = Var[f(X)]$ of the distribution of output $f(x)$ taking into account the input uncertainty are (MUCM, 2011, Haylock (1997), p43-44):

$$\begin{aligned} E[f(X)] &= \int_X f(x)dG(x) \\ Var[f(X)] &= E[f(X)^2] - \{E[f(X)]\}^2 \\ E[f(X)^2] &= \int_X f(x)^2 dG(x) \end{aligned} \tag{3.8}$$

The analytical expressions for these integrals, given in Haylock (1997) and MUCM (2011), take advantage of similarity of the squared exponential correlation function, the normal priors on $G(x)$, and the multivariate-normal posterior distribution of $f(x)|\sigma^2, f(D)$. The expressions are very long, and so are not reproduced here, but are cheap to evaluate.

3.2.6 Sensitivity Analysis

The purpose of a sensitivity analysis is to understand how the model output responds to changes in inputs. Historically, these have been conducted by assessing the change in output for small changes in input at some specified point of interest (a local sensitivity analysis) (Saltelli et al., 2004). The partial derivatives of the approximated function at this point are often used for this purpose (ibid). This is problematic in the case where the whole input space is potentially of interest, particularly if the model is non-linear, in which case the derivatives at one point are not representative of the rest of the input space (Saltelli et al., 2008). Thus global measures of model sensitivity that summarise the behaviour of the outputs across the input space are to be preferred (ibid).

Although there are various methods for conducting sensitivity analyses (Saltelli et al., 2004, 2008), variance-based methods provide a way to utilise emulators to maximise efficiency (Oakley & O'Hagan, 2004). Sensitivity Analysis is defined by Saltelli et al. (2004, p45) as "The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to difference sources of uncertainty within the model input". This definition provides a link to the uncertainty analysis described in the previous section (ibid), which defined a method for finding the expectation of the total uncertainty due to inputs in our model. The methods described below aim to partition this uncertainty between inputs (Oakley & O'Hagan, 2004).

The principal variance-based measure used in this thesis is the *sensitivity variance* V_w, where w here identifies the input or collection of inputs which we are interested in apportioning variance to (MUCM, 2011). This measures the reduction in overall variance that would result from knowing the value(s) of x_w (Oakley & O'Hagan, 2004, p761):

$$\begin{aligned} V_w &= Var\{f(x)\} - E\{Var(f(x)|x_w)\} \\ &= Var\{E(f(x)|x_w)\} \\ &= E\{E(f(x)|x_w)^2\} - E\{f(x)\}^2 \end{aligned} \tag{3.9}$$

Dividing by the total variance $L = Var\{f(x)\}$, we get a scale-invariant *sensitivity index* $S_w = V_w/L$ (Oakley & O'Hagan, 2004). Where the set of inputs w contains only 1 input, we obtain the main effect for that input, and the additional reductions obtained for combinations of inputs are the joint (interaction) effects (Oakley & O'Hagan, 2004; Bijak et al., 2013). The expectations of the conditional variances for input subsets are given in closed form by MUCM (2011), assuming normal priors on the inputs and squared exponential correlation functions.

Sensitivity analyses can be extremely useful tools in analysing ABM and assessing their robustness. Firstly, just knowing which inputs are important in a simulation and which are not is helpful in understanding the processes involved - if simulation is not sensitive to a given parameter, then that parameter can safely be ignored (Grimm & Railsback, 2005). Secondly, given that ABMs may require assumptions regarding the values of some parameters due to lack of data, finding that outputs are not that sensitive to changes in such parameters helps justify these modelling choices (ibid). Thirdly, understanding which inputs are contributing most to our uncertainty helps target where we should gather more information in order to increase the precision with which we can estimate outputs of interest (Oakley & O'Hagan, 2004).

3.2.7 Probabilistic Calibration Using Emulation

Emulation techniques can be used to *calibrate* a simulation using empirical data (Kennedy & O'Hagan, 2001a). This will be discussed in detail in Chapter 5. The idea is that we can learn about 'true' values of unobserved inputs by examining what values of these inputs result in simulation outputs that match reality. Input parameters are divided into two groups; those which are known or measured and those which must be calibrated. Probability distributions for the calibration parameters can be estimated in a unified Bayesian model taking into account all of the sources of uncertainty discussed above, including prior uncertainty about parameter value, model discrepancy, and observation error (ibid).

An alternative calibration method using emulation is *History Matching*, so-called because it identifies simulation input spaces which provide plausible matches to real (historical) data series (Vernon et al., 2010). Emulator predictions are used to rule out implausible areas of the parameter space by examining their compatibility with empirical observations, taking uncertainty into account (ibid). New simulation runs are then sought in the remaining 'non-implausible' area until no further reduction in space can be achieved, or the available computing resources are exhausted (ibid). Again, more detail and an example is provided in Chapter 5.

3.3 Value of Emulation Techniques for the Anaylsis of Agent-Based Models

The preceding sections of this chapter have discussed one framework for the design, analysis and calibration of computer experiments associated with Managing Uncertainty in Complex Models community (MUCM, 2011). There are some obvious synergies between these techniques and the more general approaches to analysing ABM proposed by

Werker & Brenner (2004), Boero & Squazzoni (2005) and Grimm et al. (2005) discussed in Chapter 2.

1. Uncovering the relationships inherent in the simulation model itself and examining its compatibility with theory, the *analytical adequacy* phase, becomes easier with the use of experimental design and emulation techniques. Efficient designs increase the ease with which one can explore the parameter space, and having a fast approximation to the simulator output allows the modeller to develop a comprehensive picture of the response of the ABM to its parameters. Furthermore, the availability of analytical global sensitivity measures gives a quick metric to assess the theoretical importance of various inputs.
2. The process of comparing a model to its real world target is facilitated by emulation. For example, the two stage process advocated by Werker & Brenner (2004), in which empirical data is used to reduce the input space before 'abductive' methods are used to find the set of generating input values, has a natural implementation in a Bayesian emulation framework. The first stage can be achieved by eliciting (prior) or estimating (posterior) probability distributions for all relevant inputs (Oakley, 2002). The second stage can be achieved by using these distributions as priors on input parameters, and using the calibration methods sketched above, and described in more detail in chapter 5, to find the posterior distribution for the calibration parameters.

The next two chapters aim to provide simple demonstrations of these methods in action on demographic simulation applications. Chapter 4 describes the analysis of a replication of the Agent-Based Model of Billari et al. (2007), and Chapter 5 calibrates a micro-simulation using both the full Bayesian calibration techniques of Kennedy & O'Hagan (2001a) and the History Matching method of Vernon et al. (2010).

4

Analysis of a Demographic Agent-Based Model of Marriage

4.1 Emulating the 'Wedding Ring' Model

Building on the work conducted in Bijak et al. (2013), a brief example of the use of emulators to examine the behaviour of a simple agent-based model of partnership formation is now presented. The model in question is a re-implementation of the 'Wedding Ring' model of Billari et al. (2007), with the addition of demographic data for the UK. In particular, fertility and mortality data from 1950-2011 is introduced, together with a starting population taken from UK census data (see Bijak et al., 2013, for data sources). The model itself aims to show how aggregate age-at-marriage patterns can be built up from the effect of social pressure on individual partner search intensities (Billari et al., 2007). The work extends upon Bijak et al. (2013) in three ways.

1. It utilises a Latin Hypercube design to gain more information with fewer runs.
2. Additional simulation runs are used to validate the emulator using the approach of Bastos & O'Hagan (2009).
3. The heteroskedastic emulation techniques of Boukouvalas (2010) are applied in order to attempt to better fit the surface.

The focus of this section will be on explaining the use of the emulator, but a brief description of the model follows in order to aid understanding. A fuller exposition can be found in Bijak et al. (2013), as well as in the original paper by Billari et al. (2007), and the model code is available at `https://www.openabm.org/model/3549/version/2/view`. Individuals within the simulation reside on a ring and have a number of 'relevant others' who form their social network. The proportion of these others who are married enters a function that determines the radius within which an individual searches for partners. The sigmoid shape of this function is controlled by two parameters a and b [1]. An additional parameter 'spatial distance' or sd controls the distance within which individuals draw social network members.

We analyse here a single output quantity, the average proportion of agents married over the course of the simulation. The results presented here differ slightly from those presented in previous work, as a larger starting population is used, and a Latin Hypercube Sample rather than a grid sampling design defined the set of input points. As the output data is a proportion, transformations might have been considered to ensure the predicted values remain bounded between $[0, 1]$. However, given that the output data does not approach either bound, the data was left untransformed (Gelman et al., 2014).

The simulation is interesting as it involves the phase transition type behaviour we associate with Agent-Based models. As parameters a and b decrease, individuals become

[1] In Bijak et al. (2013) and Billari et al. (2007), the two parameters were α and β, and a and b where reserved for other model quantities. However, to avoid confusion with the emulator parameters, a and b are preferred here.

more responsive to pressure from other agents in their social network. Relatively small changes in parameters values result in the system changing behaviour from situations where most agents stay unmarried, to those involving almost everyone marrying (Bijak et al., 2013). This provides a useful example as to what extent Gaussian Process emulators can deal with this kind of transition, given that they assume smoothness in the parameter space and have been criticised for this reason in the analysis of ABM (Chalabi, 2014).

4.2 Input Design and Emulator fit

Firstly, a training set was obtained by generating a Latin Hypercube sample, initially lying in $[0, 1]$. Two hundred design points were specified, each consisting a three values, one for each simulation input. Furthermore, an additional 50 points were obtained for the purposes of validating the emulator. Following Bastos & O'Hagan (2009), these consisted of 25 additional space filling points, chosen to maximise distance from the existing points, and 25 points relatively close to those in the original sample. Such choices maximise the value of the validation sample, as evaluating error at a range of distances from existing points better tests the estimated values of the roughness hyper-parameters (ibid). The simulation was then run at all of these points, obtaining $N = 200$ training set input and output pairs, and $m = 50$ validation pairs. Note that as the LHS sample is generated to lie between $[0, 1]$, it must be scaled for purposes of input to the simulation. First, a range of possible input values must be specified for each parameter, representing our best guess (prior knowledge) of what the most extreme reasonable values for these parameter might be. Then the following transformation is applied to each LHS point to get to the required scale:

$$x'_{i,j} = x_{i,j}(high_j - low_j) + low_j \tag{4.1}$$

Where $x'_{i,j}$ is the j^{th} element of the i^{th} input to the simulation; $x_{i,j}$ the corresponding LHS input in the range $[0, 1]$; and $high_j$ and low_j represent the relevant range endpoints for that input.

A Gaussian Process emulator is fitted to the training data. Note that the original $[0, 1]$ scale input design is used for fitting purposes in order that the roughness parameters can be estimated accurately, and furthermore can be easily compared. This difference in scale must be taken into account later when interpreting the parameters. The R Statistical Computing Language (R Development Core Team, 2015) was used for all estimation. To estimate the emulator hyper-parameters, the mode of the joint marginal likelihood of the roughness parameters and the hyper-parameter α was first found numerically using the built-in R function `optim`. Several starting points were trialled to avoid a local maximum being chosen. Values of β, σ^2 and τ^2 follow given these hyper-parameters, and the full emulator is obtained. The values of the fitted parameters are given below in Table 4.1.

TABLE 4.1: Parameter Estimates for emulator of 'Wedding Grid' model

Parameter Name	Estimates
$\beta = (\beta_0, \beta_a, \beta_b, \beta_{sd})$	(0.694, -0.3, -0.221, 0.006)
$\Omega = (\omega_a, \omega_b, \omega_{sd})$	(27.65, 31.7, 0.03)
σ^2	0.00574
τ^2	0.00025

As can be seen, the value of the parameters of the mean function and the roughness parameters are much smaller for the sd input than for a and b, indicating that changes in this input lead to small changes in the output, while the variability associated with inputs is considerably larger than uncertainty relating to simulation stochasticity τ^2.

4.3 Emulator Validation

Before we can be confident that our emulator accurately represents our simulator, we should attempt to check its prediction against the validation set. Bastos & O'Hagan (2009) propose several metrics to assess emulator validate, two of which are reported below. Emulator predictions and variances are obtained for the validation input points, and standardised residuals calculated for each: these are displayed in Figure 4.1.

The values of these residuals appear to be relatively reasonable; most lie within the range $[-2, 2]$. However, such metrics do not take into account the correlation between residuals implied by the Gaussian covariance structure. The Mahalanobis distance can better represent emulator validity taking this into account, and is calculated through the formula (Bastos & O'Hagan, 2009, p429):

$$MD = (y_{cv} - E(f(X_{cv})))^T (V(f(X_{cv})))^{-1} (y_{cv} - E(f(X_{cv}))) \tag{4.2}$$

where y_{cv} indicates the outputs for the m validation points, and $E(f(X_{cv}))$ and $V(f(X_{cv}))$ represent the emulator mean and variance estimates at these points. This value, multiplied by $(n-q)/(m(n-q-2))$, can be compared to the quantiles of an F-distribution with m and $n-q$ degrees of freedom (Bastos & O'Hagan, 2009). Small values indicate under-confident predictions, in that the predictive distributions were too wide given the actual differences between simulators and emulators, while high values indicate the opposite. For this emulator, the 90% intervals for the relevant scaled F-distribution are

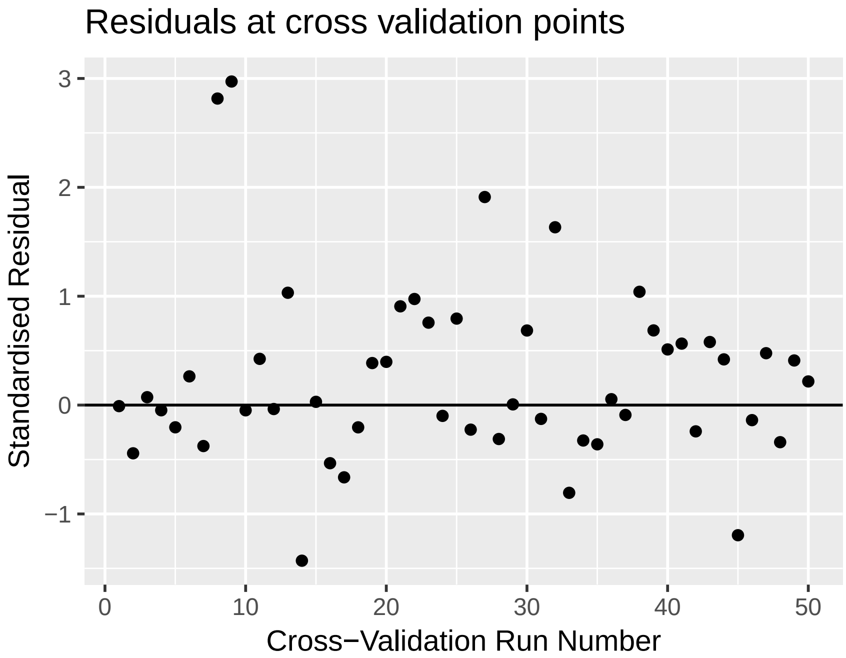

Figure 4.1: Plot of standardised emulator residuals at validation points

30.9 and 74.9, and the calculated Mahalanobis distance is 38.6, suggesting the emulator is reasonably accurate in quantifying its uncertainty about unknown points. In practice, if high values for Mahalanobis distances are obtained, suggesting a poorly fitting emulator that is overconfident about its ability to represent the simulator (ibid), outputs at more design points could be collected in order to attempt to obtain a better fit (ibid).

4.4 Predictions and Uncertainty and Sensitivity Analyses

Rebuilding the emulator using the validation points as well, predictions are obtained for a range of values for the first two parameters, using the formulas above. These predictions are displayed in Figure 4.2 and Figure 4.3 and the 95% predictive intervals resulting from uncertainty due to simulation stochasticity and emulator uncertainty are also shown. Each plot is a slice through the parameter space taken at mid-point values for the other two parameters. The bivariate predictions are displayed in Figure 4.4, again at the midpoint of the social distance parameter. As an aside, the fitted emulator allows many predictions to be generated with little computational expense for such plots, and so the use of animations using time as an additional visualisation dimension could be considered to show the effect of all inputs in one plot. This might assist with the task of understanding simulation behaviour, the need for which was highlighted in Chapter 2, as long as the emulator fits the underlying simulation reasonably well.

As well as showing predictions, Figure 4.3 also shows the outputs for a set of 250 test runs of the simulation, specified to lie on the same slice of the parameter space. As can

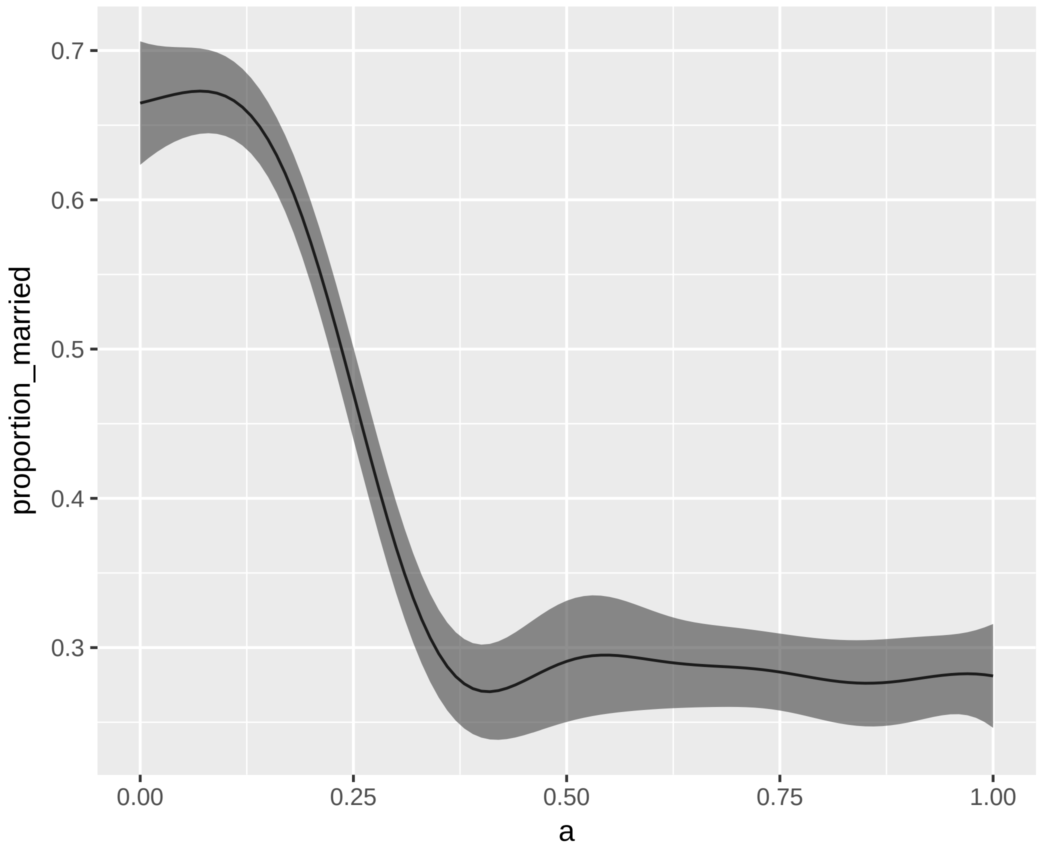

FIGURE 4.2: Plot of emulator predictions by parameter a with 95% predictive interval

be seen, despite passing the validation test, the emulator performs quite poorly close to the phase transition, and in general is under-confident about its ability to predict elsewhere, as away from the phase transition, all points lie well inside the predictive interval.

To conduct uncertainty and sensitivity analyses, assumptions about the distribution of the inputs must be made. For convenience, we assume normal distributions around the midpoint of the input ranges with variances such that the distributions span the input range, $X_i \sim N(0.5, 0.02)$. In other applications, these distributions could reflect substantive prior knowledge. In such a case, the uncertainty analysis allows us to infer how input uncertainty feeds through to uncertainty about outputs. Given the chosen distributions, then, the predicted mean simulator output was 0.33508, and the variance of this estimator was close to 0. The expectation of the overall simulator variance was 0.00874.

Similarly, a sensitivity analysis is conducted to examine how sensitive the simulator is to changes in the various inputs, given the probability distributions of these inputs. The findings are summarised in Table 4.2.

As previously reported in Bijak et al. (2013), the parameters a and b, controlling the way marriage search intensity responds to social pressure, are most significant in causing

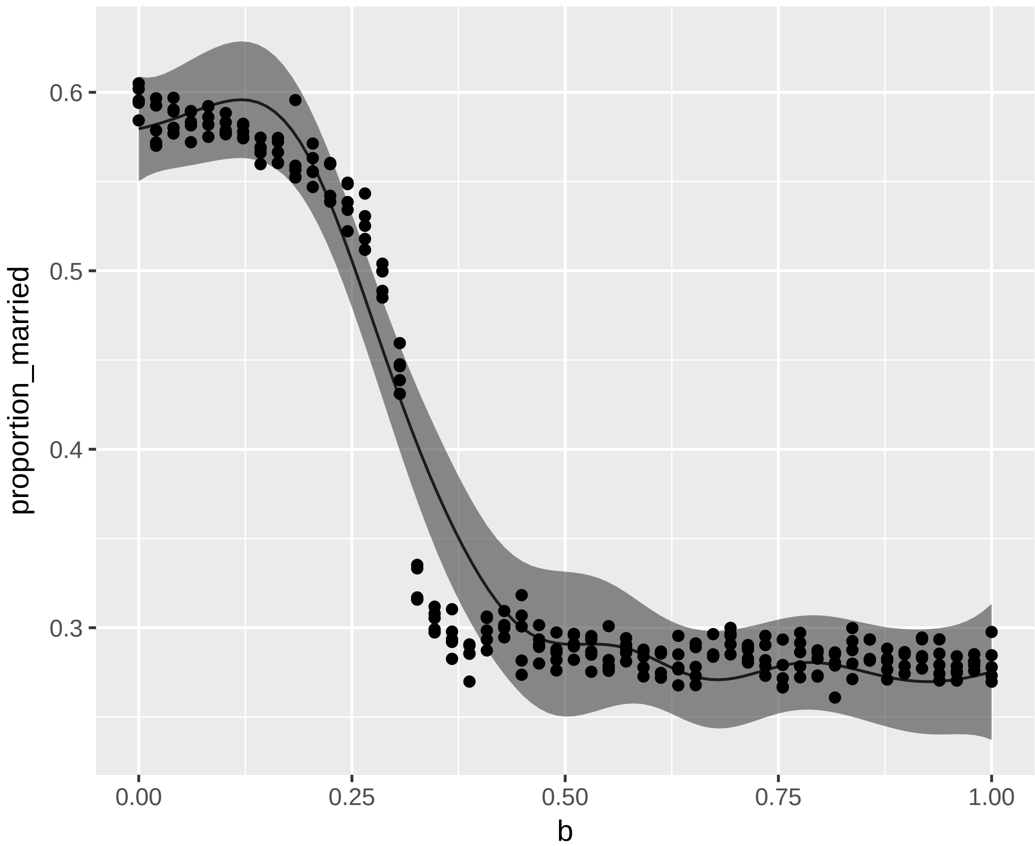

FIGURE 4.3: Plot of emulator predictions by parameter b with 95% predictive interval

changes in outputs. The numbers in the table can be interpreted as proportions of total output variance (excluding the stochastic variance associated with the nugget) explained by each input or combination of inputs. The first row describes the proportionate reduction in uncertainty associated with observing each input alone, while the symmetric matrix in the remaining rows describe the additional reduction gained from knowing the values of two inputs together. Diagonals are not defined, as inputs do not interact with themselves. The numbers above the diagonal, representing main effects and all possible two-way interactions, do not sum exactly to 1, as a small amount of residual variance is found in the three way interaction and not reported.

TABLE 4.2: Sensitivity Analysis

	a	b	Spatial Distance
Main Effect	0.357	0.461	0.001
Interaction w/ a	NA	0.176	0.001
Interaction w/ b	0.176	NA	0.002
Interaction w/ Spatial Distance	0.001	0.002	NA

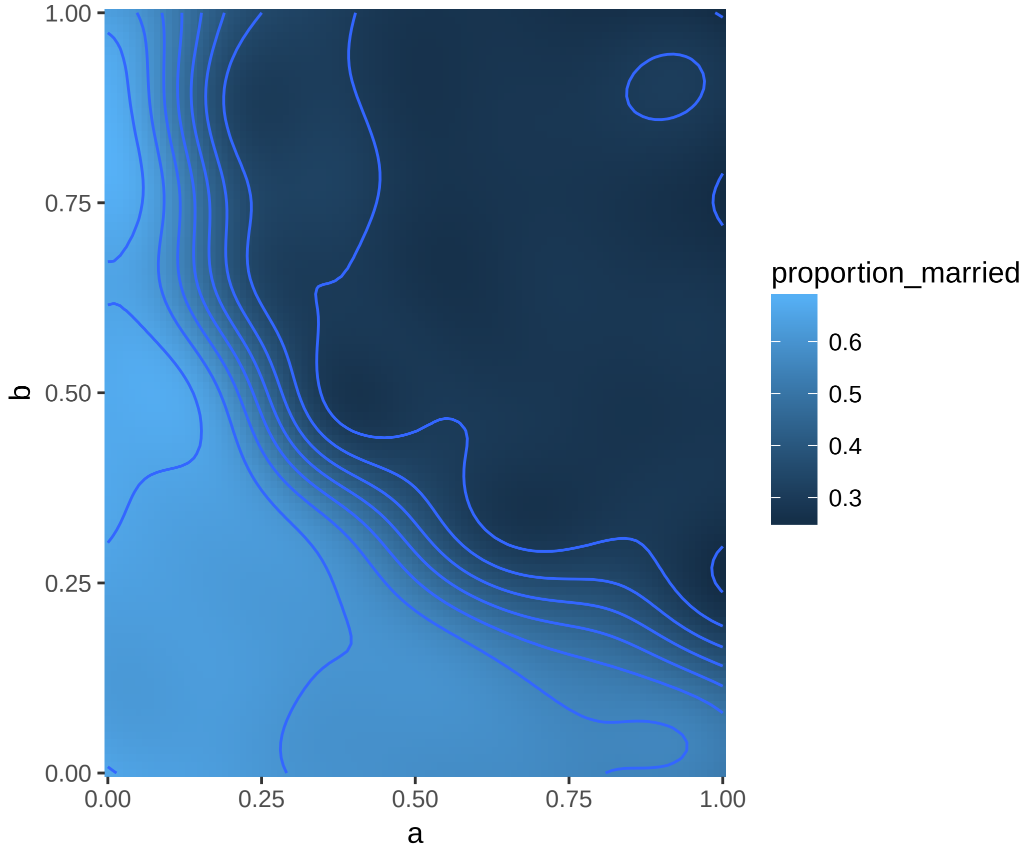

FIGURE 4.4: Bivariate plot of emulator predictions by parameters a and b

4.5 Heteroskedastic Emulation

As observed from Figure 4.3, the default emulator fails to adequately capture behaviour around the phase transition, which is considerably different from elsewhere in the space. In particular, both variance and correlation lengths might be expected to be different across the boundary seen in the contour plot in Figure 4.4. A number of approaches could potentially be useful in addressing this. The Treed Gaussian Process methods of Gramacy (2005) would allow different models to be fitted to different areas of the parameter space (see also Heard, 2014), but this approach is most convenient for axis-orientated discontinuities. Chalabi (2014) suggests the use of Polynomial Chaos expansions to model uncertainty in Agent-Based Models (O'Hagan, 2013), but this is not attempted here.

4.5.1 Concept

A simplified version of the heteroskedastic emulation approach of Kersting & Plagemann (2007) and Boukouvalas (2010) is adopted for the analysis in this section with the aim that the increased flexibility of allowing non-constant variances will improve emulator performance. This method collects repeated simulation data at a some or all of the design points, and estimates the sample mean and variance at these points. An emulator is first fitted to the data, and used to generate estimates of variances at design points where only one observation has been collected. Next, an emulator is fitted to the log variances (both those estimated directly from the sample, and those obtained in the previous step), with suitable adjustments made for the log transformation. This done, the smoothed variance predictions for the design points are used to build a emulator for the simulation mean by providing an estimate of the variability of the sample mean estimate. An iterative process follows where mean estimates are used to generate new variance estimates and vice versa until some convergence criterion is reached (ibid).

Two emulators are thus used in predicting new points; mean predictions can then be generated using the usual formulas, while variance predictions from the variance emulator allow predictions of the heteroskedastic stochastic variance to be added to that relating to uncertainty about the mean (ibid).

4.5.2 Application and Results

A slightly simplified version of the process described is used here. Outputs from multiple simulation runs are obtained at *every* design point, rather than at only a selected group. This suggests a reduction in the importance of the iterative stage, which is mainly important for updating estimates of mean and variances at points which have only one simulation value. Thus, only one iteration of the above algorithm is conducted.

To fit the heteroskedastic emulator, 400 design points with 6 repetitions at each point were obtained, and the appropriate variance and mean emulators fitted. Plots of the outputs by the a and b parameters are displayed in Figure 4.5 and Figure 4.6.

By visual inspection, the heteroskedastic emulator gets closer to the shape of the underlying function than the homoskedastic equivalent, but this may be due to the inclusion of double the number of design points. Figure 4.6 shows that the emulator mean prediction appears to be closer to the simulation outputs around the phase transition than in the homoskedastic case, and allows tighter predictive intervals over the rest of the parameter space.

However, the emulator still is not as successful at predicting at the phase transition - the globally-estimated correlation lengths work to reduce uncertainty at these quickly-changing areas, and it appears as if the variance emulator over-smooths variances across the phase transition, perhaps because not enough sample points lie in this narrow band. The phase transition occupies only a very small part of the total area of the parameter space, and while the emulator can point to the location of this transition, with only very few observations of the region it may struggle to represent the system's behaviour there fully.

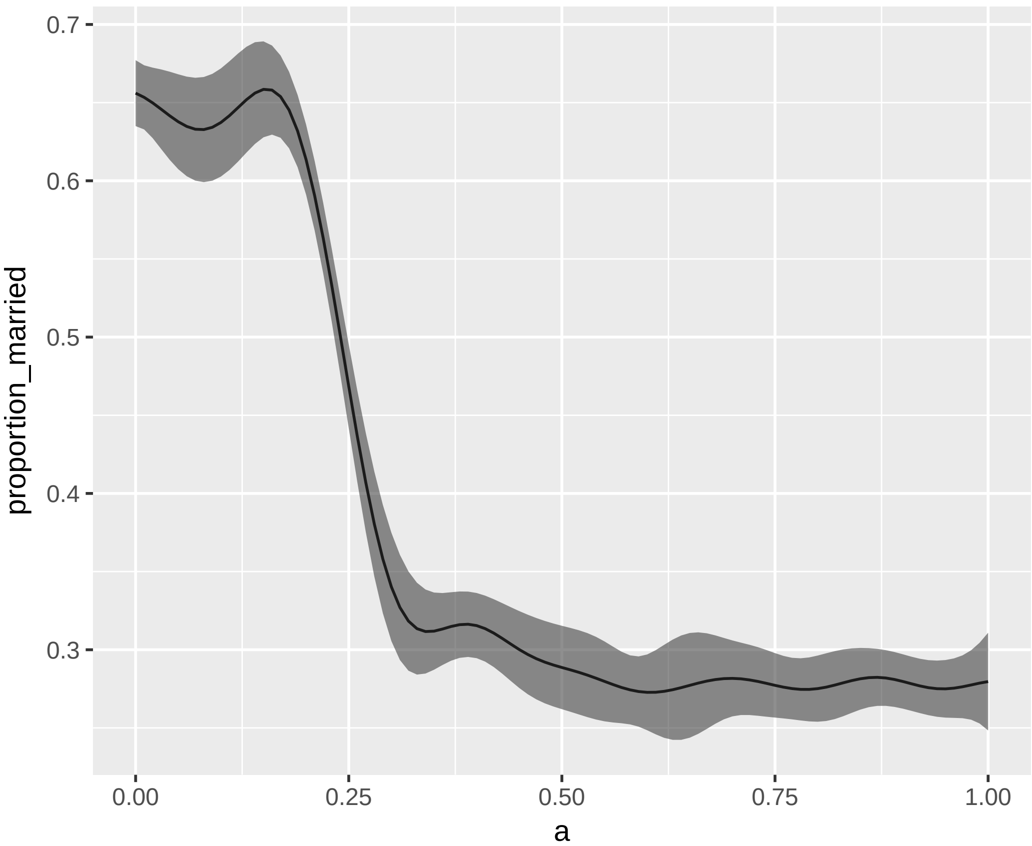

FIGURE 4.5: Plot of heteroskedastic emulator predictions by parameter a with 95% predictive interval

One approach to improve this performance might be to collect more repetitions per point to reduce noise in sample variance estimates - however, without improving the sampling scheme this will still lead to under-sampling of the transition. A better strategy might be to attempt to oversample the interesting area using sequential design of experiments principles, which Boukouvalas (2010) also discusses. Boukouvalas (2010) samples the areas where uncertainty is high, but taking the approach of Luke (2007) might prove more successful. He uses a genetic algorithm to sample areas with large values for function derivatives. Given that we have analytical expressions for the emulator derivatives (Oakley, 1999), a second wave of sampling designed to maximise information about parameter combinations that are predicted to result in high slopes may help with this problem.

However, there is still the problem that a meta-model that assumes smoothness may struggle to mimic this type of system successfully. Adopting correlation function forms that decay more rapidly might reduce the effect of this assumption (Rasmussen & Williams, 2006), and so this is another direction for future investigation.

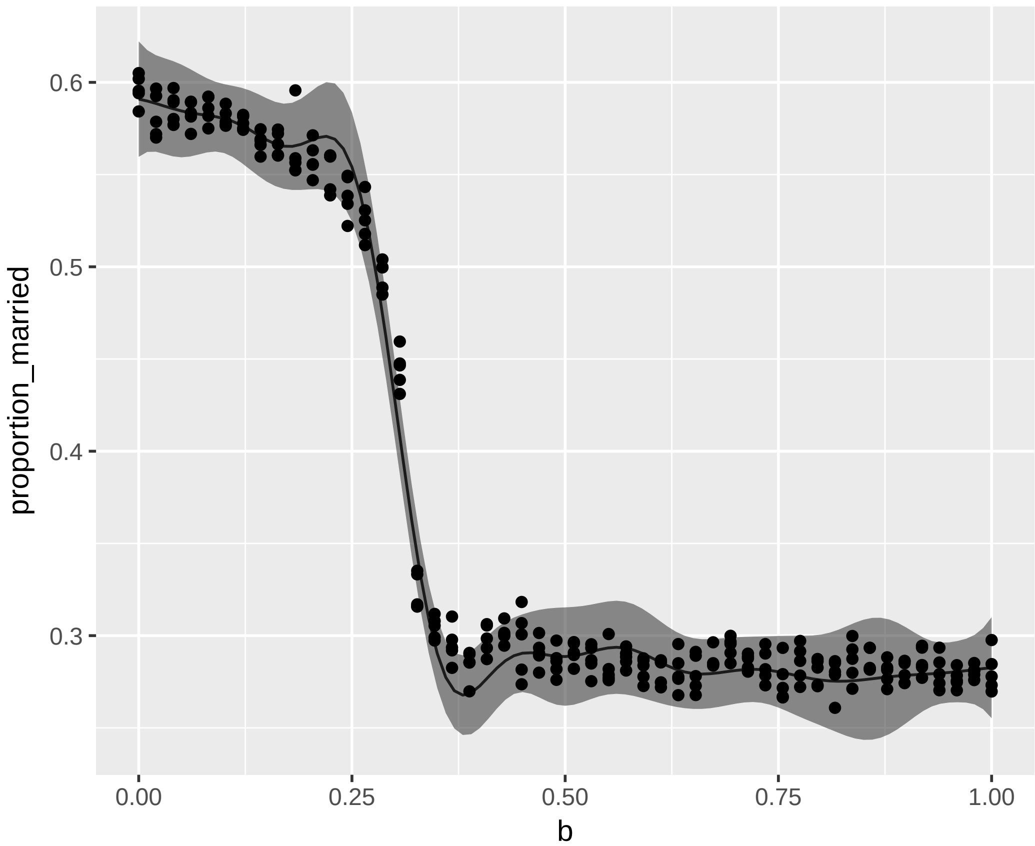

FIGURE 4.6: Plot of heteroskedastic emulator predictions by parameter b with 95% predictive interval

4.5.3 Conclusion

This chapter has demonstrated the use of Gaussian Process Emulators to analyse agent based models in demography through generating predictions and conducting uncertainty and sensitivity analyses. In Chapter 2, a need for methods both for analysing the internal behaviour of demographic Agent-Based Models and for relating them to external data was identified. This chapter shows how GP emulators can meet the first of these needs, and advocates the use of heteroskedastic emulators to best capture simulation stochasticity.

By examining the predictions of the meta-model, the above analysis identifies the relationships between the two parameters a and b. These parameters define the response of agents to increases in the proportion of their networks who are already partnered. As was identified in Bijak et al. (2013), a phase-transition occurs in parameter space, marking a divide between simulations where many agents stay unmarried, and those where initial marriages quickly induce a cascade of others.

The flexibility of the Gaussian process enables this dynamic to be captured by the emulator without a need for any prior knowledge about the shape of response surface, other than the correlation assumptions of the Gaussian process prior. The relaxation of the assumption of homoskedasticity inherent in the standard emulators using the heteroskedastic emulation approach of Boukouvalas (2010) is believed to be important for the analysis of demographic Agent-Based Models in future. Given that Agent-Based Models often exhibit emergent behaviour (see Chapter 2), whereby feedback loops caused by multiple random interactions between individuals lead to unpredictable macro-level behaviour, it is very probable that for many such models volatility in some areas of the parameter space will be considerably higher than in other areas. Failing to take this into account will likely lead to poorly fitting metamodels.

Further work is needed in examining how to provide better fits around phase transitions, however, and sequential sampling designs and varying the correlation function are the two suggested ways to proceed.

5

Calibration of a Demographic Micro-simulation of the UK

5.1 Calibrating Social Simulations

The previous chapter showed how emulation techniques can aid us with the analysis of Agent-Based Models, providing a computationally cheap way of approximating simulation outputs at any combination of input points, and furthermore allowing an analysis of the global sensitivity of outputs to the inputs of interest. However, as has been discussed in Chapter 2, once we go beyond merely investigating the implications of various assumptions using simulation, we also want to know about the relationship between our simulation and what we have observed empirically (Werker & Brenner, 2004).

This requires a further extension of the techniques discussed in the previous chapter to include the calibration problem (sometimes called the inverse problem (Grimm & Railsback, 2005)). This involves attempting to find values of simulation inputs that lead to simulation outputs that closely match reality. Given that social simulations are by their nature simplifications of and abstractions from whatever it is they wish to study, it is necessary also to account for the presence of *discrepancy* between model and reality in this relationship (Kennedy & O'Hagan, 2001a).

5.2 Simulation Setup

Although this thesis is focused on approaches for Agent-Based Models, for the purposes of this chapter the demographic micro-simulation of Zinn & Himmelspach (2009) is used. Micro-simulation and Agent-Based Models have similar approaches in that they both stochastically simulate at the individual level, the former using fixed probabilistic relations, the latter by modelling some element of decision making. In practice, it is also difficult to definitively categorise models as exclusively one or the other, as many Agent-Based models rely on estimated transition probabilities in certain sections of their modelling (e.g. Geard & McCaw, 2013), while micro-simulations have often included decision-type rules where data is unavailable, in particular for questions of partnership formation (e.g. Hammel et al., 1979; Zinn, 2011). Bijak et al. (2013) discuss the commonalities and differences between the approaches in more detail (Section 2). The relative lack of individual level interactions in models at the micro-simulation end of the spectrum, however, might make them less prone to the sorts of non-linearities we often see in ABMs (Epstein & Axtell, 1996; Gilbert & Troitzsch, 2005), and this must be borne in mind when considering the methods applied here; the micro-simulation calibration attempted here is an *easier* problem than we are likely to face when attempting to do the same thing with Agent-Based Models.

The micro-simulation used is MicCore (Zinn & Himmelspach, 2009), a fast Java-based platform for building continuous-time simulations with arbitrary collections of states and rates. MicCore also has an interface to the R Statistical Computing Language

(R Development Core Team, 2015) in the form of an R-package (Zinn, 2015). All that is required is to specify sets of transitory and absorbing states, the functions that define the probabilities of transitions between these states, and the size and distribution over states of the initial population.

The starting population for the simulation was generated to reflect the age and sex composition of the UK in 1950 using data from the Human Mortality Database (Human Mortality Database, 2011). The inputs to the simulation are parameters for functions defining transitions rates for fertility and mortality by age adapted from those described in the examples provided with the MicSim R package (Zinn, 2015)[1]. The fertility function is a Hadwiger function taking the following form (Chandola et al., 1999, p320):

$$m_f(x) = a\frac{b}{c}\left(\frac{c}{x}\right)^{\frac{3}{2}} exp\left(-b^2\left(\frac{c}{x}+\frac{x}{c}-2\right)\right) \tag{5.1}$$

where $m_f(.)$ are the fertility rates, x is age; and a, b, and c are input parameters. Roughly speaking, parameter a works to scale the distribution multiplicatively, c controls the location of the peak, while b is related to the distribution's shape.

Mortality is treated similarly simply as a Gompertz mortality function (eg Lee et al., 2014):

$$m(x) = d * exp\left(e * x\right) \tag{5.2}$$

where $m(x)$ describe the mortality rates, and d and e are further parameters describe the slope and intercept of the log mortality function.

This is clearly a very simple model of population change, one that would not generally require microsimulation to analyse. However, it is used as a basis point for the examination of how to calibration stochastic demographic simulations with a view to preparing the ground for the complex Agent-Based model which will be analysed in the next chapter. This model is also appealing because it provides a test case for calibrating in the presence of *model discrepancy* in such a context (Kennedy & O'Hagan, 2001a). The calibration target observation is the distribution of population by age in 2011, with data again drawn from the Human Mortality Database (Human Mortality Database, 2011). Clearly, providing fertility and mortality rates that are constant over time will not be able to replicate exactly this distribution. To begin with, we know that mortality rates have decreased over time, and that fertility has fluctuated between the end of the second world war and the 1970s (the baby boom and bust). Secondly, immigration is completely ignored in the simulation, and so it is expected that the model will

[1]My thanks to Sabine Zinn for providing the MicCore Jar file and supporting R code in order to facilitate the work in this chapter.

significantly underestimate the size of the UK population in 2011, given the waves of immigration to the UK since the second world war.

From a substantive point of view, then, this model is a gross simplification, and no attempt is made to claim that it is fit for the purpose of prediction or explanation. It is used here merely to demonstrate the possible utility of emulator-based calibration methods for demographic agent-based models. Furthermore, the existence of discrepancy and the fact that there is some knowledge about the direction and extent of it gives us an opportunity to examine how the ability to formally incorporate such prior knowledge helps us deal with the presence of non-modelled processes in demographic simulations (Brynjarsdóttir & O'Hagan, 2014), which are likely to be present in Agent-Based calibration problems as well as in microsimulations.

5.3 Full Bayesian Calibration

Two calibration methods are examined in this chapter. This section attempts a full Bayesian calibration of the simulation in order to obtain a probability distribution for the unknown variables following the approach of Kennedy & O'Hagan (2001a). The following section uses History Matching (Vernon et al., 2010), introduced briefly in Chapter 3, to narrow down the area of the input space deemed 'non-implausible' by a process of iterative refocusing. In both cases, we wish to calibrate the input parameters to obtain a good match for the distribution of population over age in 2011. Given that age distributions are likely to be important to many demographic agent based models, this may hopefully be instructive to future work in the field.

The essence of Kennedy and Hagan's (2001a) approach to the calibration problem is to divide input parameters into two groups, and relate these to observations by means of a calibration equation. The first group are termed 'control' or *'location'* inputs, and these are known for every empirical data point we collect. The term location relates to the early use of Gaussian Processes in geostatistics, where location variables would correspond to physical coordinates, which are generally known for every measurement taken. In what follows, it therefore seems natural to consider age a location input, as values corresponding to empirical measurements and to simulation outputs are known. Input variables are denoted as x in line with the previous chapter.

The second group of inputs are *calibration* parameters. These are not observed for real world measurements, but could be considered to have 'true' values in reality which we wish to find. These are denoted θ and can be given some prior distribution, either vague or reflecting some substantive knowledge about their expected values. Kennedy & O'Hagan (2001a) also differentiate between these true values and the various corresponding 'guesses' of θ used in the training sample, denoted as t_i. Thus, the i^{th} training set simulator run is now denoted $f(x_i, t_i)$, while its correlation with the j^{th} run is denoted $c((x_i, t_i), (x_j, t_j))$. It is also now necessary to distinguish between the training set D_1 (previously just D) and the points at which observations are gained D_2, which contains only location parameters. Note that the notation follows Kennedy & O'Hagan (2001a) as closely as possible.

The calibration equation is (Kennedy & O'Hagan, 2001a, p435):

$$z(x) = \rho f(x, \theta) + d(x) + \epsilon(x) \tag{5.3}$$

where $z(x)$ is the true process at location x, $f(x, \theta)$ is the simulator output at 'real' best input values θ, ρ a regression coefficient scaling the simulation output and ϵ is the observation error. The other element of the calibration equation is the model discrepancy term $d(x)$. This represents the mismatch between the simulator and reality *given that the simulator is run at the best input values of the calibration parameters.* This captures the idea that the simulator is a simplification of reality and may not match reality exactly even if it where fully calibrated. This discrepancy function over the location inputs is modelled as another Gaussian process, priors for the parameters of which must be elicited from the relevant modellers and domain specialists (Oakley, 2002; Vernon et al., 2010). Modelling this discrepancy function is very important, as without it, posterior distributions for the true values of θ may be biased or falsely precise (Brynjarsdóttir & O'Hagan, 2014). However, this is not an easy task, and in practice it may be very difficult to elicit these types of distribution, particularly as we may have no prior clear idea of the simulator will behave in any case.

Given the above framework, the stacked vector of simulation outputs and observations $\zeta = (f(D_1), z(D_2))^T$ can be modelled as a function of the emulator and discrepancy function as follows (Kennedy & O'Hagan, 2001a, p437-438):

$$\begin{gathered} E(\zeta|\theta, \beta, \phi) = m_\zeta(\theta) = H(\theta)\beta \\ H(\theta) = \begin{pmatrix} H_1(D_1) & 0 \\ \rho H_1(D_2(\theta)) & H_2(D_2) \end{pmatrix} \\ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \end{gathered} \tag{5.4}$$

$$V(\zeta|\theta,\beta,\phi)) = \begin{pmatrix} V_1(D_1) & \rho C_1(D_1, D_2(\theta))^T \\ \rho C_1(D_1, D_2(\theta)) & \lambda I_n + \rho^2 V_1(D_2(\theta)) + V_2(D_2) \end{pmatrix} \tag{5.5}$$

where ϕ indicates the collection of all correlation hyper-parameters, λ denotes the observation error and ρ a regression coefficient scaling the simulation output. Note that now there are additional subscripts on the basis functions, which may differ between the emulator and discrepancy function, and on the correlation and variance functions. A subscript of 1 indicates this parameter or function belongs to the base emulator, and a subscript of 2 indicates the discrepancy function.

Kennedy and O'Hagan (2001a) proceed by first estimating the base emulator parameters from the simulations outputs, ignoring the discrepancy and observations. Then, conditional on these values, they obtain approximate expressions for the mean and variance of the distributions of the remaining parameters ρ and λ and discrepancy hyper-parameters by integrating over the prior distribution of θ, $p(\theta)$, assuming this to be normal. This distribution is maximised to obtain plug-in estimates for the discrepancy roughness, mean parameters and variance. Finally, conditional on all these estimates, sampling methods are used to draw from the posterior distribution of θ (Kennedy & O'Hagan, 2001b).

For this work, MCMC methods are used instead to estimate all parameters in a similar manner to Qian & Wu (2008), and as suggested by MUCM (2011). More specifically, Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS) implemented by `stan` sampling software are used (Neal, 2010; Stan Development Team, 2015; Gelman et al., 2014). HMC utilises derivatives of the log-posterior at each sampling step in order to inform the choice of the next point, thus traversing the posterior more quickly than in random-walk Metropolis-type samplers (Gelman et al., 2014).

5.3.1 Model and Prior Specification

A Latin hypercube sample of 100 points is initially constructed across the 5 calibration parameters, with parameter range end points chosen to represent the most extreme values that still seem possible. All inputs are scaled to lie between 0 and 1. The simulator outputs are plotted in Figure 5.1. Age is treated as a control parameter, and because it is vital to our output, it is oversampled. This is done by repeating the design across the 4 quartiles of the range of age considered (1-70), giving a total of 400 input points, with 4 times as many unique age values as for the other variables. More than 400 inputs leads to more unwieldy calculations, as the matrix inversions needed with each Hamiltonian step scale with $\mathcal{O}(N^3)$; that is, the number of calculations required to invert the matrix is proportional to a cubic function of the number of input points, for any reasonably large N, so that the computation time slows down disproportionately as we add more points.

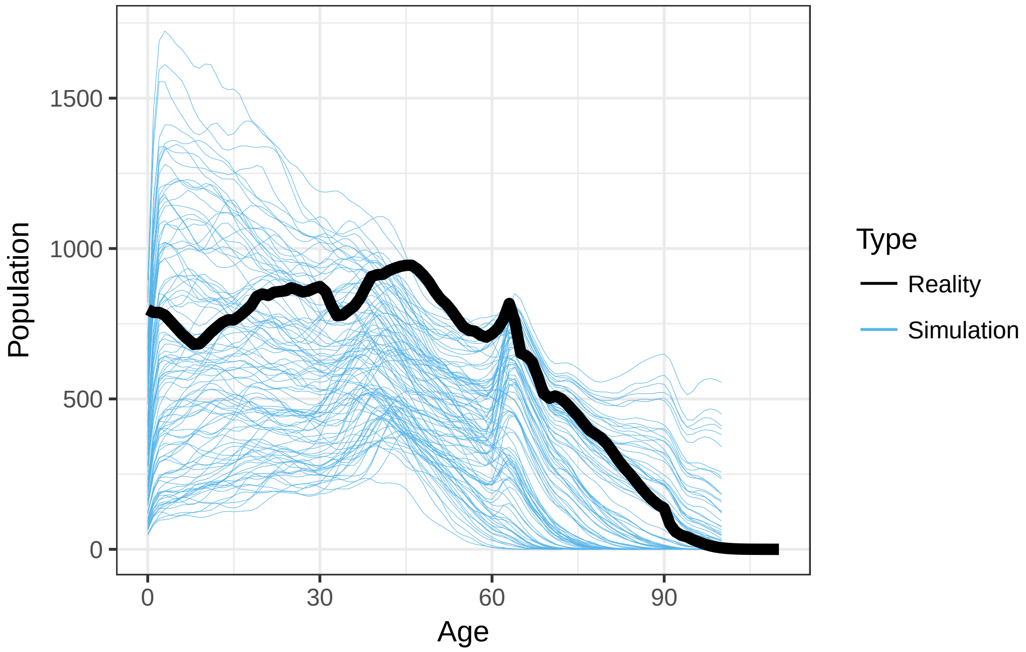

FIGURE 5.1: Plot of population by age for initial LHS sample, compared to observed values for the UK in 2011, scaled appropriately

In terms of the model itself, some simplifications are made from the Kennedy & O'Hagan (2001a) formulation described above. Firstly, ρ is chosen to equal 1, as the observations used to calibrate against are scaled to reflect the difference between the simulation starting population and the UK population size in 1950, and so we expect the simulation outputs and target values to be of the same scale. The observation error λ is set to zero in this case, as it is expected to be very small in relation to the simulation stochasticity and discrepancy and so only adds complication in this case. Simple linear mean functions are used for the emulator, as before, while the discrepancy function is chosen to be zero-meaned.

Simulation and observations values are scaled down further by the same factor so that their maximum is around 10 to avoid overflow problems and to aid prior specification; a more principled approach would be to scale so that their mean is zero and standard deviations equal to unity (Kennedy, 2004), and future work will take this approach. Vague normal priors are set for all mean function coefficients, while following Gelman (2006), half-Cauchy priors are used for the variance and roughness parameters. These are Cauchy distributions centred on zero, but constrained to be non-negative. Scale parameters of 5 are used for the emulator variance and roughness parameters, giving them extensive tails, while 1 is used for the emulator nugget parameter, reflecting the expectation that it will be considerably smaller than the variance of the function itself.

The exception to use of Cauchy priors is the discrepancy function variance; this is much more tightly specified using a half-normal distribution with mean 0 and variance 2. This is a non-standard prior for a variance parameter, but it is used to reflect the expectation that the simulator will be able to replicate the real values relatively closely. Longer-tailed or vaguer priors have been seen to lead to identification problems, with model discrepancy explaining more of the variance than is reasonable given what is known about the simulation and the system under study. Finally, the calibration parameters are modelled as uniform over the range of sampled values, reflecting agnosticism as to their probable values.

5.3.2 Results

Three parallel Hamiltonian Monte Carlo chains were run in `stan`, each for 1000 iterations and 500 runs for each chain were used for adapting the HMC leapfrog step size and other sampling parameters, while the last 500 where kept for posterior inference. The potential scale reduction convergence diagnostic of Brooks & Gelman (1998), known as $\hat{R}$ (Rhat in the table below) is used to assess convergence. This is based on the idea that, upon convergence, samples from each chain should be indistinguishable, and likewise, samples from the beginning and end of an individual change should also have identical properties. The $\hat{R}$ diagnostic assesses the extent to which this is the case, and should approach 1 for converged chains. A value larger than about 1.1 suggests more samples might be advisable (Gelman et al., 2014).

The table below shows posterior summaries for the model parameters, where *sigmaf*, *sigmad* and *sigmaN* describe the emulator, discrepancy and nugget variance respectively, and *n_eff* corresponds to the effective posterior sample size accounting for auto-correlation.

TABLE 5.1: HMC sample summary

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
betas[1]	3.31	1.55	0.14	3.38	6.33	675.7	1
betas[2]	-2.2	1.85	-5.9	-2.19	1.75	696.7	1
betas[3]	-1.55	0.7	-2.85	-1.56	-0.13	914.1	1
betas[4]	-0.63	0.99	-2.49	-0.67	1.39	1102	1
betas[5]	6.19	0.88	4.44	6.22	7.88	1122	1
betas[6]	0.84	0.99	-1.1	0.82	2.81	1200	1
betas[7]	0.01	0.89	-1.72	0.01	1.8	1389	1
sigmaf	7.06	2.01	4.04	6.68	11.74	473	1.01
omegas[1]	24.22	5.16	17.34	23.16	37.99	435.5	1.01
omegas[2]	0.32	0.09	0.17	0.31	0.53	966.1	1.01
omegas[3]	0.82	0.21	0.45	0.81	1.27	969.5	1

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
omegas[4]	0.63	0.15	0.39	0.61	0.96	557.6	1.01
omegas[5]	1.09	0.24	0.7	1.06	1.63	1007	1
omegas[6]	0.64	0.19	0.36	0.6	1.11	838.1	1
sigmaN	0.06	0.01	0.04	0.06	0.08	739.3	1.01
omegas2[1]	13.76	23.42	0.46	2.99	83.49	135.2	1.01
sigmad	2.22	1.2	0.44	2.04	5.16	395.6	1
thetas[1]	0.46	0.27	0.03	0.46	0.93	397.4	1.01
thetas[2]	0.35	0.2	0.04	0.32	0.82	360.1	1
thetas[3]	0.57	0.13	0.3	0.58	0.8	302.7	1.01
thetas[4]	0.53	0.22	0.12	0.5	0.97	369.5	1.01
thetas[5]	0.11	0.12	0	0.07	0.46	290.9	1
log prob.	31.8	3.9	23.32	32.16	38.57	293.1	1.01

Assessed by the $\hat{R}$ metric, all samples appear to have converged to an acceptable extent. Visual inspection of the sample trace-plots (Figure 5.2) seem to indicate reasonable mixing and stationarity of the chains for *theta*, with similar patterns evident across other parameters. The first 14 rows of the table down to the estimate for *sigmaN*, the nugget variance parameter, refer to the underlying emulator. As a robustness check, estimates of these parameters were obtained with only the simulator training dataset using the methods described in the previous two chapters, and these closely matched the mean of the posterior samples summarised in the first column of the summary table. The parameters of the discrepancy function *sigmad* and *omegas2[1]* inform us about the mismatch between the simulation and reality. The value of the discrepancy variance *sigmad*, effectively the range of this mismatch, is a significant proportion of the total emulator variance, with a 50% central interval between 1.32 and 2.97. The roughness parameter *omegas2[1]* has a rather more skewed distribution, with most of the density below 10, indicating relatively smooth discrepancies, but with the possibility of much rougher surfaces.

The aim of this process is to inform about probable values of the calibration parameters *theta*. Referring back to interpretation of these parameters, *thetas[1]* and *thetas[2]* relate to the Gompertz mortality function in Equation 5.2, while *thetas* 3 to 5 correspond to a, b and c in the Hadwiger fertility function (Equation 5.1). From the posterior summaries and the density plot in Figure 5.3, it appears that our calibration procedure has been rather more informative about the fertility parameters than about mortality; positive probability is assigned across the entire possible range of *thetas* 1-2 and the distribution of *thetas[1]* is particularly flat. In contrast, more is revealed about the fertility function, with more precise posteriors evident for *thetas[3]* and *thetas[5]*.

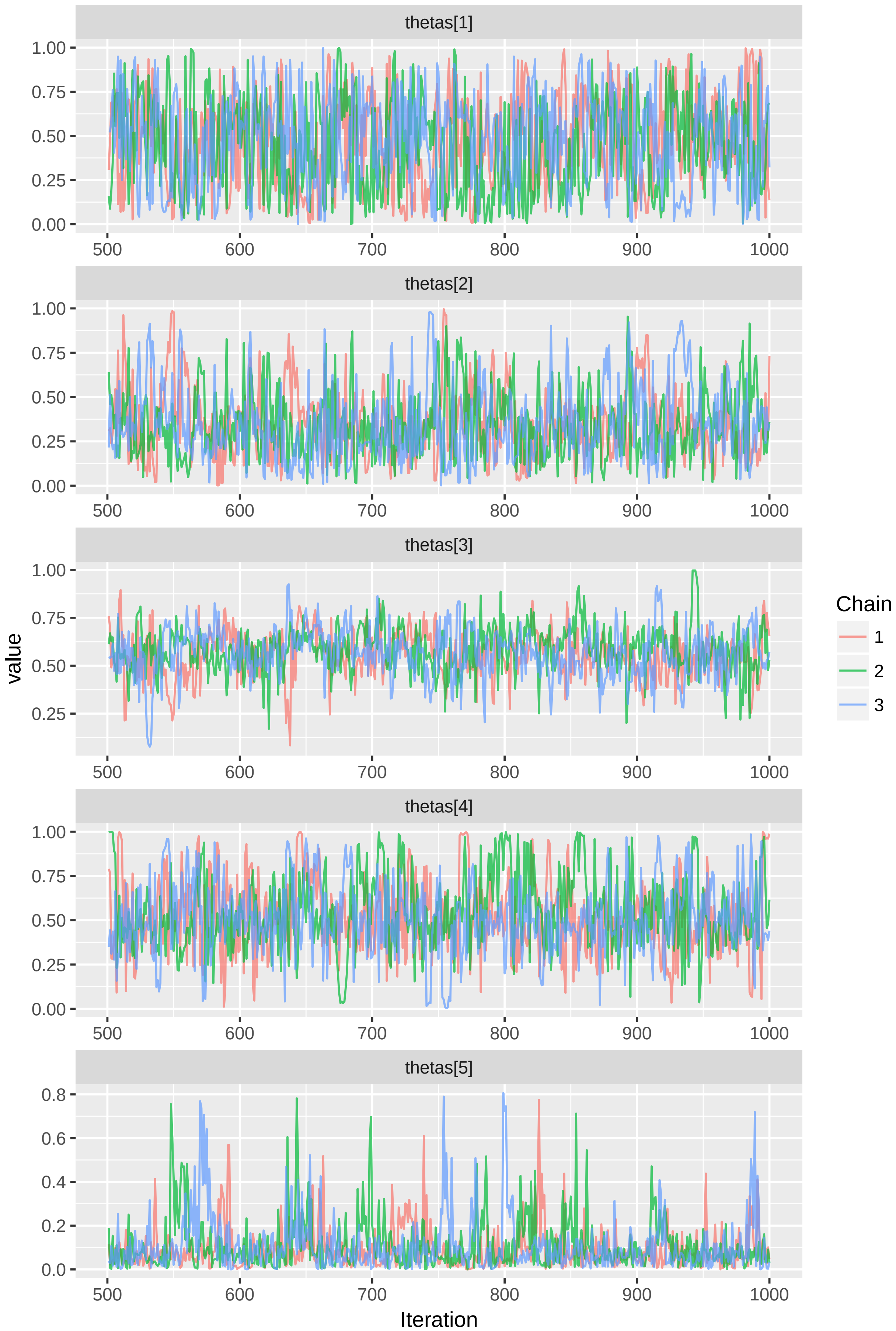

FIGURE 5.2: HMC sample trace plots for calibration parameters theta

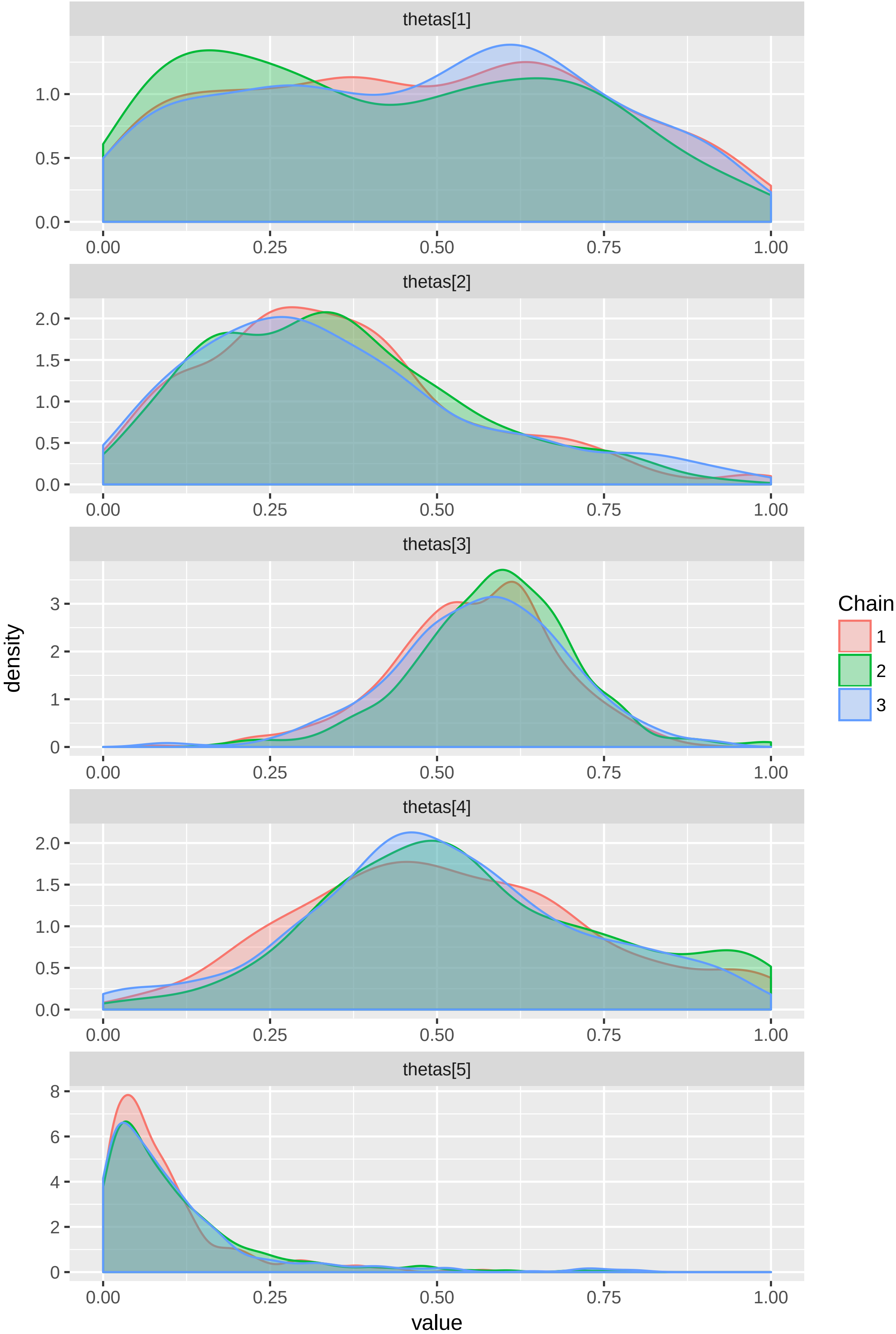

FIGURE 5.3: Posterior density plots for calibration parameters theta

5.3.3 Calibration Assessment

The aim of a calibration process is to attempt to reduce uncertainty about the possible input values that may have generated observed patterns, taking into account possible model discrepancy. Although posterior distributions for the calibration parameters θ have been obtained, these have given us almost no information about the mortality function. Even for fertility, we have not reduced our uncertainty to a great extent, considering that the original input ranges were chosen to include all parameters values thought remotely possible. This is largely because not all the available information about the age distribution is utilised.

With outputs for only 4 age points per design point, our uncertainty about the rest of the age distribution is relatively high, meaning we can not learn much about our input parameters. This is particularly true for mortality, which only has a discernible effect for the upper reaches of the age distribution. Recall we only take one sample in the upper age quartile (greater than 54) per design point, and so we may learn next to nothing about old-age effects for many combinations of input points.

This discussion suggests that an alternative approach may be required that is more flexible about how output measures are chosen and sampled. This section has assumed a Gaussian correlation structure across age, and treated it as an input parameter to the emulator. The next section discusses a 'History Matching' approach (Vernon et al., 2010) that builds multiple emulators corresponding to different areas of the age distribution, and for the purposes of calibration, ignores the correlations between them.

5.4 History Matching

History Matching is so called because of its historical use in inference about oil reserves; geospatial models of oil location, volumes and dynamics can be iteratively calibrated until model outputs match histories of oil well pressures and production (Cumming et al., 2012, Andrianakis et al. (2015)). This approach has been used to calibrate models of galaxy formation (Vernon et al., 2010), traffic congestion (Boukouvalas et al., 2014), and disease transmission (Andrianakis et al., 2015), to give just some examples. The last of these was a pure Agent-Based Model, suggesting the technique may be appropriate for use in demographic ABM.

5.4.1 Description

History matching makes use of an 'implausibility' metric that gives values to input points that reflects how safely we can rule out that a given input point has generated the observed outputs, given our uncertainty about simulation and measurement error, together with our assumptions about model discrepancy (Andrianakis et al., 2015).

A fitted emulator (or emulators) is used to calculate this quantity for a large range of possible calibration parameter values. Formally, the implausibility is defined as (Andrianakis et al., 2015, p8)

$$I(x)^2 = \frac{(z(x) - E(d(x)) - E(f(x)))^2}{Var(f(x)) + Var(d(x)) + Var(\epsilon(x))} \tag{5.6}$$

where $I(x)$ is the implausibility at input points x, $Var(\epsilon(x))$ represents observation error, $E(d(x))$ and $Var(d(x))$ are the discrepancy mean and variance respectively, and the former is often assumed to be zero. Thus, implausibility is a function of the squared distance between observation and emulator prediction, scaled by the uncertainty. For more than one output, the multivariate equivalent is defined as (Andrianakis et al., 2015, p8):

$$\begin{aligned} I(x) =& (z(x) - E(d(x)) - E(f(x)))^T (V_{all})^{-1} (z(x) - E(d(x)) - E(f(x))) \\ V_{all} =& V(f(x)) + V(d(x)) + V(\epsilon(x)) \end{aligned} \tag{5.7}$$

with V now representing matrices rather than scalars.

Once these values have been calculated, any values that fall outside a predefined cut-off point are rejected as implausible. The value of the cutoff point can be determined by choosing an appropriate quantile from the upper tail of a chi-squared distribution with degrees of freedom equal to the number of outputs is suggested by Williamson et al. (2013). This rejection of implausible points generally greatly reduces the 'non-implausible' area of the input space [2]. An additional 'wave' of simulation runs from the reduced space can then be taken, and a new emulator built. These steps can be repeated until a plausible subset of the input space is identified. This process of *iterative refocusing* can act to calibrate the simulation, although unlike in the previous section, a distribution over the calibration parameters is not obtained. This approach seems well suited to demographic ABM applications, as it is relatively intuitive, easier to apply than the full Bayesian equivalent and also will gives a well-defined response in the case that no outputs fall within the model discrepancy tolerance, in contrast to the Kennedy & O'Hagan (2001a) method (Williamson et al., 2013; Brynjarsdóttir & O'Hagan, 2014).

5.4.2 History Matching Specification

In order to capture the response of the simulation across the age distribution, emulators are constructed at 15 evenly spaced points from age 2 to age 72. The same initial

[2] Non-Implausible or Not-Yet-Ruled-Out space is used in the literature rather than the more intuitive 'plausible', as a failure to rule out a space is not the same as considering it plausible. It may well be subsequently ruled out in future iterations of the process.

input design is used for the other parameters as in the previous specification. The correlations between the emulators are ignored, as is the case in Boukouvalas et al. (2014) and Andrianakis et al. (2015), meaning that the variance matrix $V(f(x))$ is diagonal. Although correlations will clearly exist between outputs at different ages, and their inclusion would result in more efficient reduction of the parameters space, they are ignored here for simplicity, as specifying or estimating these correlations is difficult in practice (Boukouvalas et al., 2014). Similarly, diagonal assumptions for the discrepancy variance are also used. While correlation between different outputs has been successfully elicited by Vernon et al. (2010), this is a complex process and is not attempted for this simple example.

A mean discrepancy function is included in this exercise, however. As discussed above, we expect our simulation output to be considerably lower than reality, as we are modelling the UK as a system closed to immigration, and we do not allow for the 'baby boom' of the mid-twentieth century. To account for this, discrepancy is specified to include two 'humps', one centred on the baby boom generation at 45.5 years, and one to reflect working age migrants, centred about 25 years. Simple squared exponential functions are used for each element. These are very simplistic assumptions, and more care would be taken to elicit meaningful prior distributions as in Vernon et al. (2010) and Oakley (2002) if calibrated simulation were to be put to subsequent use. The discrepancy variance was set to 200, reflecting the a priori intuition about what an expectation of simulator mismatch centred on the mean would be. The overall shape of the discrepancy function can be seen in Figure 5.4.

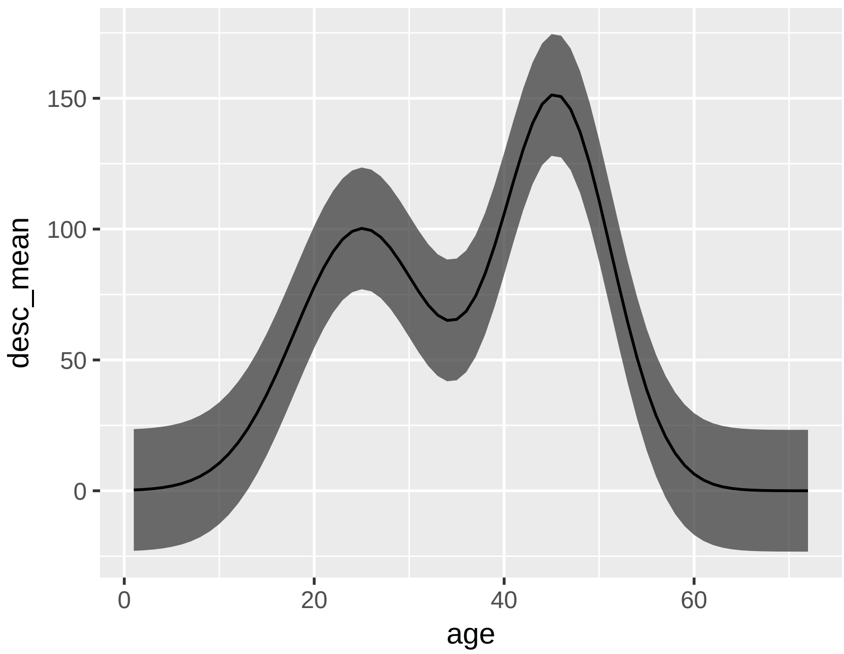

FIGURE 5.4: Shape of the descrepancy function

5.4.3 Results

Firstly, the 15 emulators were fitted independently, and the cross validation checks of Bastos & O'Hagan (2009) discussed above were applied. The Malanhobis distance for all but one emulator fell within a 99% expected interval based on the cross-validation sample. The final emulator was not far outside this range, and so was deemed acceptable for these purposes. Next, a sample of 100,000 input points was drawn uniformly at random from the ranges over which the emulators had been fitted. This is a somewhat smaller sample than is ideal, as Andrianakis et al. (2015) use 10^8 points, although the dimensionality was much higher in their case. Implausibility measures for these points were calculated using Equation 5.7, again assuming measurement error to be negligible.

The correlation plot below (Figure 5.5) shows the locations of the small number of non-implausible input points remaining after the cut-off was applied. Note the scales on the axis; the range of inputs has been greatly reduced by the use of history matching through just one wave. Note θ_2 and θ_5 in particular, the reduced range matches with that proposed by the full calibration methods in the previous section.

Figure 5.6 plots the estimated emulator means for a sub sample 1000 of 100,000 inputs, together with the location of the non-Implausible values following the first wave. These are in line with expectations in that they match up closely to the observations less the discrepancy function. An additional wave of simulations could now be drawn to get a more accurate picture of simulation behaviour in the reduced space, however, it seems likely that little further reduction in input space would be made.

5.5 Conclusions from two calibration methods

This chapter has demonstrated two methods for the calibration of demographic simulation, using the MicSim microsimulation of Zinn & Himmelspach (2009). The first utilises Hamiltonian Monte Carlo to sample from the posterior distribution of the calibration parameters, while the second iteratively rules out areas of the calibration input space that are deemed implausible. The latter, history matching, is preferred as a method for future calibration of Agent Based Simulation for a number of reasons.

1. History Matching requires fewer assumptions about the nature of the outputs in question. For instance, one could combine history matching with the use of a Poisson data model (Diggle et al., 1998; Rasmussen & Nickisch, 2015), or with heteroskedastic emulators as in Boukouvalas et al. (2014). This could conceivably be done for full Bayesian calibration, although it would involve a further computation within Hamiltonian Monte Carlo steps and may hinder convergence.

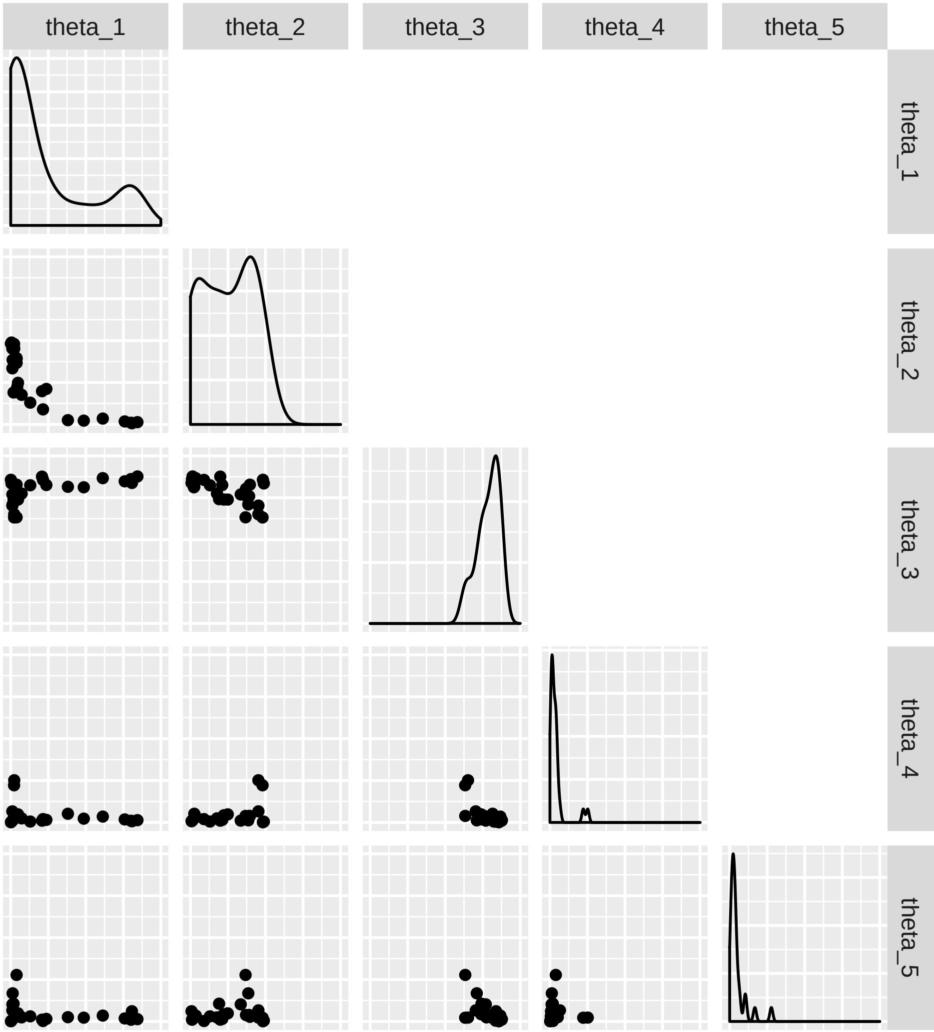

FIGURE 5.5: Plot of location of non-implausibile points from history matching calibration in [0,1] input space, from a sample of 100,000 points

2. History Matching provides an easier way to treat the key demographic variables of age and sex, particularly in the case were outputs by age are only available in grouped form. In this example, the use of continuous time microsimulation meant that population could be sampled for one year age groups centred on any value of age. In practice observations are only likely to be available for age groups, and simulations are often discrete time and may only output agent age by year. While inputs to the full calibration model must generally be continuous, by using separate emulators for relevant age group, and potentially for each sex, this problem can be side-stepped.

3. History Matching is much computational quicker, because Hamiltonian Monte Carlo sampling is not required. The original Kennedy & O'Hagan (2001a) method did not utilise Hamiltonian Monte Carlo, but did use other numerical methods to approximate the posterior distributions.

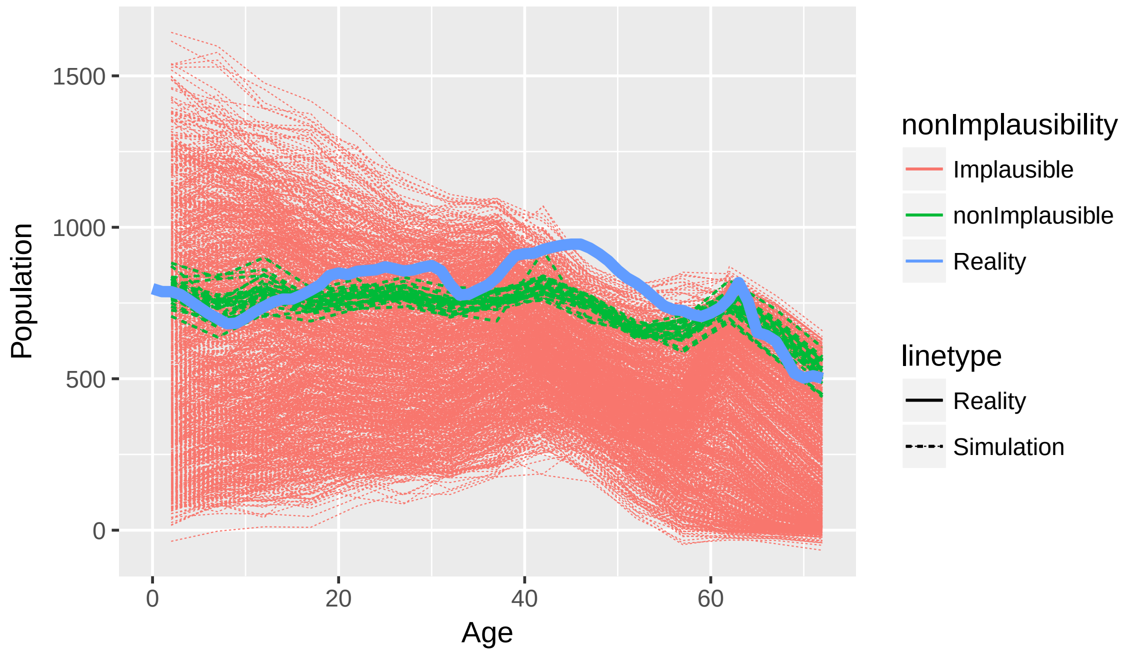

FIGURE 5.6: Plot of sub-sample of implausible and non-implausible emulator predictions from 100,000 simulated input points

4. With History Matching, it is considerably easier to specify discrepancy function priors, particularly if cross-output correlations need to be specified. In the Kennedy & O'Hagan (2001a) model, priors need to be specified for a discrepancy Gaussian process, which are not easy for layman to interpret, although Oakley (2002) do provide methods for doing just this, and Brynjarsdóttir & O'Hagan (2014) also discuss how a such specifications might be made in terms of specifying constraints and derivatives.

Of course, the two methods are not mutually exclusive: one option is to attempt to first reduce the range of inputs considered and then to attempt full calibration (Andrianakis et al., 2015). Furthermore, the Kennedy & O'Hagan (2001a) method does have advantage of obtaining probability distributions that can then be used for calibrated prediction. A final note of caution is to point out the calibrating a model does not guarantee that it is valid (Oreskes et al., 1994); it may be that model misspecification means that the calibrated model generates the right output with the wrong mechanism. This is case for the microsimulation example given here, as a consequence of the poor fertility model that does not differentiate between parity. Plotting the fertility rate functions generated by the modes of the calibrated distribution against their empirical rate equivalent shows how poorly the calibration has performed (Figure 5.7); however this is more a consequence of the simple (but demonstrative) choice of model than the calibration process itself.

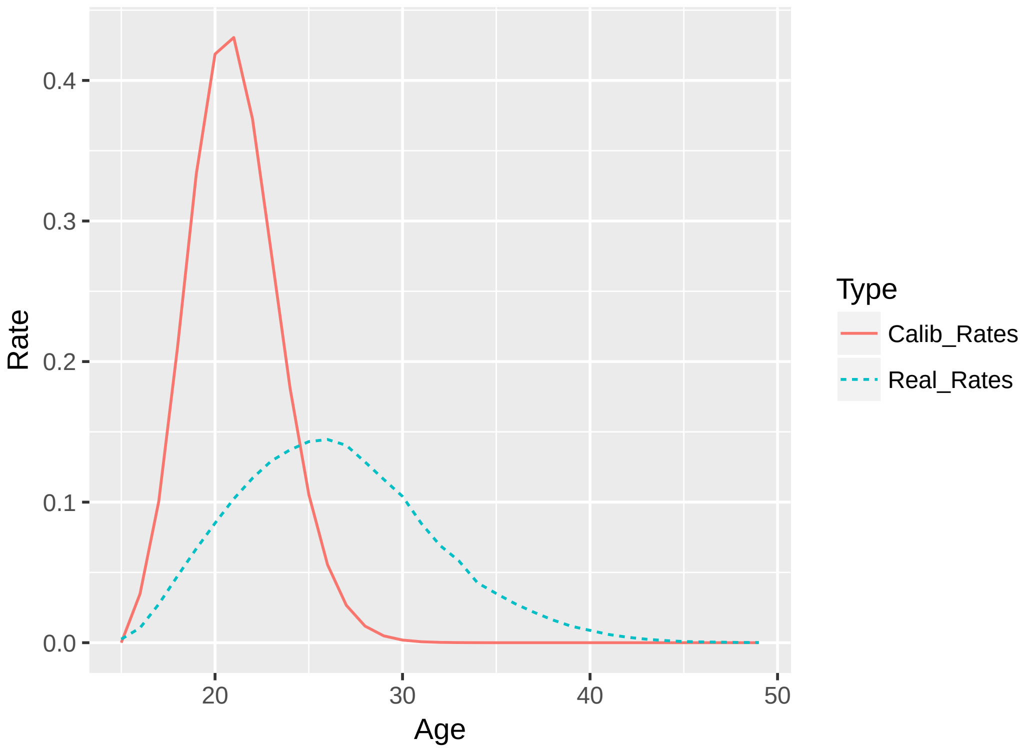

FIGURE 5.7: Plot of mean calibrated fertility rates against empirical 1980 UK equivalent

6

An Emulation Case Study: The Easterlin Effect

6.1 Motivating an Agent-Based Model of Intergenerational Ferility

The preceding segments of this thesis have demonstrated the efficacy of Gaussian Process Emulators in relatively abstract, 'toy' cases. This chapter puts them to use in a substantive case study, with the aim of calibrating an Agent-Based Model of intergenerational fertility. This model aims to examine the plausibility of an existing theory which relates cycles in macro-level fertility to an individual-level desire to delay childbearing until a level of well-being commensurate with that achieved by one's parents (Easterlin, 1987).

This phenomenon (dubbed the 'Easterlin effect' (Pampel & Peters, 1995)) does not occupy the minds of many fertility researchers at present, because it is of mostly historical interest; it is deemed to have operated in a period in the early 20th century to the late 1970s. However, it is an interesting target for a simulation in this manner because it is a simple theory that explicitly describes links between the micro- and macro-levels in a demographic process, and because it has clear implications which can be tested (the generation of cycles in fertility). Furthermore, the 'relative well-being' mechanism through which the childbearing decision relies on individual relationships between agents - specifically, those between parents and children. Modelling this mechanism directly in an Agent-Based Model, although not without its own challenges and shortcomings, avoids some of the specification and endogeneity problems found in more traditional empirical approaches to the problem, highlighted in review by Macunovich (1998) and Waldorf & Byun (2005), while recognising micro-foundations not present in the mathematical models of Lee (1974), Wachter (1991) and others.

However, as previously argued, attempting to model a complicated social phenomenon of this nature often requires the inclusion of several interconnected processes, each of which is governed by its own set of parameters, and each of which is a potential source of error and uncertainty. Thus, the use of statistical emulator techniques is important for allowing a systematic analysis and calibration of the simulation described in the subsequent sections.

6.2 Description and Theory

The Easterlin effect purports to explain the existence of distinctive wave-like patterns in the time series of births in the United States. It is associated with the the work of Richard Easterlin, who developed his theory in a series of articles and books from the late 1950s onwards (e.g. Easterlin, 1962, 1966, 1975, 1987). These patterns are different from the expected generational cycles, whereby large cohorts have relatively more children than smaller cohorts simply because they include more prospective parents, and thus 'echoes' of larger cohorts are seen in population age structures at intervals of approximately a generation's length.

Instead, Easterlin is describing waves of double this period, whereby a child's fertility is likely to be negatively correlated with that of his parents' generation, but positively correlated with his grandparents'. Examining Figure 6.1, net of the upward trend, broad arcs are visible in the plotted time-series of births, until some flattening off after 1980.

Easterlin posits that generational cycles such as those described are the result of a link between birth cohort size and later fertility, a link that is caused by reduced opportunities in many spheres for those in larger cohorts, and thus an increase in the probability that those in such cohorts will not feel well-off enough to start a family.

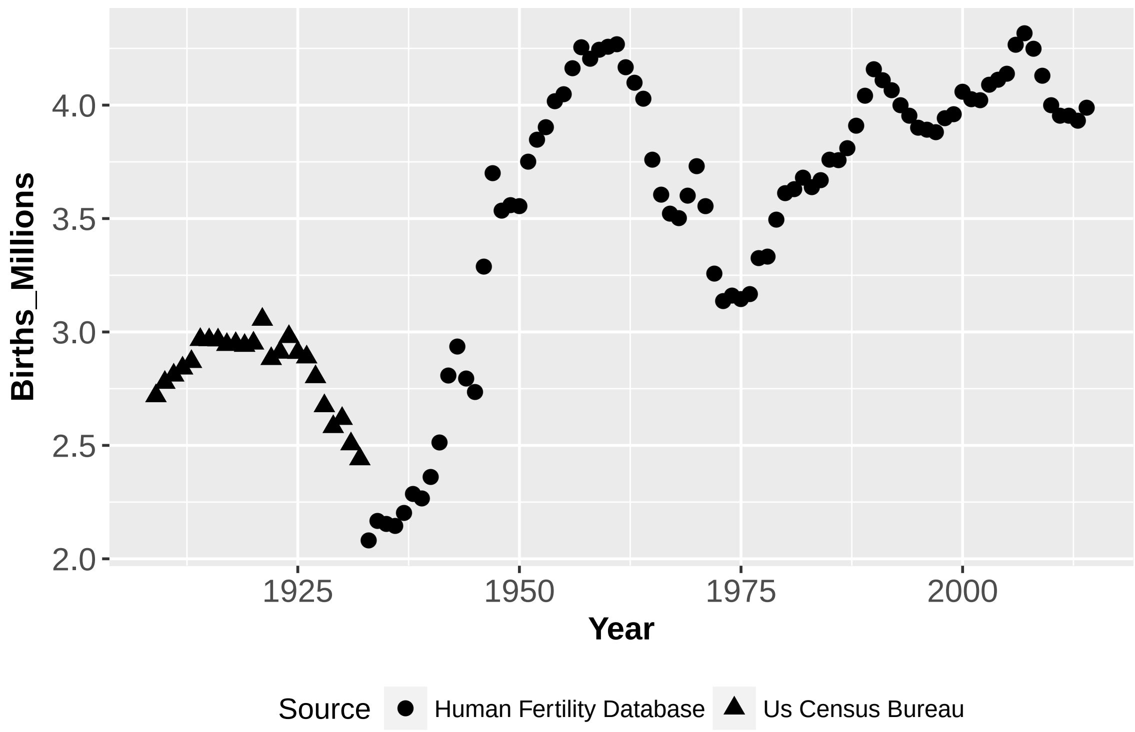

FIGURE 6.1: Time Series of US Births since 1909. Source: US Census Bureau, 1975; Human Fertility Database, 2015

To breakdown his thesis further, it is possible to identify four specific assumptions upon which it rests (Easterlin, 1987).

1. Individuals wish to start a family only once they have reached a level of well-being that satisfies their aspirations.
2. These aspirations are set relative to some reference group, rather than in terms of some absolute standard of well-being.
3. The relevant reference group against which individuals set their aspiration is their parents.

4. Individual well-being is negatively correlated with cohort size, in particular due to increased competition in the labour market.

This last statement contains a subsidiary assumption about the labour market, namely, that older and younger workers are imperfectly substitutable, or else relative cohort size would not effect the success of young workers (Macunovich, 1998). The fact that older workers are likely to have obtained greater experience in their chosen career (equivalently - developed higher levels of job- or trade-specific human capital) goes some way to laying the ground for such an assumption; often, higher level jobs require experience to be performed effectively.

Easterlin also hypothesises that competition and crowding in two other areas of life may lead to poorer outcomes for those in large cohorts (Easterlin, 1987). Firstly, larger families lead to greater divisions of parental resources, both monetary and in terms of time and attention. Secondly, crowding in educational establishments may also be problematic; if school capacity building and teacher training lags behind demand, then it is likely that smaller cohorts will be at an educational advantage. However, this chapter focuses only on intra-cohort competition as it manifests itself through the labour market. Easterlin's hypothesis was somewhat heterodox in economic theory at the time because, rather than taking consumption preferences as given, it considered them as malleable and something to be explained (Easterlin, 2004). Preferences are described as determined through a process of socialisation - being exposed to certain levels of consumption during later teen-hood make one desire to attain at least this level of well-being.

The combination of these factors, then, is hypothesised to results in cycles of two generations in length, as small cohorts have little job market competitions, and therefore outperform their parents' generation, and therefore give birth to relatively more children. The resulting larger cohorts are then less successful due to the increased competition for jobs and resources they face, and therefore are condemned to lower average fertility, and so on (Easterlin, 1987).

The relative lack of immigration during the period in question due to strict immigration laws is also described as providing the conditions necessary for such cycles to develop (Easterlin, 1978). If immigration was subject to few restrictions, labour shortages due to smaller cohorts would not lead to a bidding-up of wages, but instead to a greater inflow of immigrants to meet the demand. Similarly, an relatively large cohort might lead to a cessation of such flows, leading to little change in the experience of labour market conditions between cohorts, regardless of size.

Cohort sizes are also described as having implications for female labour force participation, albeit in the opposite direction from those that held for men.The change in sign is a reflection of the reality of the more detached role of women in the labour market in the mid-twentieth century (Easterlin, 1978), largely due to the limited opportunities available to them as a result of discrimination and oppressive social norms. Older and younger women were supposed more substitutable as labour sources due to the 'non-career' roles they were typically obliged to take. Thus, scarcities in young male labour may lead to better employment prospects for them, but may lead to a withdrawal from the labour market of their spouses, as they were expected to devote themselves to motherhood, and an increase in employment among older women, who may be incentivised by higher wages to fill the gaps. Conversely, poorer prospects for the spouses of younger female cohorts may require them to work more - particularly as they are less likely to start families, with knock on effects for the demand of older female age groups.

6.3 Empirical Evidence in the literature

In part due to its apparent success in describing the American fertility boom and bust, Easterlin's theory has received a lot of attention in the demographic literature. Broadly speaking, this work can be split into two parts; studies that focus on the macro-level, looking at demographic and economic time-series and mathematical models of the process, and those which centre on the micro-level, and use survey data to investigate the relationships between parent's wealth, income, cohort size and fertility. Pampel & Peters (1995) and Macunovich (1998) provide comprehensive reviews of such studies.

6.3.1 The Easterlin Effect at the Macro-Level.

Evidence for the Easterlin hypothesis is mixed. At the macro-level, in terms of empirical data, it is clear that since the early 1980s the observed cycles have vanished, as can be seen in Figure 6.1. A number of reasons for this breakdown have been suggested. Increases in immigration and unemployment together with less secure working arrangements, more temporary work and increased female labour force participation are just some of these (Pampel & Peters, 1995). Indeed, Easterlin's original analyses (e.g. 1962, 1967, 1978) were very much concerned with the interplay of cohort size, immigration, and labour force conditions in different sub-groups in determining fertility, due to his initial association of fertility cycles with immigration-linked Kuznets waves in economic activity.

Despite this later failure of the theory's predictions, the relationship between cohort size and later fertility appears to hold strong for much of the twentieth century - fluctuations in fertility from about 1900-1970 follow relatively closely those of lagged cohort size (Wachter, 1991). Using birth-order life-tables constructed by the Human Fertility Database Project, it is possible to examine how fertility rates differ by birth order as the swing from high fertility to low fertility progresses between 1963 and 1976 (Figure 6.2).

Birth-order life-tables describe the rate at which those with n children transition to having $n+1$ children, based on data for single period. Those with n children are described in demographic terminology as being at *parity* n. Note that rates at all parities decline over the period, so that if relative income is responsible for the sharp downturn, it is likely to act on all decisions to have a child, not just the initial transition to parenthood.

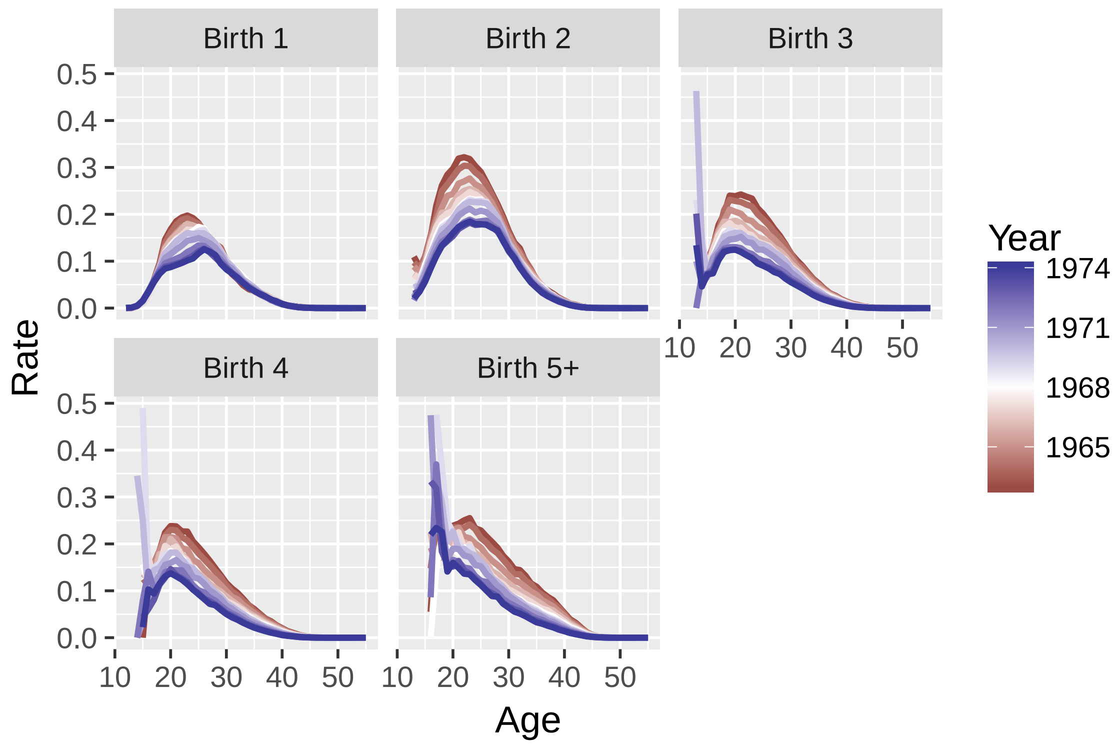

FIGURE 6.2: US Birth Rates by Birth Order. Source: Human Fertility Database, 2015

A considerable amount of work has been done creating macro-level mathematical models of the Easterlin process, most notably by Lee (1974), Frauenthal & Swick (1983), Wachter & Lee (1989) and Wachter (1991). These examine formal, dynamical systems models where fertility rates are damped by larger cohorts and similarly boosted in the presence of smaller ones. Thus, fertility is affected to various extents by the past course of births, allowing deviations from the equilibrium path (Lee, 1974; Wachter & Lee, 1989; Wachter, 1991). In general, these models involve formulations of the type shown in Equation 6.1.

$$B(t) = \int l_s m_s B(t-s) M\left[\frac{B(t-s)}{e^{r(t-s)}}\right] ds \tag{6.1}$$

where $B(t)$ denotes the number of births at time t, l the survivorship ratio, m a fixed component of fertility, and r the equilibrium growth rate. The significant part of the model is the expression $M[.]$, which is a function of the sizes of past cohorts $B(t-s)$ relative to the equilibrium trend $e^{r(t-s)}$. Various forms of this function have been analysed to see if the resulting system can sustain limit cycles, and whether the parameters on $M[.]$ are similar to those estimated from US data (Lee, 1974; Wachter & Lee, 1989; Frauenthal & Swick, 1983). Wachter (1991) analyses a family of models of this nature, with fertility depending on either the whole of the labour force or on various subsections or cohorts. In general, he suggests that the existence of self-generating Easterlin cycles is possible, but that it is quite unlikely that the macro-forms specified are responsible for the patterns observed in the US (Wachter, 1991).

A study by Waldorf & Byun (2005) reviewed macro-level studies in a systematic manner using the tools of meta-analysis. The findings were mixed, with supportive results being more likely in the US, and highly dependent on the control variables used and the specific functional forms used. Interestingly, they suggest that (publication) biases towards negative findings (disproving the thesis) were more likely the more recently the study was published, while for older studies, the opposite bias was suggested. Furthermore, the authors found that the use of income as a control or explanatory factor substantially changed the direction of results, and highlighted endogeneity concerns in the use of this variable.

The studies discussed above utilise purely macro variables and do not discuss the underlying mechanism relating to fertility damping. Furthermore, they generally assume structural heterogeneity (that is, the parameters remain constant) across time.

6.3.2 The Easterlin Effect at the Micro-Level.

A large number of studies based on survey data have been carried out examining whether the Easterlin effect holds at the micro-level. Macunovich (1998) provides a review of these studies (together with some macro-analyses), and suggests that the evidence is mixed, and many studies do not find an association between relative income and fertility, particularly outside of the U.S. However, she believes that these negative findings are due in part to data problems and specification errors in the models used to test the theory. In particular, measurement of past parental wealth or income is often difficult, and where it is present as a data-set variable it is often only available for one parent. Furthermore, the assumption that a simple threshold, defined by material well-being in teen-hood, will suffice as the operationalisation of the theory is also critiqued by Macunovich; instead, some function of the past well-being may also be consistent with the theory.

More recently, research has been conducted into the possibility that happiness and satisfaction is often a prerequisite for family formation, Parr (2010), for instance found that fertility is related to prior satisfaction with life, although he does not discuss the cohort effect in determining such satisfaction. Teitelbaum & Winter (2013) describe a different perspective somewhat commensurate with Easterlin's thesis; that declines in fertility can be understood as a risk management strategy. Thus, even if not directly affected by the detrimental effects of being born into a large cohort, perceptions of the risk of suffering unemployment or stagnant wages may still relate cohort size to lower fertility indirectly.

Overall, there is no clear picture in the literature as to whether Easterlin cycles were indeed behind the boom and busts in fertility observed in the US between about 1900 and 1980, in part because of the difficulty in generalising from such a short-lived phenomenon, data problems, difficulties in specifying the model, and so on. This paper approaches the problem from the opposite direction, as will be discussed presently.

6.4 An Agent-Based Approach

The focus of this paper is on attempting to 'grow' Easterlin cycles from the bottom up (Epstein & Axtell, 1996). That is, by specifying the micro-level behaviour of agents in a simulation in line with the mechanisms suggested by Easterlin, we attempt to recreate the observed cyclical patterns. If we can succeed in doing so under reasonable assumptions, we can assume that Easterlin's micro-level specification is plausible. Thus, the research question is as follows: 1. What behavioural rules and micro-level conditions are sufficient to generate Easterlin-like waves in fertility?

6.4.1 Methodology

In order to examine the interplay between the population and individual level inherent in Easterlin's theory, and to capture the influence of population heterogeneity and intergenerational links on the relative well-being thesis, a discrete time agent-based simulation was built. This allows us to formalise the particular mechanism suggested by Easterlin, maintaining the specific links between generations that result in the cycles observed in the data.

The model simulates the life histories of agents as they are born, age, find jobs, start a family, and eventually die. The model is relatively simple; the idea is not to capture every element of social life with absolute fidelity, but rather to distil these elements to the essences which are required to study the question in hand. However, an attempt is made to include population heterogeneity, particularly in earnings, and explicit relationships between agents as integral parts of the model, in a way that is more difficult in other modelling paradigms.

The logic of the overall approach is as follows. The simulation represents a idealised abstraction of the hypothesis given by Easterlin; if it is successful in replicating the waves in fertility of similar period to those seen in the twentieth century, then we can consider the theory plausible. However, the process of building an Agent-Based Model is far from trivial. Many assumptions about specific encodings of individual behaviour must be made, and additionally, many of these encodings will involve parameters for which we do not know the true empirical value. This means that the simulation must be calibrated in an attempt to match the empirical results. This requires the use of statistical techniques, as finding distributions of well-fitting parameters in high dimensional spaces with non-linear response profiles, as is common in Agent-Based Models, can be difficult. Gaussian Process emulators are thus used in order to identify plausible parameter ranges given empirical observations, in line with work by Vernon et al. (2010) and Boukouvalas et al. (2014). The next section discusses the architecture and specification of the model in detail.

6.5 Overview of the Simulation Model

6.5.1 Principles and Process

Writing a simulation model inevitably requires many choices. The open-ended nature of the problem means a potentially infinite space of realisations of the concepts described by Easterlin and others. There are also many technical choices relating to the structuring of the code and the technology used. This section describes some of the choices made and guiding principles used for the development of the simulation under discussion. Most points will be straightforward to many who have developed simulations before or have programming experience.

The choice of programming language or modelling platform is foremost among these. Proprietary systems such as `simul8` and `anylogic` are to be discounted because these place a barrier to the scrutiny of the work; if scientists are unable to run the model themselves, they cannot properly question its assumptions. Object-oriented systems are also preferred, because they can more concretely represent the concept of an agent as a discrete entity, possessing state and the possibility of action and reaction to events.

Netlogo provides an intuitive and surprisingly powerful platform for agent-based modelling, and has the advantage of being widely used amongst in the ABM community. However, its limitations as a domain-specific language mean it is hamstrung against more full-fledged and general-purpose languages, although there remains the possibility of extending the model through underlying Java code (Rossiter, 2015). There exist a range of Java-based tool-kits for Agent-Based modelling, most notably Repast and Mason, which provide well-used and robust frameworks for the construction of simulations. Being constructed in a compiled language, they also have a considerable speed advantage over interpreted languages, but at the cost of greater verbosity and more development time.

The choice was made to code the simulation in the programming language `python` for several reasons. The ease of expression in the language and its interpreted nature means that more time can be spent testing and analysing than coding, and that the code should be relatively easy to understand. The existence of large numbers of third party libraries allows much preliminary analysis, plotting and data-processing to be undertaken in `python`. Finally, the familiarity of the author with the language also facilitates the process of developing the model.

A loose attempt is made to follow the principles set out by Rossiter (2015) regarding the underlying architecture. In particular, this requires separation between the 'domain model' and the supporting architecture responsible for running the simulation, manipulating parameters, collecting data on model state for analysis, and logging information for debugging, and a hierarchical separation between layers of code. Many improvements still need to be made to meet the standards set out by Rossiter (2015), however.

The need to easily vary simulation behaviour at run-time in order to conduct experiments as described in previous chapters was also a key consideration. The implementation of the model is also designed to be easily extensible and modifiable (Axelrod, 2003). This requires that good object-oriented principles are adhered to, so that functional elements of the model are described by separate classes[1] wherever possible ('separation of concerns'). This naturally forces the simulation to become more modular, allowing the swapping in and out of 'modules'[2] when required. This also facilitates automated unit testing of classes (Rossiter, 2015), although at present only very few tests exist for the simulation reported here.

Time is modelled as discrete, in the sense that simulation time is 'stepped' from one period to the next before each agent conducts their scheduled actions. This is in contrast to continuous time models, and in particular, discrete event simulations (e.g. Zinn & Himmelspach, 2009), where events can be scheduled at any point in (pseudo) continuous time, and the simulation steps from event to event. Although discrete event simulation has the advantage of avoiding problems regarding simultaneity, so that there are explicit orderings to events, stepping from period to period rather than from event to event is easier to implement. Additionally, randomised orderings of agent actions should eliminate most concerns over simultaneous action.

The extensive nature of the mechanisms Easterlin described as driving his cycles, incorporating labour market participation, wage setting, fertility, partnership and other elements besides, mean that a complete description in a single model is probably undesirable, at least in a project undertaken by a single programmer/analyst. The decision

[1]Classes are a programming concept central to object-oriented programming, and can be thought of as templates defining a set of characteristics (variables and functions/methods) which any object of that class will share. An object created according to such a template is then called an instance of that particular class. A more rigorous definition is given in Gamma et al. (1995).

[2]Here the use of the term *modules* is used informally and generally refers to a `python` class, rather than the more specific `python` concept of a module, which may contain several classes.

is thus made to treat certain elements of the model as exogenous, and in particular, the simulation is so engineered so that wages and unemployment are assumed to respond to cohort size exogenously, and in the way Easterlin described. This allows the focus to remain on the key element, that of fertility choice.

6.5.2 Simulation Structure

With the above descriptions in mind, a simulation was built to reflect the key assumptions of Easterlin's theory, namely, that fertility is related to relative income, which is in turn related to cohort size. This sub-section provides a high-level overview of the simulation structure, while the next section provides a detailed description of each element of the simulation.

There is an emerging consensus within the ABM community that simulation should be documented in a unified format, called the ODD+D protocol, with acronym referring to Overview, Design, Details + Decisions (Groeneveld et al., 2016). While admirable, this format is somewhat rigid, and ignores the considerable diversity of the field of agent-based modelling. Furthermore, it leads to considerable duplication, as some ODD+D descriptions are generally relegated to an Appendix, and do not form part of the main description of the model in the body of text. For these reasons, an ODD+D description of the model is not provided. Instead, detailed verbal descriptions of the workings of the simulation are provided, and the code itself is provided for replication purposes. The code is freely available as a `git` repository at `https://bitbucket.org/jhilton/easterlinphd`, and requires only `python` together with some easily downloadable libraries to run.

Figure 6.3 displays the overall structure of the code, with major classes displayed as boxes, with relationships between classes displayed with arrows. Solid arrows denote that one class contains or owns an instance of the another class, while dotted arrows indicate communication between classes. Classes for which multiple instances are created within a simulation (for instance, agents) are indicated in grey. The diagram is not intended as a formal Unified Modelling Language (UML) diagram, commonly used in software engineering, but just as an aid to understanding the details of the simulation.

Agents are created by the 'agent factory' class, which determines the agent's initial characteristics. Each agent has both a fertility and employment class. The separation of these processes into different classes contained within each agent instance makes it easier to extend the model and trial different specifications and sub-models relating to these processes, as these can simply be written as new fertility sub-classes, and the particular fertility model used in any given simulation can be chosen when running the simulation.

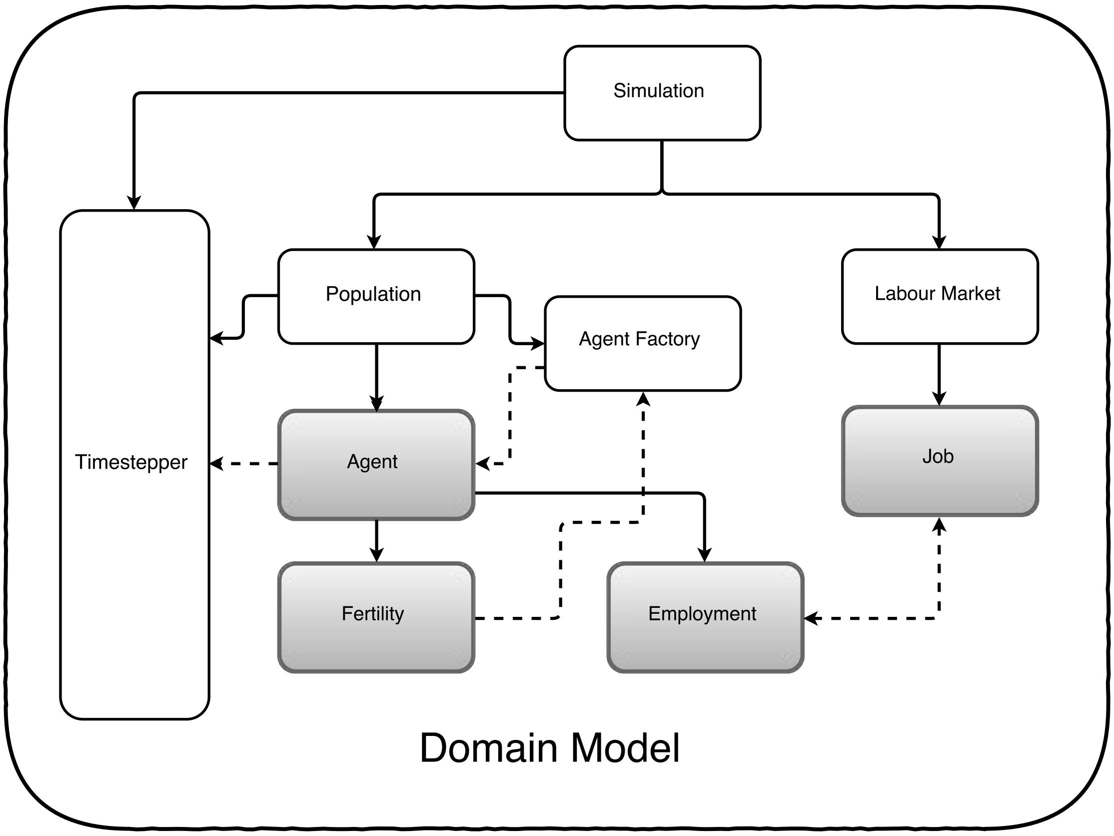

FIGURE 6.3: Schematic map of major classes in the simulation

As well as the simulation itself, a degree of supporting architecture is required to gather information from the simulation, and govern the running of simulation experiments. In line with Rossiter (2015) and with the principle of separating concerns, this code is kept separate from the 'domain model' as far as possible. Figure 6.4 describes these additional classes, which in principle could be re-used for other similar simulations. The experiment and design point classes define a set of simulations to be run in line with a particular experimental design (e.g. a Latin Hypercube Sample), and create corresponding parameter dictionaries (`python` objects from which values associated with particular "keys" or labels can be extracted) which are later used to run the simulation. The statistics collection class is also configurable so that instances can collect more or less data from the simulation as it runs.

6.6 Model Specification

A description of the key elements of the simulation is now offered, together with how they relate to the ideas in Easterlin's thesis. The simulation initialisation process is also described, together with a summary of how the model executes.

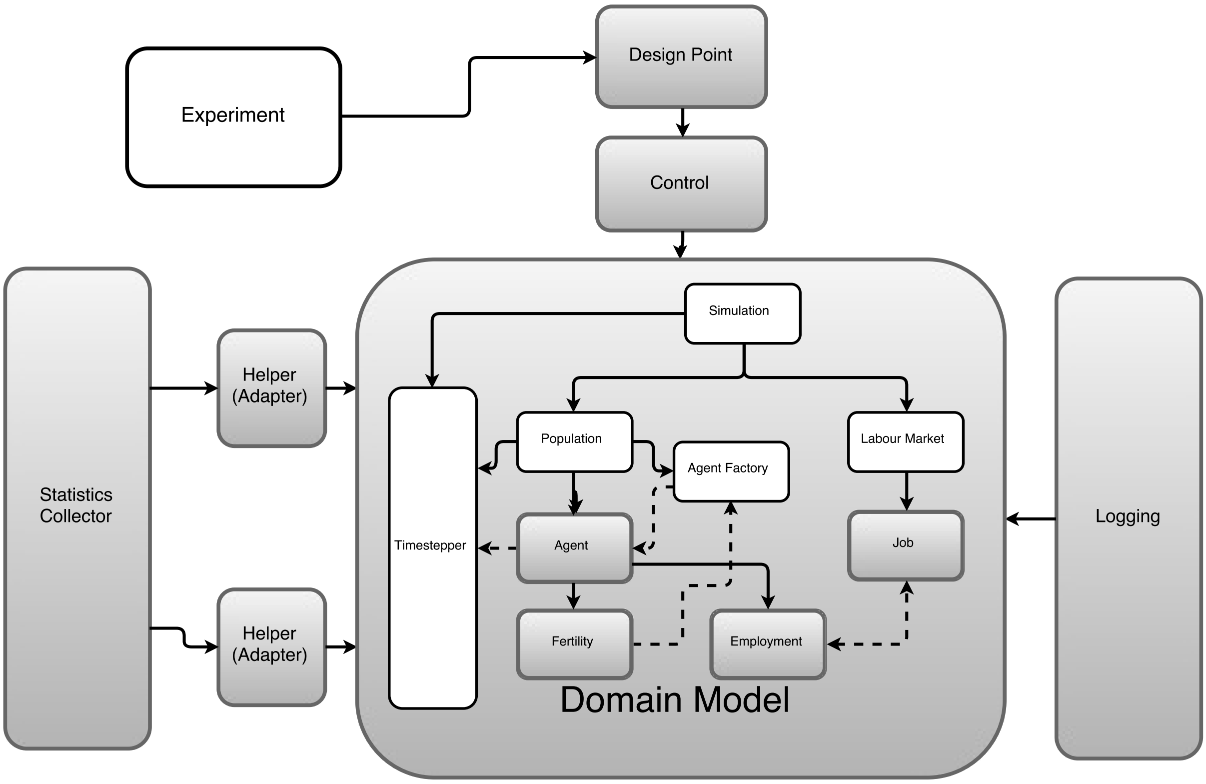

FIGURE 6.4: Diagram of classes governing experimental design and statistics collection.

6.6.1 The Agent

The basis of the model is the agent. Males and Females are represented by different sub-classes in the code, although they share many of the same methods. Instances of the agent class keep track of the key state variables in the model: age, the identity of an agent's immediate family and an agent's individual 'skill' level, reflecting their ability to earn a wage in the market. This is drawn from a normal distribution centered around 0.5, and truncated close to zero and at one:

The class also contains methods[3] relating to some of the exogenous demographic processes not of central interest to the research, including mortality and partnership. Mortality is described by a Gompertz function, as displayed in Equation 6.2, while mortality below aged 30 is assumed zero for the sake of simplicity, as it is expected that this will not make a noticeable difference to the outcome of the model. Differences in mortality by sex are ignored in the simulation, and while the Gompertz model parameters are configurable, they are kept fixed for the experiments detailed below.

$$m(x) = a_m \exp(b_m x) \quad \forall x : x > 30 \tag{6.2}$$

[3] A method is a set of instructions (a function) specific to a given class.

where a_m and b_m are further parameters describing the slope and intercept of the log mortality function.

Partnership is also considered exogenous, but is rather more complicated in its implementation. Following Zinn & Himmelspach (2009), agents enter the marriage market according to an age-specific hazard modelled by a double exponential function as shown in Equation 6.3. Once in the marriage market, agents undergo a matching process, whereby females in the market, by order of entry to the market, are able to choose the most compatible male partner by age and skill, determined by a Euclidean distance measure. The ideal degree of age difference between partners is a parameter in the model, but is held constant at 3 years for the purposes of this work. The age-specific schedule is also parametrisable, but is set up to ensure marriage is relatively early, in line with the norm in the early twentieth century, and to ensure that the major determinant of fertility is the relative income mechanism described by Easterlin, and not the rate of entry into marriage (although this could also depend on relative income and cohort size to some extent, as better-off males might be more suitable mates and more inclined to set up a family (Easterlin, 1987)).

$$n(x) = a\,\exp(-\alpha_n(x-\mu) - \exp(-\lambda_n * (x-\mu))) \tag{6.3}$$

6.6.2 Population

The population class exists mainly as holder for the collection of agents, and coordinates their actions and interactions. Agents are held as a `python` list object within the class, and this list is iterated through allowing agents to undertake their relevant yearly actions. The order of iteration is randomised every time-step so that no particular agent always 'acts' first.

6.6.3 Labour Market

More complex is the labour market. Earlier experiments with this simulation attempted to include labour market processes as an endogenous part of the model (see also Fagiolo et al., 2004), with firms and agents setting desired prices for labour based on their knowledge about supply conditions, which could be learnt from interactions with other agents. However, while this is an interesting avenue for research in its own right, it does not form the central focus of this investigation, which is more concerned with fertility decisions.

The labour market element of the simulation, then, is intended to reflect the ideal conditions under which Easterlin-like cycles might be expected to flourish. In particular, wages and employment opportunities are set up so that those in smaller cohorts are more likely to get jobs, and the jobs they do get will be better paid. Note that it is by no means certain that this is how the labour market does behave - while the supply of labour is expected generally to affect labour, other factors, such as firms' capital investment responses to changes in labour supply and the possibility of increased immigration, might be expected to mitigate this effect. Furthermore, exogenous processes such as technological advances and longer-term changes to the relative market power of capital and labour also may interfere with this relationship (Cahuc et al., 2014).

To turn to the specific implementation details, the number of jobs of the economy is held to be proportional to the population, with a weighting towards those of working age in line with the consumption patterns identified by Lee et al. (2011). An agent's capability to produce in a given job is defined by a fixed productivity function, which varies according to the amount of experience an agent has in the labour market, the agent's predetermined 'skill' level, and the 'difficulty' of the job. Experience (defined as the number of past years of employment) is assumed to affect productivity in line with the Mincer model of lifetime earnings (Mincer, 1974; Cahuc et al., 2014). More specifically, productivity increases with experience at a decreasing rate, before declining for higher values, reflecting the acquisition of human capital throughout an individual's life while allowing depreciation of same as retirement approaches (ibid).

Skill and difficulty are related to productivity in such a way that higher skilled agents have a productivity advantage when undertaking more difficult tasks (Cahuc et al., 2014, section 10.2.2). The specific functional form of productivity is given in Equation 6.4, with s, d and e representing skill difficulty and experience respectively. The final choice of function is somewhat arbitrary, but has a basis in economic literature and results in a realistic log-normal distribution of wages for many populations and parameter choices. Including graduation in jobs and earnings within the labour market allows the potential to examine how distributional aspects might affect the workings of the mechanism identified by Easterlin.

$$p(e, s, d) = \beta d \exp(\alpha(s - d) + s + \gamma e - \delta e^2) \tag{6.4}$$

Realised wages in the simulation are deterministically related to the size of a particular cohort by means of an adjustment to this productivity function. The relative size of each working-age birth cohort is calculated, and a Gaussian kernel centred on an agent's age is used to take a weighted average of nearby age groups to arrive at a measure of relative cohort size. The final wage (given in Equation 6.5) is then realised as the product of the exponential of this weighted cohort size f and the productivity p given by Equation 6.4. In addition the elements l and m allow for the possibility of linear and exponential growth in time t, reflecting the effect of economic growth. Thus, the effect of supply of labour on wages is treated as given and exogenous in this model.

$$w(p, f) = lt + \exp(\zeta f + mt)p \tag{6.5}$$

A number of factors are ignored in the model. Specific relationships between earnings, consumption, savings, and demand for labour are at present ignored, for the sake of focusing on the demographic rather than the economic elements of the phenomenon.

6.6.4 Employment

The employment class controls the labour market behaviour of the agents themselves. Agents begin to apply for jobs when they pass the age of 16, and retire from the market at age 65. Agents apply at random to vacant posts, and are offered jobs if they are the best candidate, assessed according to their contribution. Agents accept the job if it is the best offer they receive. The number of applications is drawn randomly each turn and depends upon an agent's employment status; those without jobs apply more widely. The mean number of applications made while employed or unemployed are tunable parameters in the model. A certain amount of 'churn' is also introduced, meaning that a small number of jobs are created and destroyed every time step, representing random exogenous shocks. At present, the simulation is run with only male breadwinners in an attempt to reflect the reality of the period, although female labour market participation is possible within the simulation, and increases commensurate with those occurring in the later twentieth century could be modelled in future work.

6.6.5 Fertility

The fertility class is central to the simulation behaviour. The modular nature of the simulation allows for various specifications to be examined. In line with Cioffi-Revilla (2010), a series of models are specified starting from the very simple. The simplest such model m^0, described in Equation 6.6, just adapts a schedule of age-specific fertility rates $\mu_f(x)$ according to a relative cohort-size measure C, in a similar way to is described in macro-models in the literature such as Wachter (1991) and Lee (1974). This base-case model ignores the employment elements of the model, and allows a check that self-generated Easterlin-like cycles are indeed possible within the confines of the model as we have described it.

$$p(b_{xt}) = \mu_f(x)\exp(\eta C) \tag{6.6}$$

The model m^1 requires definition of aspiration levels for individual agents. These are defined with references to an agent's father's earnings at some formative age in their childhood (by default, at 15). Thus, in this case, an agent 'remembers' this value (denoted y), and an age-specific fertility function is adjusted according to a ratio of this value (averaged over the male and female members of a partnership) and the breadwinner's own earnings w_i. The log of this ratio replaces C in Equation 6.6 above.

For each of these models the age-specific fertility schedule is defined by a Hadwiger function (Chandola et al., 1999), as shown in Equation 6.7. The parameters of this function are fixed at values which allow a Total Fertility Rate (TFR) of just over 2 to be maintained, in absence of relative-cohort size adjustment. This means the population size stays relatively stable, making for easier analysis.

$$\mu_f(x) = a\frac{b}{c}\left(\frac{c}{x}\right)^{\frac{3}{2}} \exp\left(-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right) \tag{6.7}$$

These models provide a simplistic representation of the relationship between relative cohort size and income. However, they rely on adjusting an already calibrated fertility schedule according to relative cohort size. A better model would attempt to build the fertility from the bottom up.

An alternative model m^2 provides a different approach. Instead of allowing transitions between parities on a random basis, according to the ratio of income w to aspiration y, the model instead allows only a deterministic relationship, and a heterogeneous distribution of crucial decision-making parameters across agents provides for different behaviour across agents. Agents are endowed with an individual value γ_i determining what fraction below their parents' past earnings y_i they deem sufficient to start a family. Thus, the indicator $I_{i,1}(t)$ describing whether an agent has their first child in year t is defined as follows:

$$\begin{aligned} I_{i,1}(t) \text{ if } &: \\ & w_{it} > (1-\gamma_i)y_i \\ \gamma \sim \text{Uniform}(0, \gamma_{max}) & \end{aligned} \tag{6.8}$$

The individual preferences relating to consumption are uniformly distributed across the population, with a minimum value at 0 and the maximum determined by a variable parameter. Transitions to higher parities follow a similar process, except that we include both an adjustment for the notional cost of additional children as a proportion of income χ, and also allow for agents to limit their fertility once they reach their desired family size $par_i^{desired}$. The distribution of these preferences for numbers of children are determined according to another parameter. Equation 6.9 describes these conditions as the product of two indicator functions - on relating to relative income, the second to parity.

$$I_{i,2+} \text{ if } : \quad I\left\{(1-\chi(par)w_i) > (1+\gamma_i)y\right\} \; * \; I(par < par_i^{desired}) \tag{6.9}$$

Once it has been determined that a couple decide to add to their family, a new agent is created. The newborn agent's characteristics are mostly initialised in their defaults states; age, experience, and (for females) parity, are for instance set to zero. Given the centrality of the concept of relative income, and the identification of one's parents as a comparator group, it is expected that social mobility and the correlations of income within families might impact on the fertility process modelled. The model therefore allows for intergenerational transmission of earning potential (for instance, through investment in education and parental attention) by allowing an agent's 'skill' value to be correlated with the average of it's parents $\bar{s}$. The degree of correlation is a parameter k in the model, so that:

$$\begin{aligned} \eta &\sim N(k * \phi(\bar{s}), 1-k^2) \\ s_{child} &= \Phi(\eta) \end{aligned} \tag{6.10}$$

where ϕ and Φ represent the standard normal density and distribution function respectively.

6.6.6 Initialisation

The model initialisation is a complex process. The starting population of 5000 agents receive random ages drawn according to a specified distribution, and random experiences commensurate with their age. The age distribution is chosen so that there are fewer agents at older ages, as might be expected because of mortality, and furthermore, the cohort aged between 40-60 at the start of the population is out-sized relative to those younger than it. Thus, those entering childbearing at the start of the simulation should belong to a small cohort, providing some impetus to the Easterlin-like effect that is the target for the simulation. Partnership and employment status setup amongst the initial population involves randomly assigning individuals into states for each age group (increasing with age), with probabilities designed to result in realistic population proportions. Those assigned to the 'married' and 'employed' states are matched with appropriate partners and jobs respectively. Similarly, children in the starting population are randomly assigned parents who have met their aspirations, which are again drawn at random. Some artefacts are evident in this setup, due to the difficulty of matching starting aspirations to as-yet unrealised partnerships and earning potentials, so a burn-in phase of 50 time-steps is allowed before simulation results are collected.

6.6.7 Model execution

The simulation class controls and coordinates the simulation. Time-step lengths are configurable, but one year steps are used in the result presented. Each time-step, agents are 'aged-on', and mortality, partnership search, job applications and fertility behaviour all take place, in that order. Following this, the matching in the marriage market is resolved, and similarly, job applications are assessed. The simulation code is designed so that the parameters for individual runs in an experiment can be distributed to independent instances of the program. This allows us to take advantage of the 'embarrassingly parallel' nature of the task of running multiple simulations (in that the results of each simulation are independent of each other), and utilise supercomputer resources such as the University of Southampton's `iridis 4` to minimise runtime. The simulation is run with a starting population of 5000 agents, but because some of the runs exhibit high levels of population growth and all run over 400 simulation-years, the final population size can be much higher.

6.7 Simulation Results

6.7.1 Simple probabilistic models

Starting with the simplest model m^0, simulation runs at suitably high values of the feedback parameter allow cycles in the times series of births to arise, as can be clearly seen in Figure 6.5. This is not surprising given the deterministic relationship between cohort-size and fertility in the model, and the work of Lee (1974) and Wachter (1991), who show that such birth cycles are possible under at least some formulations of the theory.

By examining some other quantities it is also possible to see that although wages and unemployment are unrelated to the fertility process for this specification, they are clearly effected by changes in cohort size, which can be seen in Figure 6.6. This lays the groundings for later models, where fertility is affected by cohort size only through the labour market.

Moving on to the model m^1, which involves an adjustment of an age-specific fertility schedule based on relative cohort size, we note that the simulation is again able to recreate the expected cycles (Figure 6.7), giving some credence to the idea that relative earnings can form a plausible intermediary between cohort size and fertility.

However, for both these models, problems remain. The distribution over parity is particularly problematic, as can be seen in Figure 6.8. One would expect parity 2 or 3 to account for the highest share of families, but this is not the case. The lack of

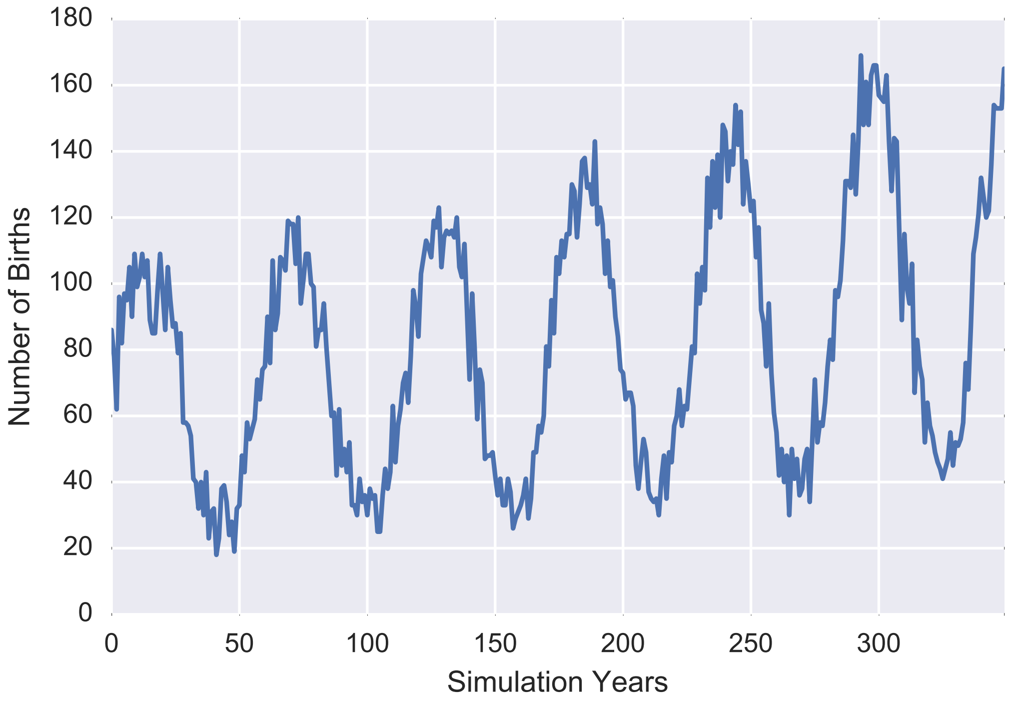

FIGURE 6.5: Plot of time-series of simulated births for model m0

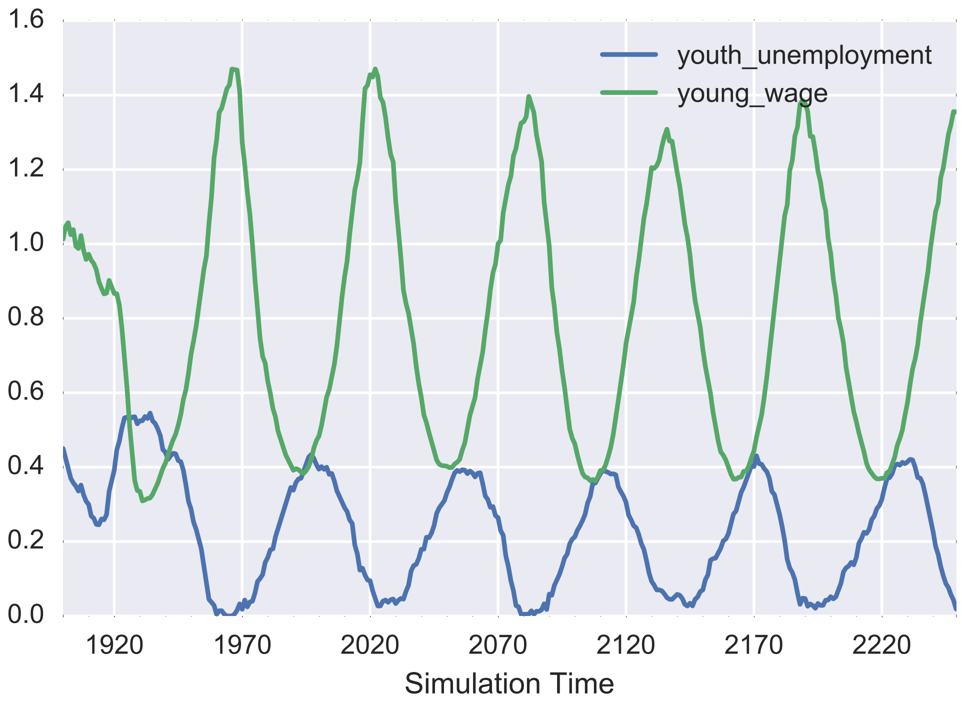

FIGURE 6.6: Plot of time-series of simulated youth unemployment and wages for model m0

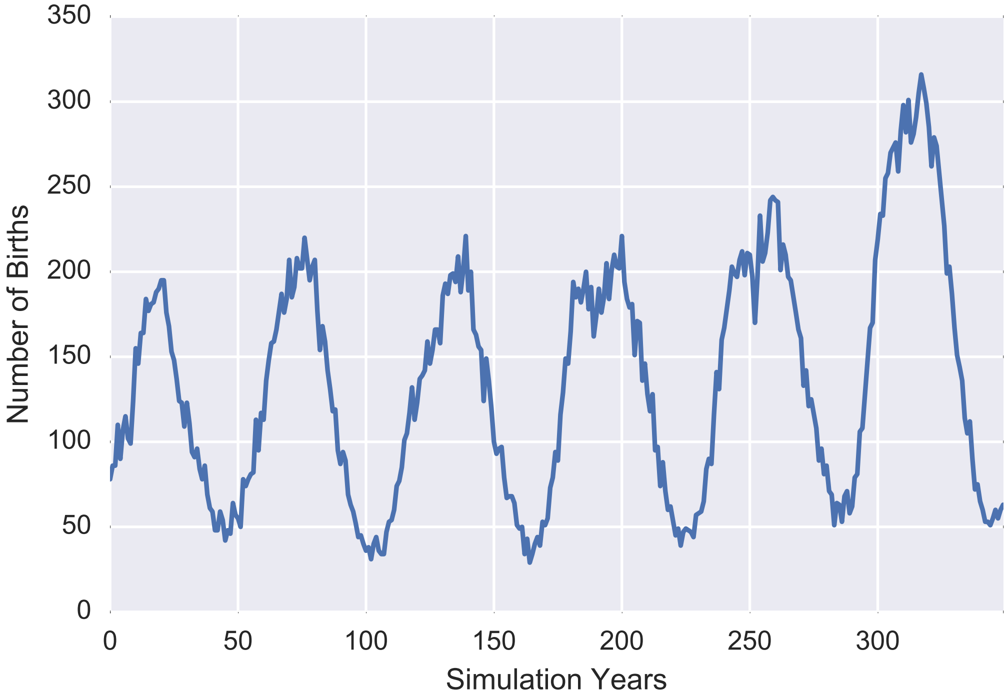

FIGURE 6.7: Plot of time-series of simulated births for model m1

distinction between births of different orders in the decision making model explains the lack of realism in the parity distribution.

Furthermore, in this model, the age-specific fertility schedule is pre-fixed, and does not emerge from individual decisions. Thus, although the variation in this schedule results from individual level comparisons between own and parental incomes, a model in which the timing of childbearing is endogenous would be preferred.

6.7.2 Heterogeneous Agents Model

Model m^2 provides a simulation that is less dependent on stochastic response and more on explicit decision making. To understand the inter-relationships of the various parameters in the model and how they impact on fertility patterns, the use of emulation techniques is advisable. Additionally, the inclusion of a reasonable number of parameters in the model, and a greater degree of uncertainty about how those parameters will impact on fertility means that we must attempt to calibrate the model to reproduce the empirical phenomena we are interested in. In order to identify plausible parameter values that reproduce empirical patterns, a large number of simulation runs were undertaken at a spread of values. More specifically, a Latin Hypercube sample of 400 points, each of which were repeated 5 times to allow an estimation of the effect of simulation stochasticity generated from the Monte-Carlo trials in the simulation. Each run involved

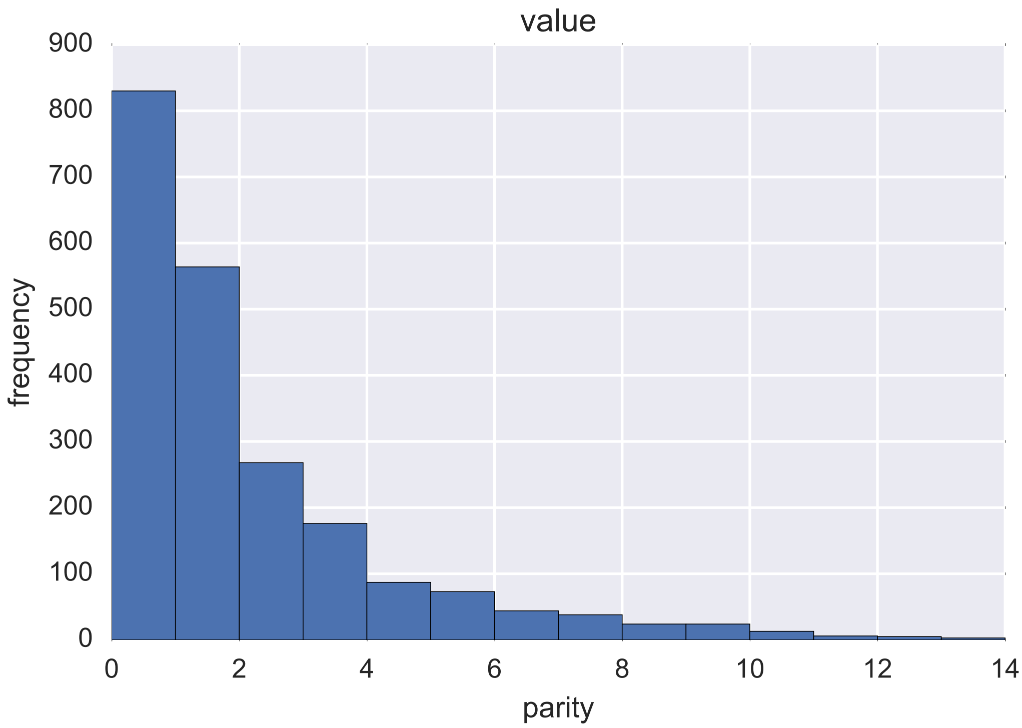

FIGURE 6.8: Plot of simulated parity distribution for a year in simulation m0

400 simulated time-steps with a starting population of 5000 agents. Seven parameters were varied within these runs, as given in the list below. More parameters could have potentially been added to this list, such as those governing the productivity function in Equation 6.4, but these were left for future investigation.

1. *Support Ratio* : Defines the relationship between weighted population size and number of jobs.
2. *Small Family Desire* : The proportion of agents who desire 2 children or fewer. The remainder of agents are split between a desire for 3 or 4 children
3. *Wage Growth Rate* : This describes the year-on-year increase in wages, excluding the increase associated with agent's increased experience. These wage increases are assumed to be driven by increases in economic production, and represented by m in Equation 6.5.
4. *Wage Elasticity* : This describes the extend to which wages respond to changes to labour supply, operationalised by the relative size of an agent's birth cohort (Given by ζ in Equation 6.5).
5. *Intergenerational Correlation* : The extent to which individuals have similar abilities to earn as their parents, assumed to be driven by investment in education and parental time investment (described as k in Equation 6.10)
6. *Child Cost* : The cost as a proportion of income of raising an additional child (parameter χ in Equation 6.9)

7. *Relative income offset* : Defines the maximum of the distribution of individual parameters describing how an agent's fertility responds to relative wage (defined in Equation 6.8 as γ_{max}).

In this initial group of runs, the majority of simulations showed population decline, due to too few agents reaching their aspired standard of living, while some others runs saw explosive growth. Thus, a heteroskedastic emulator (see Chapter 4) was fitted to the growth rates r defined by the standard growth equation $P_t = P_0e^{rt}$, and an initial screening process identified ranges of the input parameters where the population is growing. A relatively ad-hoc method was used to achieve this goal; ranges of each parameter were selected for which moderate population growth was predicted, and a new Latin-Hypercube sample was defined within these confines. Cutting the range in this manner facilitates the subsequent calibration process, in which the history matching techniques described in Chapter 5 are applied in order obtain a subset of the parameter space for which the cycles of the desired period and amplitude can be obtained, combined with a similar rate of population growth.

6.8 History Matching of Easterlin Model

6.8.1 History Matching setup

To analyse subsequent waves of simulation runs, the time series of births from each simulation run in the repeated Latin Hyper-cube designs were first de-trended by dividing by a trend estimated using LOESS smoothing, which fits a local regression function at each observation to a subset of points, with the points weighted by distance from that particular observation (Cleveland et al., 1998). In common with Wachter (1991), we therefore observe the proportional deviation in births from the trend. Next, a non-linear model is fitted to the detrended series with the following form:

$$b*_t = 1 + \alpha sin(2\pi t\phi + \psi)$$

where $b*_t$ denotes the de-trended birth sequences, ϕ and ψ describe the frequency and phase of the time series respectively, while α is related to the amplitude. To find suitable starting points for the fitting process, a discrete Fourier transform of the detrended series is taken, and the dominant frequency f (equivalently, the dominant period $p = 1/f$) extracted from the spectrum. This measure of periodicity, together with the average rate of growth r, and the amplitude of the cycles are the important results from this process. According to Wachter (1991), a period of 42 years and a growth rate of 0.008, together with an amplitude of around 20% of the overall level of births, were seen in empirical data in the United States [4]. We therefore wish to recreate these metrics in our simulated time series.

To this end, heteroskedastic emulators are fitted to three outputs; the log of the estimated period, the log of the amplitude, and the growth rate. A zero-mean discrepancy function was defined, allowing a reasonably tight distribution around the target values of the output in order that the calibration may take into account both possible differences between simulator and reality, and also uncertainty in the original estimates of the key outputs. For the period and amplitude parameters, the discrepancy distribution was specified on the log scale. Implausibility values for 10 million parameter combinations were then estimated based on predictions from these emulators. The large number of points is necessary to adequately cover the 7-dimensional parameter space.

Note that this process adds an additional layer of uncertainty to the calibration compared to other applications of the History Matching technique, in that we are applying emulators to parameter estimates resulting from a regression model, which is itself applied to simulator outputs, rather than emulating the results of the simulation directly. This reduces the dimensionality of the calibration problem, as it is then not necessary to consider the whole birth time-series as a target of the emulator. During the emulation and calibration process, the simulator can be thought of as an arbitrary set of calculations (a black box) translating inputs to outputs, which it is assumed can be represented as a Gaussian process, here the post-simulation regression estimation is considered as an additional part of these calculations.

An alternative might be to treat time as an input dimension, and include the periodic structure in the emulator itself. However, including a sinusoidal component as part of the mean function of the GP or even as part of the covariance structure (Rasmussen & Williams, 2006, p118-122) will generally involve assuming that the period of the sinusoid remains constant over the other parameters, which is not appropriate in this case.

A related problem is that it is assumed that the emulated outputs are independent; this is unlikely to be the case. This assumptions can be relaxed through the use of multivariate emulators, as discussed in previous chapters. However, this increases the complexity of the calibration operation, or otherwise requires the modeller to assume that correlations structures are the same across outputs.

Given these problems, the history matching process undertaken here must be treated with caution, as a heuristic approach to obtaining a set of input parameters compatible with target outputs. An alternative approach might be the likelihood emulation approach of Oakley & Youngman (2011), or the minimum simulated distance technique of Grazzini et al. (2013a).

[4] Wachter's metrics are used rather than attempting to re-estimate the values afresh, as Wachter had access to a longer time-series of data than currently accessible, going back to the turn of the century.

6.8.2 History Matching Results

The expected distribution of the implausibility metric is of a chi-squared distribution with degrees of freedom equal to the number of dimensions, and so we reject as implausible all values that fall beyond a cutoff point in the upper tail of this distribution (Vernon et al., 2010). Of the 10 million initial points, only around 250,000 are not rejected in this manner, suggesting that many parameter combination do not display the results desired.

Figure 6.9 is an optical depth plot showing the results of this exercise. The figure shows each bi-variate combination of parameters, splitting these into bins. Each combination is one panel in the plot, and the heat map describes for each bin, the proportion of generated points which cannot be rejected as implausible candidates for the 'true' set of parameters. Darker colours indicate that a greater proportion of points are viable in a given region.

As can be seen from the plot, higher values of the support ratio parameter are rejected altogether, as these result in too few jobs to provide the population growth targeted. Furthermore, there are some clear trade-offs between the extent to which wage responds to labour supply (Wage elasticity) and the range of the heterogeneous income offset parameter, which determines how close individuals wish to be to their parents' past income before they start a family.

However, there is still a considerable amount of the parameter space which remains non-implausible after this initial cull of implausible points, and so a second wave of history matching is required. An additional sample of 400 points is taken from the set of non-implausible points, and the simulation is run again at these points, with 7 repetitions at each point to increase the precision of the variance estimates.

The process of generating estimates and calculating implausibilities is repeated with emulators fitted on this new training sample of non-implausible points. Of the c.250,000 points that were previously viable, only around 2,700 parameter combinations remain in the set of points that may have generated the target data. Figure 6.10 displays the locations of these remaining points. Note that the colour scale indicates that for most bi-variate cells, the vast majority of points are implausible. As before, trade-offs between parameters are evident, particularly Wage Elasticity and Income Offset.

6.8.3 Results of Calibrated Simulation

From the remaining space of non-implausible parameters, 10 points are selected in order to examine the behaviour of the calibrated simulation for summary purposes. The points were chosen to maximise coverage of the remaining viable 2,700 parameter combination by generating a number of potential 10-point samples, and choosing the one which had

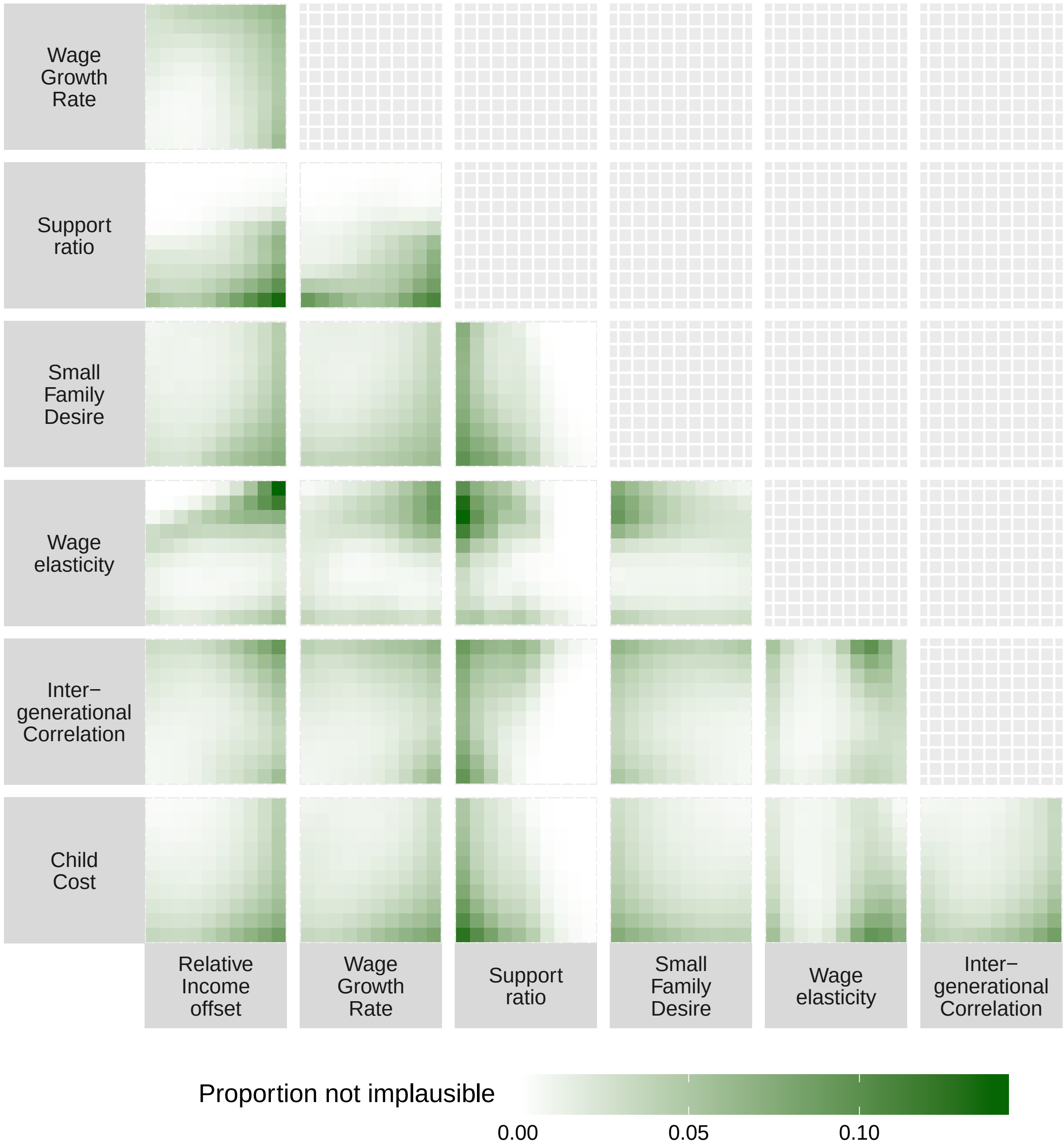

FIGURE 6.9: Wave 1 Optical Depth plot describing location of non-implausible points in parameter space

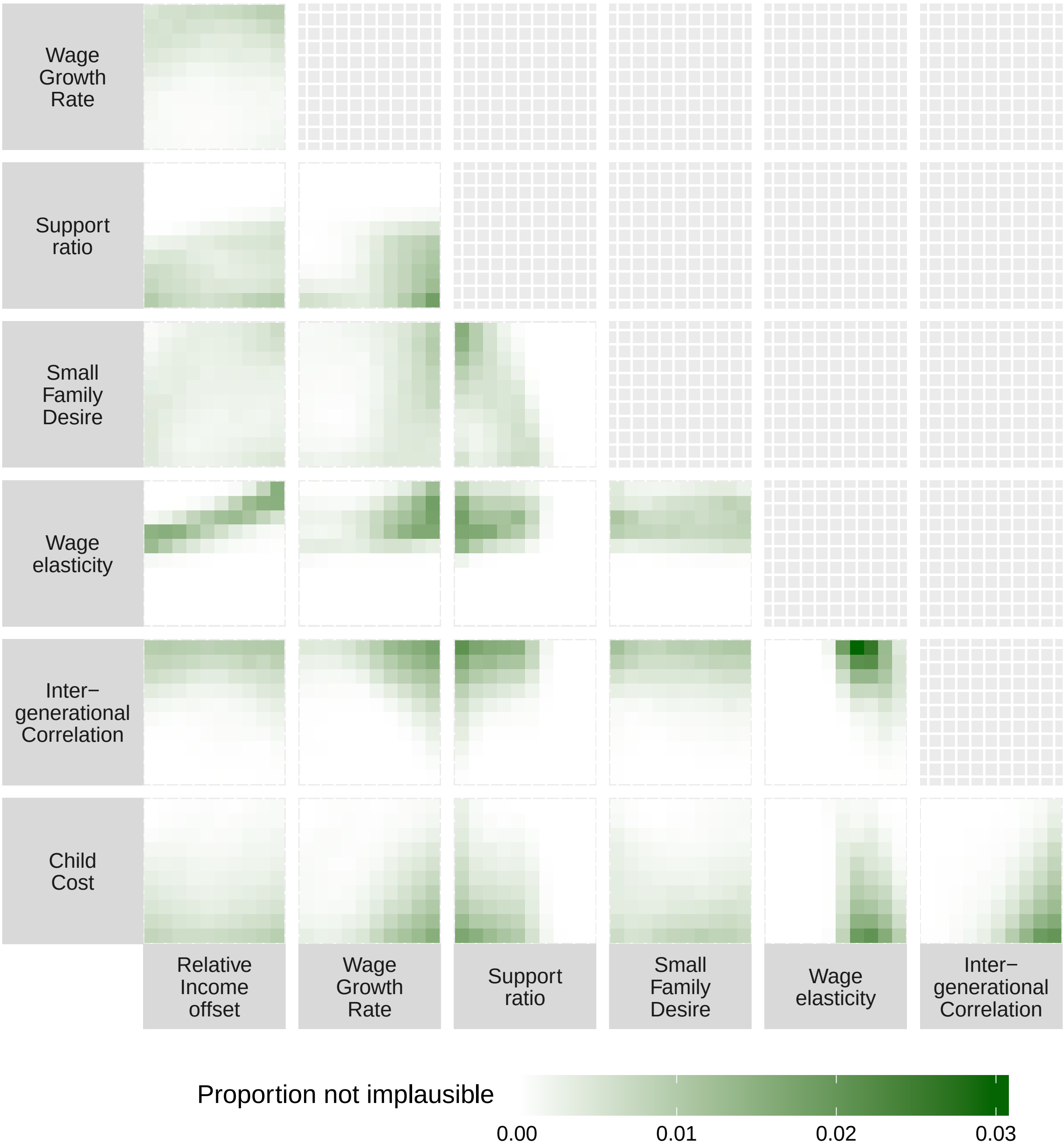

FIGURE 6.10: Wave 2 Optical Depth plot describing location of non-implausible points in parameter space

the largest minimum distance between points. The simulation was run 7 times at the selected points, and a range of statistics were collected for each run.

To briefly characterise the calibrated points, the growth rates clustered around the targeted 0.008 rate, and the detrended series mostly had the desired periodic characteristics, as is evident in Figure 6.11. Some variability between the different time series is evident; some runs appear to display very strong periodic tendencies which remain constant throughout the simulation period. Others in contrast appear to be deteriorating and may be only transitory.

Other quantities aside from those calibrated against are also of interest, in particular in allowing some degree of validation of the simulated results. Figure 6.12 displays age-specific fertility for each of the 10 calibrated points over a period of 25 simulated years. This period is chosen in an attempt to capture a peak and a trough in fertility fluctuations. Also plotted (in red) are empirical ASFRs for the US for the peak year of 1957 and the earlier low point of 1936. As can clearly be seen, the simulated fertility schedule has a peak in broadly the correct place, but the peak is more pronounced than in the empirical data, and declines too early. The shape of the fertility curve in the simulation is most likely related to the rate of growth of the productivity function with age. This helps define the rate at which individuals are able to catch-up with their parents earnings, and thus meet their aspirations. The parameters relating to this function were not included in the calibration exercise, and so future work might focus on calibrating against the rates explicitly and add the relevant parameters to the set to be varied.

6.9 Discussion

This chapter identifies some conditions under which Easterlin-like cycles in fertility can be generated, under a number of assumptions needed to abstract reality into the simulation described above. It thus contributes an existence or plausibility proof, examining how the mechanisms described by Easterlin could indeed have led to the observed cycles. Simulation provides a useful tool with which to investigate this question as it allows us to replicate the process under discussion. In a sense, it provides a more detailed formalisation of the Easterlin Hypothesis than existing mathematical treatments by Lee and Wachter are able to provide, because specification of the 'control' process (whereby fertility is restricted amongst larger birth cohorts) is at the micro-level, rather than assumed and specified in terms of elasticities on fertility rates as a function of past births. From a methodological point of view, the value of using statistical emulators to examine and calibrate simulation models is underlined by the use of Gaussian process emulators to identify areas of the parameter space which recreate the two-generation cycles described

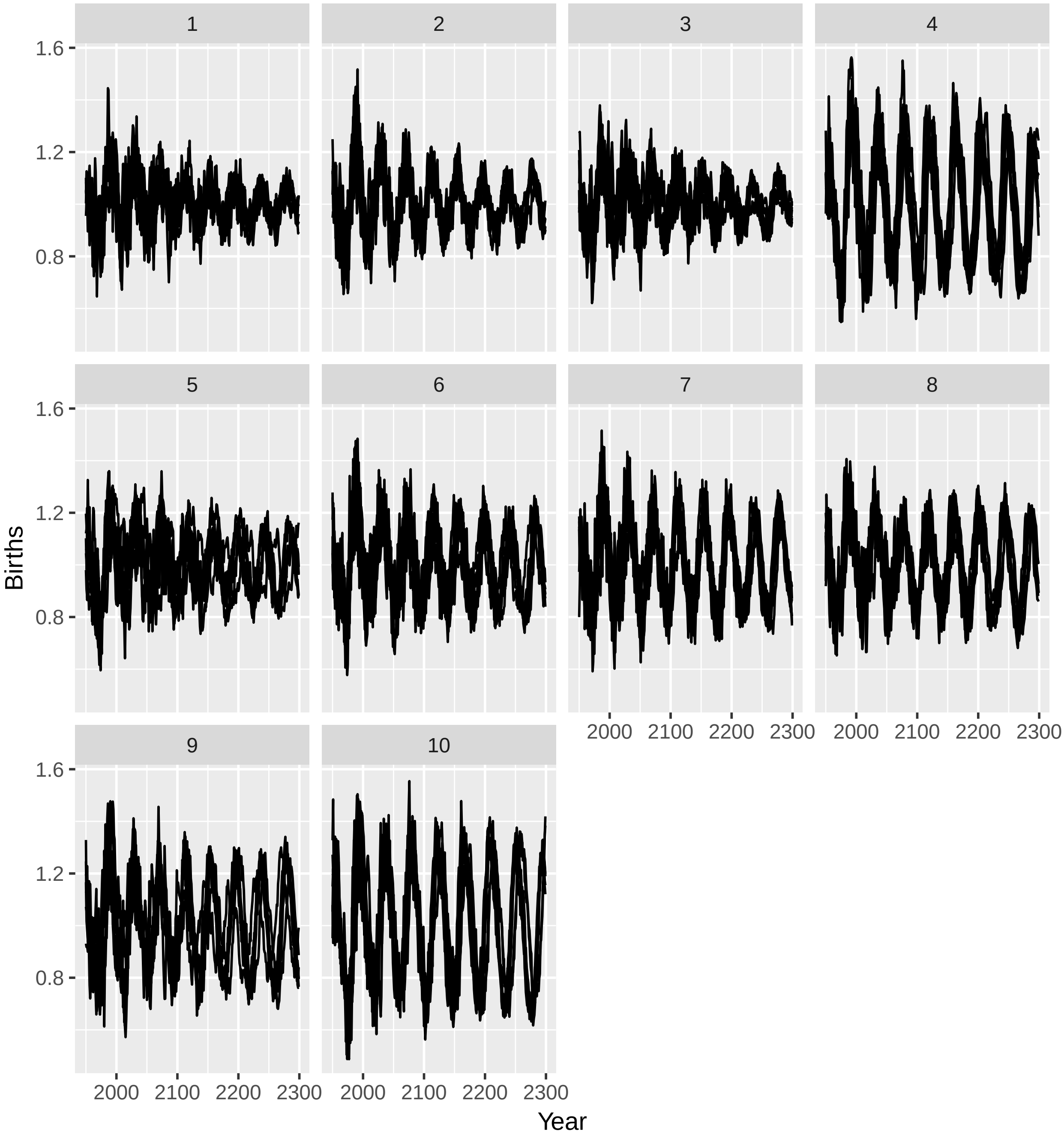

FIGURE 6.11: Simulated time-series of detrended births for a selection of non-implausible points

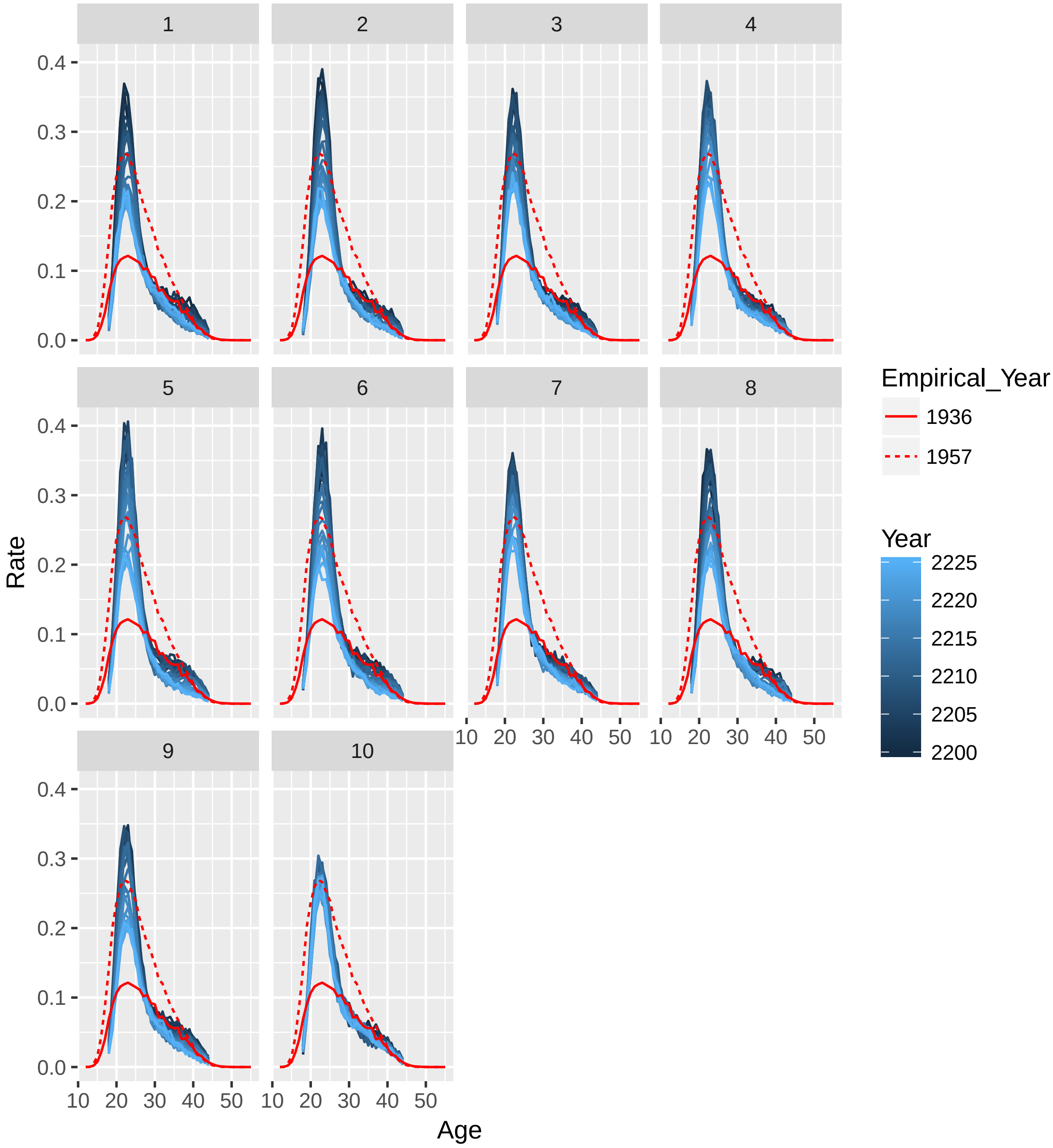

FIGURE 6.12: Comparison of Age-Specific Fertility Rates versus empirical US equivalents at peak and nadir of cycle. US Data from Human Fertility Database

by Easterlin. One benefit of the approach utilised here relative to optimisation-type calibration schemes is that a range of parameter combinations are identified, rather than a single 'best fit' point that ignores the fact that other parameter combinations may provide results well within a reasonable error, but might better reflect reality on other dimensions that are not observed or calibrated on (Oakley & Youngman, 2011).

Further work could take a number of interesting directions, particularly given the extensibility of the simulation framework. Firstly, various scenarios that have been posited to explain the demise of cyclic patterns in fertility from the mid-80s onwards may be tested within the simulated environment. For instance, female labour force participation maybe gradually increased, and likewise immigration may be introduced. It might be expected that this would result in a weakening in the relationship between birth cohort size and labour market success.

An interesting consideration is the relationship between population heterogeneity and randomness in simulations such as these. Part of the difference between models m^1 and m^2 is the reliance of the former on random trials, while the latter uses deterministic decision processes. The use of probability in model m^1 is here 'standing in' for heterogeneity in how much importance individuals place on their income relative to their wage. The model m^2 'explains' some of this randomness with a deterministic relationship and population heterogeneity, and is preferable in that it also takes into account the effect of dynamic sorting in a way that a simple probabilistic relationship would not - in a similar manner to the way that frailty models are better able to estimate hazards by accounting for differential population compositions over time.

Another direction might be to try and extend the simplistic model of decision making provided in the model to include a greater degree of time-awareness, more explicit calculation on the part of the agents as to the desirability of labour and family changes, and some explicit representation of the limits to the information available to agents (c.f. Gray et al.). More sophisticated treatment of economic factors may also be possible. In particular, the introduction of savings at the individual level could open up interesting additional areas of research, as could the introduction of firms as explicit agents in the labour market.

At present, the simulation does not display complex emergent behaviour, which is not necessary to capture the nature of the target system. However, social interaction effects may also be included in future work. At present agents interact only through the labour and marriage markets. However, moving from a measure of relative income based on parents to one using social network peers as a reference group may be one way to examine this effect. In this framework, individuals are discouraged from starting a family if they feel the earn less than their friends. Alternatively, information about the benefits and costs of family formation may be passed from early adopters to peers, so that knowing parents may in fact make you less likely to want to become one yourself.

Effects of this nature also allow the currently exogenous family size preferences to become endogenous norms, as in Aparicio Diaz et al. (2011), mutated by knowledge of peer's experiences.

However, we must beware of the 'kitchen sink' approach to simulation modelling, whereby the modeller attempts to include every possible facet and nuance of the target system (Axelrod, 2003). Adding more and more detail will generally cloud our understanding of the simulation, as well as adding more and more potential sources of error. Keeping the simulation modular in design allows us to consider these extensions in isolation, allowing us to test the implications of each extension individually (Gray et al.).

7

Summary and Conclusions

7.1 Summary

7.1.1 Statistical anaylsis of demographic simulations.

This thesis has demonstrated how Gaussian process emulators can be used to analyse and calibrate simulations of demographic processes. Simulation methodology and agent-based models in particular are valuable tools for investigating demographic behaviour for a number of reasons. Firstly, they allow demographic processes to be analysed from 'the bottom up', by specifying rules of behaviour at the micro-level, and examining the population-level consequences of such behaviour (Epstein & Axtell, 1996). Thus, observed demographic patterns can be explained in terms of micro-level behaviours and interactions (Billari, 2015). As a corollary of this multi-level structure, downward effects of the population level on individual behaviour is also observable within a simulation (Gilbert & Troitzsch, 2005). Secondly, the ability to encode interactive behaviour into such simulations opens up whole fields for study that are difficult to capture with other methodologies, including information spread, norm development and social learning (Epstein & Axtell, 1996; Squazzoni, 2012). Thirdly, psychologically plausible models of decision making can be examined and tested for plausibility by allowing them to govern agents' actions within a simulation (Epstein, 2013). Fourthly, the inherently dynamic nature of simulation means that the interaction of all these features can be examined over time, rather than in snapshot. Finally, treating simulations as pseudo experiments allows the modeller to conduct 'what-if' exercises and scenario analysis (Epstein, 2008).

However, despite these advantages, developing agent-based models and producing new demographic knowledge with them remains a difficult task. This is partly because of the difficulties in analysing results from such models. The flexibility of simulation modelling, where the limitation on what is possible becomes computational rather than analytical tractability (Leombruni & Richiardi, 2005), means that often rather complex models result, and even relatively simple models require many parameters to represent quantities we cannot observe empirically, particularly those relating to decision processes.

This flexibility also means there are many sources of uncertainty entering into the modelling process. These include uncertainty over the 'correct' structure of the model itself, pseudo-randomness in the simulation introduced by random number generation and uncertainty over the true values of inputs, as well as a lack of knowledge about points at which the simulation has yet to be run (Kennedy & O'Hagan, 2001a). These uncertainties multiply when one attempts to relate a simulation model to reality: one is then faced additionally with measurement error and uncertainty about the discrepancy between model and reality (ibid).

The concern of agent-based models with representing multiple individuals who interact repeatedly with one another, affecting each other's actions, means that ABMs are expected to exhibit *complex* behaviour, meaning that population-level patterns *emerge* from the micro-level, and are not trivially predictable from knowledge of the lower-level conditions (Epstein & Axtell, 1996). This results in non-linearities in the relationships between inputs and outputs, and therefore it is expected that this makes the analysis of such models difficult.

The combination of the above features of agent-based simulation means that statistical analysis is needed in order to ease the process of analysis and calibration. This work suggests that Gaussian process emulators provide a natural and convenient way to approach this problem.

7.1.2 Summary and Contribution of this thesis

Together, the challenges discussed in the previous section form the motivation for the work discussed in this thesis. Taken together, the work in this thesis provides a methodological template for the design and analysis of demographic agent-based simulations, providing strategies based on work from the literature on computer simulation (e.g. Santner et al., 2003; Kennedy & O'Hagan, 2001a; Boukouvalas, 2010). Furthermore, a case study develops a novel agent-based model of fertility change, and uses the techniques introduced in the rest of the text to calibrate the model, providing micro-level rules under which the relative cohort-size mechanism of Easterlin (1987) succeeds in generating cycles in births similar to those observed by Wachter (1991) in the time series of US births.

Chapter 2 set the groundwork for the later chapters by identifying some valid scientific objectives for the process of agent-based modelling, and noting that ABMs serve a variety of purposes, from simple explanatory and theoretical models, to detailed 'social simulations' with strong empirical links (Silverman, 2007). Consequently, statistical methods for the analysis of agent-based models must both enable the analysis of highly uncertain and non-linear models, and have the ability to calibrate against empirical data. The chapter also reviews current methods of analysis of agent based models, finding that many studies proceed with analysis in a relatively ad-hoc manner. The use of more sophisticated and structured approach to analysis of simulations, such as experimental design and response surface methodology, are evident amongst some Agent-Based simulation communities, particular in operational research (Kleijnen, 2008) and to some extent in ecology (Grimm et al., 2005). However, where statistical methods are used for this purpose, they often do not account for all sources of uncertainty, and typically rely on approximations of simulation behaviour using lower-order polynomials. This motivates the advocacy of more principled techniques in the rest of the thesis. The chapter also reviews existing work on demographic agent-based models, which is a small but growing body of work.

Chapter 3 focuses on setting out the mathematical details of the emulator techniques used later in the thesis. The Chapter begins with discussion of problems of experimental design, following Santner et al. (2003) and others, and in particular the advantages of using a Latin Hypercube Sample over a more intuitive grid sampling, as the former technique allows more information to gained about the process under study from fewer runs. An introduction to the Gaussian Process Emulators is then offered following the exposition by Kennedy & O'Hagan (2001a), O'Hagan (2006), and others, giving details of the processes of estimation and prediction using GPs, as well as the techniques of sensitivity analysis and probabilistic calibration. It is concluded that these techniques help with both establishing the *analytical adequacy* of a simulation model, and with attempting to compare the model against the empirical world, by allowing calibration of the model against observed data.

The next Chapter (4), gives an example of the use of emulation techniques for the first of the problems identified in the previous Chapters, that of analysing a simulation in order to better understand the processes underlying it. A reimplementation of the simulation constructed by Billari et al. (2007), an existing application of agent-based modelling to nuptiality processes, is used for this purpose. An emulator is fitted to the simulation, its validity is tested, and the sensitivity of the important parameters in the model are tested. The emulator identifies the relative lack of importance of one parameter in the model, although it is noted that the existence of a 'phase transition' in the model means that the emulator has some difficulties in fitting the underlying simulation in the area of this discontinuity.

Having discussed the analysis of demographic simulations, Chapter 5 turns to the problem of calibration. Two different methods of calibration are examined; the first is the full model-based Bayesian calibration of Kennedy & O'Hagan (2001a), while the second is the history matching technique advocated by Vernon et al. (2010) and others. The demographic microsimulation of Zinn & Himmelspach (2009), fortified with data from the UK, is used as a test case for these calibration techniques. The heteroskedastic emulators expounded by Boukouvalas (2010) are used to model the output - a technique that is likely to be important for agent based models, which are unlikely to display constant variance across the parameter space. The chapter concludes that the greater flexibility of the history matching technique means that it is better suited to the problem of attempting to calibrate demographic models where outputs over age are often the subject of interest.

The final substantive chapter (Chapter 6) discusses the development and analysis of a demographic simulation studying the mechanism identified by Easterlin (1987) as potentially responsible for cycles in fertility in the twentieth century United States. The simulation is constructed with the aim of determining what conditions can generate such patterns. Several sequential models are specified, and the effect of population heterogeneity with respect to income and the intergenerational transmission of both aspirations and earning potential are examined. Gaussian Process emulators are used to calibrate the simulation so that the growth rates and periodicity of cycle in fertility match those observed in the empirical data, providing some micro-level conditions sufficient to generate phenomena observed by Easterlin (1987).

7.2 Limitations and alternative approaches.

While it is believed that the approaches detailed in the preceding chapters have much utility in the analysis of demographic simulations, there remain a number of limitations to the application of Gaussian process emulators to the analysis of agent-based models. The most obvious is the possibility that the underlying assumptions of smoothness across proximate points in the parameter space may be violated for some models, especially in the case where discontinuous phase transitions exist. Despite this, it is argued that GP emulators will likely fit well enough to provide the modeller with an idea of the how the simulator responds to outputs, and will certainly identify the location of the phase transition, allowing scope for further investigation of the cause of the discontinuity, and its relevance to the research question.

Another potential stumbling block is the relative complexity of the technique, relative to, say, response surface methodology. As Grow (2016) points out, the latter relies mainly on mainstream regression techniques that are well understood by demographers and other social scientists who may be involved in simulation modelling. This has advantages in both executing and communicating the analysis and calibration of models. However, the existence of R packages such as DICE (Roustant et al., 2012) takes some of the burden from the modeller. The lack of assumptions about the parametric form of the response of inputs to outputs in the Gaussian Process approach, together with the ability to handle different sources of uncertainty coherently do provide reasonable justifications for choosing the more complicated method, however, particularly if the response of inputs to outputs is likely to be complex (in the technical sense), and if subjective uncertainty to relating to model discrepancy, measurement and input parameters also need to be taken into consideration.

One solution to the problem of discontinuous response surfaces, or to cases where the assumption of second-order stationarity seem to be violated, is to further refine the emulator used to account for this. The approaches of Gramacy (2007) and Rasmussen & Ghahramani (2002) bear consideration here. Gramacy (2007) investigate Treed Gaussian Processes, whereby different models of various complexities are fitted to different areas of the parameter space, with part of the estimation process being the delimiting of the domains of the separate models. The `R` package `tgp` provides methods to undertake this process efficiently using MCMC, but one problem with the approach as implemented is that the divides between models must be made parallel to the parameter axes, which is unlikely to be capture the location of phase transitions in practice. The approach of Rasmussen & Ghahramani (2002) defines an infinite Dirichelet mixture of Gaussian Processes which allow for different models ('experts') to be hold sway in different areas of the space. The latter approach would seem to have some promise for the task in hand, but the increased computational complexity, and the loss of tractability of the resulting predictive posterior may mean that meta-models based on infinite mixtures becomes too unwieldy for practical use for the analysis of computational experiments.

7.3 Challenges and Directions for Future Work

The emulator techniques described in this thesis describe only a small corner of the process of using agent-based models to contribute to the production of demographic knowledge. Future investigations into this area might focus on dealing with discontinuities and lack of stationarity, and identifying automatically the locations of phase transitions. Other alternative approaches to analysis and calibration of computational experiments not explored in this thesis include the Bayesian Melding techniques of Poole & Raftery (2000) and polynomial chaos expansion (c.f. O'Hagan, 2013). The former has been used to calibrate a transport simulation (Ševčíková et al., 2007b) as well as a simulation of whale population growth, and would thus seem to be well suited for Agent-Based demographic purposes. Future work might revolve around a comparison of various approaches on a variety of difference simulations, with the aim of providing guidance to end-users about the most suitable tool for their goal.

An important point that must be reiterated is that calibration does not obviate the need for validation. Although a calibrated model may reproduce features of empirical observations that does not mean that the mechanism described by the simulation is the one that operates in reality, because there may be a great many ways of generating such patterns (Oreskes et al., 1994). This underlines the importance of grounding the construction of Agent-Based models with strong theoretical bases, and/or with good empirical backgrounds from both qualitative and quantitative sources (Courgeau et al., 2016).

More broadly, the practice of Agent-Based Modelling faces a number of challenges. As has been indicated, Agent-Based Models are often rather theoretical, aimed at examining the consequences or plausibility of one particular assumption or theory, or, alternatively, are aimed at providing detailed and faithful representation of some real-world system of interest with the aim of predicting its (possible) future behaviour. Both avenues provide problems. Staying purely within the theoretical realm risks talking only about worlds of our own devising, and not about the world as it is, and this makes it difficult to draw conclusions or validate models. Fine-grained 'social simulations' (Silverman, 2007) require extensive resources to operate, and it is easy to fall prey to the temptation to explicitly represent more and more elements of the interlinked layers of social reality, either through data or through explicit behaviour models, until the potential sources of error overwhelm the predictive potential of the model. Finding a middle ground that remains simple enough to be interpretable and maintainable, while complex enough to be relevant, is a difficult balance.

Bibliography

Abel, G. J., Bijak, J., Forster, J. J., Raymer, J., Smith, P. W., & Wong, J. S. (2013). Integrating uncertainty in time series population forecasts: An illustration using a simple projection model. *Demographic Research*, *29*(December), 1187–1226.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Andrianakis, I., & Challenor, P. G. (2012). The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics & Data Analysis*, *56*(12), 4215–4228.

Andrianakis, I., Vernon, I. R., McCreesh, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M., & White, R. G. (2015). Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. *PLoS Computational Biology*, *11*(1), e1003968.

Andrianakis, Y., & Challenor, P. G. (2011). Parameter Estimation for Gaussian Process Emulators. Tech. rep., Managing Uncertainty in Complex Models.
URL `http://www.mucm.ac.uk/Pages/Downloads/TechnicalReports/`

Aparicio Diaz, B., Fent, T., Prskawetz, A., & Bernardi, L. (2011). Transition to parenthood: the role of social interaction and endogenous networks. *Demography*, *48*(2), 559–79.

Axelrod, R. (2003). Advancing the art of simulation in the social sciences. *Journal of the Japanese and International Economies*, *12*(3), 16–22.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., & Parker, M. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences of the United States of America*, *99 Suppl 3*(1), 7275–9.

Bastos, L. S., & O'Hagan, A. (2009). Diagnostics for Gaussian Process Emulators. *Technometrics*, *51*(4), 425–438.

Bernardi, L. (2003). Channels of social influence on reproduction.

Bijak, J. (2011). *Forecasting International Migration in Europe; A Bayesian View*. Dordrecht: Springer.

Bijak, J., Hilton, J., Silverman, E., & Cao, V. D. (2013). Reforging the Wedding Ring: Exploring a Semi-Artificial Model of Population for the United Kingdom with Gaussian Process Emulators. *Demographic Research*, *29*(27), 729–766.

Billari, F., Fent, T., Prskawetz, A., & Aparicio Diaz, B. (2007). The "Wedding-Ring": An Agent-Based Marriage Model Based on Social Interactions. *Demographic Research*, *17*, 59–82.

Billari, F., Fent, T., Prskawetz, A., & Scheffran, J. (2006). Agent Based Computational Modelling: An Introduction. In F. Billari, T. Fent, A. Prskawetz, & J. Scheffran (Eds.) *Agent Based Computational Modelling*, chap. 1, (pp. 1–14). New York: Physica-Verlag.

Billari, F., Ongaro, F., & Prskawetz, A. (2003). Introduction: Agent-Based Computation Demography. In F. Billari, & A. Prskawetz (Eds.) *Agent Based Computational Demography*, chap. 1. Heidelberg: Physica-Verlag.

Billari, F. C. (2015). Integrating macro- and micro-level approaches in the explanation of population change. *Population Studies*, *69 Suppl 1*(April), S11–20.

Boero, R., & Squazzoni, F. (2005). Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *Journal of Artificial Societies and Social Simulation*, *8*(4), 6.

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. *PNAS*, *99*(suppl. 3), 7280–7287.

Booth, H. (2006). Demographic Forecasting: 1980 to 2005 in review. *Interational Journal of Forecasting*, *22*, 547–581.

Boukouvalas, A. (2010). *Emulation of random output simulators*. Phd thesis, Aston University.
URL http://eprints.aston.ac.uk/15776/

Boukouvalas, A., Sykes, P., Cornford, D., & Maruri-Aguilar, H. (2014). Bayesian precalibration of a large stochastic microsimulation model. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), 1337–1347.

Brenner, T., & Werker, C. (2007). A Taxonomy of Inference in Simulation Models. *Computational Economics*, *30*(3), 227–244.

Brooks, S., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Brown, D., Page, S., & Riolo, R. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographic Information Science*, *19*(2), 153–174.

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, *30*(Special issue: Bayesian methods in inverse problems), 1–32.

Buchanan, M. (2009). Meltdown modelling. *Nature*, *460*(August), 680–682.

Cahuc, P., Carcillo, S., & Zylberberg, A. (2014). *Labor Economics*. Cambridge, MA: MIT Press, 2 ed.

Carley, K. (2002). Simulating society: The tension between transparency and veridicality. *Proceedings of Agents*, (October), 1–16.

Casini, L., Illari, P., Russo, F., & Williamson, J. (2010). Recursive Bayesian Nets for Explanation, Prediction and Control in Cancer Science. In A. Fred, J. Felipe, & H. Gamboa (Eds.) *Bioinformatics*. Valencia.

Chalabi, Z. (2014). Quantifying Uncertainty in Agent Based Models. In *Uncertainty in Complex Models*. Sheffield.

Chandola, T., Coleman, D. a., & Hiorns, R. W. (1999). Recent European fertility patterns: fitting curves to 'distorted' distributions. *Population studies*, *53*(3), 317–329.

Chwif, L., Barretto, M. R. P., & Paul, R. J. (2000). On Simulation Model Complexity. *Proceedings of the 2000 Winter Simulation Conference*, (1989), 449–445.

Cioffi-Revilla, C. (2010). A methodology for complex social simulations. *Journal of Artificial Societies and Social Simulation*, *13*(1), 7.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1998). Local regression models. In J. M. Chambers, & T. Hastie (Eds.) *Statistical models in S*, chap. 8. Wadsworth and Brooks/Cole.

Coale, A. (1973). The demographic transition reconsidered. In *Proceedings of the International Population Conference*, (pp. 53–72). Liege: Ordina Editions.

Courgeau, D. (2012). *Probability and Social Science: Methodological Relationships between the Two Approaches*. Dordrecht: Springer.

Courgeau, D., Bijak, J., Franck, R., & Silverman, E. (2016). Model-Based Demography: Towards a Research Agenda. In A. Grow, & J. Van Bavel (Eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chap. 2, (pp. 29–52). Basel: Springer.

Cumming, J., Gosling, J. P., Oakey, J., & Seheult, A. (2012). Multilevel modelling of an oil reservoir simulator. In *Uncertainty in Complex Models*. Sheffield. URL `http://mucm.ac.uk/UCM2012/Forms/Downloads/Gosling.pdf`

Di Paolo, A. E., Noble, J., & Bullock, S. (2000). Simulation Models as Opaque Thought Experiments. In *Seventh International Conference on Artificial Life*, (pp. 497–506). Cambridge MA: MIT Press.

Diggle, P. J., Tawn, J. a., & Moyeed, R. a. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(3), 299–350.

Easterlin, R. A. (1962). The American Baby Boom in Historical Perspective. *NBER Occasional Paper*, (79).

Easterlin, R. A. (1966). On the relation of economic factors to recent and projected fertility changes. *Demography*, *3*(1), 131–153.

Easterlin, R. A. (1975). An economic framework for fertility analysis. *Studies in family planning*, *6*(3), 54–63.

Easterlin, R. A. (1978). What will 1984 be like? Socioeconomic implications of recent twists in age structure. *Demography*, *15*(4), 397–432.

Easterlin, R. A. (1987). *Birth and Fortune: The Impact of Numbers on Personal Welfare*. Chicago: University of Chicago Press, 2nd ed.

Easterlin, R. A. (2004). *The Reluctant Economist*. Cambridge: Cambridge University Press.

Edmonds, B., & Moss, S. (2004). From KISS to KIDS an anti-simplistic ' modelling approach. In P. Davidsson et al. (Ed.) *Multi-Agent Based Simulation: Springer Lecture Notes in Artificial Intelligence*, (pp. 130–144). Springer.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social*, *11*(4).

Epstein, J. M. (2013). *Agent_Zero: Towards Neurocognitive Foundations for Generative Social Science*. Princeton: Princeton Univeristy Press.

Epstein, J. M., & Axtell, R. (1996). *Growing Artifical Societies; Social Science from the Bottom Up*. Washington DC; Cambridge MA: Brookings Institution Press; MIT Press.

Fagiolo, G., Dosi, G., & Gabriele, R. (2004). in an Evolutionary Model of Labor Market and Output Dynamics. *7*(2), 157–186.

Fent, T., Aparicio Diaz, B., & Prskawetz, A. (2013). Family policies in the context of low fertility and social structure. *Demographic Research*, *29*(November), 963–998.

Forrester, A., Sobester, A., & Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. Chichester: Wiley.

Franck, R. (2002). General Introduction. In *The Explanatory Power of Models: Bridging the Gap Between Empirical and Theoretical Research in the Social Science*, chap. 1, (pp. 1–8). Boston: Kluwer Academic Publishers.

Frauenthal, J., & Swick, K. (1983). Limit cycle oscillations of the human population. *Demography*, *20*(3), 285–298.

Galán, J., & Izquierdo, L. (2009). Errors and artefacts in agent-based modelling. *Journal of Artificial Societies and Social Simulation*, *12*(1).

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design Patterns Elements of Reusable Object-Oriented Software*. Reading, Mass.: Addison-Wesley.

Geard, N., & McCaw, J. (2013). Synthetic Population Dynamics: A Model of Household Demography. *Journal of Artificial Societies and Social Simulation*, *16*(1-8).

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. Abingdon: CRC Press, third ed.

Gilbert, N. (2008). *Agent-Based Models*. London: Sage Publications.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Maidenhead: Open University Press, 2nd ed.

Gramacy, R. (2007). tgp: an R package for Bayesian nonstationary, semiparametric nonlinear regression and design by treed Gaussian process models. *Journal of Statistical Software*, *19*(9).

Gramacy, R., & Lee, H. (2012). Cases for the nugget in modeling computer experiments. *Statistics and Computing*, *22*(3), 713–722.

Gramacy, R. B. (2005). *Bayesian Treed Gaussian Process Models*. Phd, University of California, Santa Cruz.

Gray, J., Hilton, J., & Bijak, J. (????). Choosing the Choice: Refections on Modelling Decisions and Behaviour in Demographic Agent-Based Models.

Grazzini, J. (2012). Analysis of the emergent properties: stationarity and ergodicity. *Journal of Artificial Societies and Social Simulation*, *15*(2), 7.

Grazzini, J., Richiardi, M., & Sella, L. (2013a). Indirect estimation of agent-based models. An application to a simple diffusion model. *Complexity Economics*, *2*(1).

Grazzini, J., & Richiardi, M. G. (2013). Consistent Estimation of Agent-Based Models by Simulated Minimum Distance. Tech. Rep. 130, Laboratorioa Riccardo Revelli.

Grazzini, J., Richiardi, M. G., & Sella, L. (2013b). Analysis of Agent-based Models.

Grimm, V., & Railsback, S. F. (2005). *Individual-based Modelling and Ecology*. Princeton, NJ: Princeton Univeristy Press.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science (New York, N.Y.)*, *310*(5750), 987–991.

Groeneveld, J., Klabunde, A., O'Brien, M. L., & Grow, A. (2016). How to Describe Agent-Based Models in Population Studies. In A. Grow, & J. Van Bavel (Eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chap. 9, (pp. 237–256). Basel: Springer.

Grow, A. (2016). Regression Meta-Models for Sensitivity Analysis in Agent-Based Computational Demography. In A. Grow, & J. Van Bavel (Eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chap. 7, (pp. 185–210). Basel: Springer.

Grow, A., & Van Bavel, J. (2014). Assortative Mating and the Reversal of Gender Inequality in Education in Europe An Agent-Based Model. (pp. 1–24).

Hammel, E., McDaniel, C., & Wachter, K. (1979). Demographic consequences of incest tabus: a microsimulation analysis. *Science*, *205*(4410), 972–977.

Hankin, R. (2005). Introducing BACCO, an R bundle for Bayesian analysis of computer code output. *Journal of Statistical Software*, *14*(16).

Haylock, R. (1997). *Bayesian inference about outputs of computationally expensive algorithms with uncertainty on the inputs*. Phd, University of Nottingham.
URL http://eprints.nottingham.ac.uk/13193/1/338522.pdf

Heard, D. (2014). *Statistical Inference Utilizing Agent Based Models*. Doctor of philosophy, Duke University.
URL https://stat.duke.edu/sites/default/files/people/theses/dph11.pdf

Hills, T., & Todd, P. (2008). Population Heterogeneity and Individual Differences in an Assortative Agent-Based Marriage and Divorce Model (MADAM) Using Search with Relaxing Expectations. *Journal of Artificial Societies and Social Simulation*, *11*(4 5).

Hilton, J., & Bijak, J. (2016). Design and Analysis of Demographic Simulations. In A. Grow, & J. Van Bavel (Eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chap. 8, (pp. 211–235). Basel: Springer.

Human Mortality Database (2011). Human Mortality Database.
URL `http://www.mortality.org/cgi-bin/hmd`

Kamiński, B. (2015). Interval metamodels for the analysis of simulation InputOutput relations. *Simulation Modelling Practice and Theory*, *54*, 86–100.

Kashyap, R., & Villavicencio, F. (2016). The Dynamics of Son Preference, Technology Diffusion, and Fertility Decline Underlying Distorted Sex Ratios at Birth: A Simulation Approach. *Demography*, *53*, 1261–1281.

Keilman, N., Pham, D. Q., & Hetland, A. (2002). Why population forecasts should be probabilistic - illustrated by the case of Norway. *Demographic Research*, *6*(15), 409–454.

Kennedy, M. (2004). Descriptions of the Gaussian process model used in GEM-SA.
URL `ctcd.group.shef.ac.uk/gem.html`

Kennedy, M., & O'Hagan, A. (2001a). Bayesian calibration of computer models. *Journal of the Royal Statistical Society*, *63*(3), 425–464.

Kennedy, M., & O'Hagan, A. (2001b). Supplementary details on Bayesian calibration of computer models. Tech. rep.
URL `http://www.isds.duke.edu/{~}fei/samsi/Oct{_}09/01Sup-KenOHa.pdf`

Kersting, K., & Plagemann, C. (2007). Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24th Internationl Conference on Machine Learnign*.
URL `http://dl.acm.org/citation.cfm?id=1273546`

Keyfitz, N., & Caswell, H. (2005). *Applied Mathematical Demography*. New York: Springer, third ed.

Kirman, A. (2011). *Complex Economics: Individual and Collective Rationality*. Abingdon: Routledge.

Kirman, A. P., & Vriend, N. J. (2001). Evolving market structure: An ACE model of price dispersion and loyalty. *Journal of Economic Dynamics and Control*, *25*(3-4), 459–502.

Klabunde, A. (2013). Computational Economic Modelling of Migration. In *Agent Based Modelling and the Decision to Migrate*. Rostock.

Kleijnen, J. P. (2008). *Design and Analysis of Simulation Experiments*. New York: Springer.

Knio, O. M., & Le Maître, O. P. (2006). Uncertainty propagation in CFD using polynomial chaos decomposition. *Fluid Dynamics Research*, *38*(9), 616–640.

Kniveton, D., Smith, C., & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, *21*, S34–S40.

Lee, R. (1974). The formal dynamics of controlled populations and the echo, the boom and the bust. *Demography*, *11*(4), 563–585.

Lee, R. (2001). Demography Abandons Its Core.
URL `www.ceda.berkeley.edu/Publications/pdfs/rlee/FormalDemog.pdf`

Lee, R., Lee, S.-h., & Mason, A. (2011). Charting the Economic Life Cycle. *Population (English Edition)*, *34*(2008), 208–237.

Lee, S. J., Boscardin, W. J., Kirby, K. a., & Covinsky, K. E. (2014). Individualizing Life Expectancy Estimates for Older Adults Using the Gompertz Law of Human Mortality. *PLoS ONE*, *9*(9), e108540.

Leombruni, R., & Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications*, *355*(1), 103–109.

Luke, S. (2007). Finding Interesting Things : Population-based Adaptive Parameter Sweeping. In *Proceedings of the 9th annual conference on genetic and evolutionary computation*, (pp. 86–93). London.

Macunovich, D. J. (1998). Fertility and the Easterlin hypothesis: An assessment of the literature. *Journal of Population Economics*, *11*(1), 53–111.

Macy, M. W., & Willer, R. (2002). FROM FACTORS TO ACTORS : Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, *28*(1), 143–166.

Marks, R. (2012). Analysis and Synthesis: Multi-Agent Systems in the Social Sciences. *The Knowledge Engineering Review*, *27*(02), 1–31.

Marks, R. E. (2011). Validation and Model Selection : Three Similarity Measures Compared. *Complexity Economics*, *0*(1).

Massey, D. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population index*, *56*(1), 3–26.

Mincer, J. (1974). Schooling, Experience and Earnings.

Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford: Oxford University Press.

Moss, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social ...*.

MUCM (2011). Managing Uncertainty in Complex Models Toolkit.
URL `mucm.aston.ac.uk`

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, *118*(2), 549–599.

Murphy, M. (2006). The Role of Assortative Mating on Population Growth in Contemporary Developed Societies. In F. C. Billari, T. Fent, A. Prskawetz, & J. Scheffran (Eds.) *Agent Based Computational Modelling*, (pp. 61–84). Physica-Verlag.

Neal, R. (2010). MCMC using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. M. Meng (Eds.) *Handbook of Markov Chain Monte Carlo*. Chapman and Hall / CRC Press.

Ní Bhrolcháin, M., & Dyson, T. (2007). On Causation in Demography : Issues and Illustrations. *Population and Development Review*, *33*(1), 1–36.

Oakley, J. (1999). *Bayesian Uncertainty Analysis for Complex Computer Codes*. Ph.D. thesis, University of Sheffield.

Oakley, J. (2002). Eliciting Gaussian process priors for complex codes computer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *51*(1), 81–97.

Oakley, J., & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, *89*(4), 769–784.

Oakley, J., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B*, *66*(3), 751–769.

Oakley, J. E., & Youngman, B. (2011). Calibration of a Natural History Model of Bowel Cancer Progression Using Likelihood Emulation. *arXiv*, (pp. 1–24).

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*(10-11), 1290–1300.

O'Hagan, A. (2013). Polynomial Chaos : A Tutorial and Critique from a Statistician's Perspective.
URL `http://www.tonyohagan.co.uk/academic/pub.html`

O'Hagan, A., & Oakley, J. E. (2004). Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering and System Safety*, *85*(1-3), 239–248.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*, 641–646.

Pampel, F. C., & Peters, H. E. (1995). The Easterlin effect. *Annual review of sociology*, *21*, 163–94.

Parr, N. (2010). Satisfaction with life as an antecedent of fertility. *Demographic Research*, *22*, 635–662.

Pepelyshev, A., & Oakley, J. (2009). On the choice of correlation function and cross-validation for Gaussian processes. Tech. rep., Managing Uncertainty in Complex Models.
URL `http://www.mucm.ac.uk/Pages/Dissemination/TechnicalReports.html`

Poole, D., & Raftery, A. E. (2000). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association*, *95*(452), 1244.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson & Co.

Potts, J. (2000). *The New Evolutionary Microeconomics*. Cheltenham: Edward Elgar.

Qian, P. Z. G., & Wu, C. F. J. (2008). Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments. *Technometrics*, *50*(2), 192–204.

R Development Core Team (2015). R: A Language and Environment for Statistical Computing.
URL `http://www.r-project.org/`

Rasmussen, C., & Ghahramani, Z. (2002). Infinite mixtures of Gaussian process experts. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.) *Advances in Neural Information Processing Systems*, (pp. 881–888). MIT Press.

Rasmussen, C. E., & Nickisch, H. (2015). The GPML Toolbox version 3.6. Tech. rep.
URL `http://www.gaussianprocess.org/gpml/code/matlab/doc/`

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge MA: MIT Press.

Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2013). Integrated Modeling of European Migration. *Journal of the American Statistical Association*, *108*(503), 801–819.

Rehm, M. (2012). *Migration and Remittances: An Agent-Based Model*. Phd, New School for Social Research.

Rossiter, S. (2015). Simulation Design : Trans-Paradigm Best-Practice from Software Engineering. *Journal of Artificial Societies and Social Simulation*, *18*(3), 9.

Rossiter, S., Noble, J., & Bell, K. (2010). Social simulations : improving interdisciplinary understanding of scientific positioning and validity. *Journal of Artificial Societies and Social Simulation*, *13*(1), 10.

Roustant, O., Ginsbourger, D., & Deville, Y. (2012). DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodelling and Optimisation. *Journal of Statistical Software*, *51*(1).

Saltelli, A., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Chichester: Wiley.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice*. Chichester: John Wiley and Sons.

Santner, T. J., Williams, B. J., & Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer.

Ševčíková, H., Raftery, A. E., & Waddell, P. a. (2007a). Assessing uncertainty in urban simulations using Bayesian melding. *Transportation Research Part B: Methodological*, *41*(6), 652–669.

Ševčíková, H., Raftery, A. E., & Waddell, P. A. (2007b). Assessing uncertainty in urban simulations using Bayesian melding. *Transportation Research Part B: Methodological*, *41*(6), 652–669.

Silverman, E., Bijak, J., Courgeau, D., & Franck, R. (2014). System-Based Modelling: A New Paradigm for Demography. In *Agent-Based Modelling in Population Studies*. Leuven.

Silverman, E., Bijak, J., Hilton, J., Cao, V. D., & Noble, J. (2013). When demography met social simulation: a tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation*, *16*(4), 9.

Silverman, E., Bijak, J., & Noble, J. (2011). Feeding the Beast : Can Computational Demographic Models Free Us from the Tyranny of Data? In *Advances In Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*, (pp. 747–757). MIT Press.
URL `http://eprints.ecs.soton.ac.uk/22839/`

Silverman, E. S. (2007). *Methodological Investigations in Agent-Based Modelling*. Phd thesis, University of Leeds.

Smets, F., & Wouters, R. (2002). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, *1*(5), 1123–1175.

Spielauer, M. (2007). Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: survey and review. *International Journal of Microsimulation*, *1*(1), 35–53.

Squazzoni, F. (2012). *Agent-Based Computational Sociology*. Hoboken, NJ: Wiley, first ed.

Stan Development Team (2015). Stan Modeling Language Users Guide and Reference Manual.
URL `http://mc-stan.org/index.html`

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian Computation. *PLoS Computational Biology*, *9*(1).

Teitelbaum, M. S., & Winter, J. (2013). *The Global Spread of Fertility Decline: Population, Fear, and Uncertainty*. New Haven: Yale University Press.

Troitzsch, K. (2009). Not all explanations predict satisfactorily, and not all good predictions explain. *Journal of Artificial Societies and Social Simulation*, *12*(1), 10.

Urban, N. M., & Fricker, T. E. (2010). A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Computers & Geosciences*, *36*(6), 746–755.

Van Imhoff, E., & Post, W. (1998). Microsimulation methods for population projection. *Population. English selection*, *10*(1), 97–138.

Vernon, I., Goldstein, M., & Bower, R. G. (2010). Galaxy formation: A Bayesian uncertainty analysis. *Bayesian Analysis*, *5*(4), 619–670.

Wachter, K. W. (1991). Elusive Cycles: Are there Dynamically Possible Lee-Easterlin Models for U.S. Births? *Population Studies*, *45*(1), 109–135.

Wachter, K. W., & Lee, R. D. (1989). U.S. Births and Limit Cycle Models. *Demography*, *26*(1), 99–115.

Waldorf, B., & Byun, P. (2005). Meta-analysis of the impact of age structure on fertility. *Journal of Population Economics*, *18*(1), 15–40.

Werker, C., & Brenner, T. (2004). Empirical Calibration of Simulation Models. Tech. rep., Max Planck Institute of Economics.
URL `www.econstor.eu`

Wiegand, T., Revilla, E., & Knauer, F. (2004). Dealing with uncertainty in spatially explicit population models. *Biodiversity and Conservation*, *13*, 53–78.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications of Pure and Applied Mathematics*, *13*, 1–14.

Willekens, F. (2016). The Decision to Emigrate: A Simulation Model Based on the Theory of Planned Behaviour. In A. Grow, & J. Van Bavel (Eds.) *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*, chap. 10, (pp. 257–300). Basel: Springer.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, *41*(7-8), 1703–1729.

Wood, S. N. (2006). *Generalised Additive Models: An Introduction with R*. Boca Raton: Chapman and Hall / CRC Press.

Zaidi, A., Harding, A., & Williamson, P. (Eds.) (2009). *New frontiers in microsimulation modelling*. Farnham: Ashgate.

Zinn, S. (2011). *A Continuous-Time Microsimulation and First Steps Towards a Multi-Level Approach in Demography*. Doktor-ingenieur, Rostock University.

Zinn, S. (2015). MicSim: Performing Continuous-Time Microsimulation.
URL `http://cran.r-project.org/package=MicSim`

Zinn, S., & Himmelspach, J. (2009). Mic-Core: A Tool for Microsimulation. In E. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls (Ed.) *Proceedings of the 2009 Winter Simulation Conference*, (pp. 992–1002). IEEE.
URL `http://ieeexplore.ieee.org`