



Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

Discussion Papers in Economics and Econometrics

CONGRUENCE AND ENCOMPASSING

Christophe Bontemps
and
Grayham E Mizon

No. 0107

This paper is available on our website
<http://www.soton.ac.uk/~econweb/dp/dp01.html>

CONGRUENCE AND ENCOMPASSING

Christophe Bontemps

INRA,
Toulouse, France.

Grayham E. Mizon *

Economics Department,
University of Southampton, UK.

June 1996. This revision January 2001.

Abstract

This paper considers the relationship between congruence and encompassing. Congruence is defined formally, and though it is not testable directly, it can be tested indirectly via tests of misspecification. Empirically more than one model can appear congruent, but that which encompasses its rivals is dominant, will encompass all models nested within it, and accurately predict the misspecifications of non-congruent models. These results are consistent with a general-to-specific modelling strategy being successful in practice. An empirical example illustrates these points.

JEL classification: C12, C51, C52

Keywords: data generation process, empirical model, general-to-specific modelling, encompassing, misspecification, specification

*The first version this paper was written whilst the first author was visiting the EUI between October and December 1995, and he wishes to thank the researchers and students of the Department of Economics for their warm hospitality, and especially Massimiliano Marcellino for many discussions on encompassing. Valuable comments were received from John Aldrich, Grant Hillier, Maozu Lu, Mark Salmon, Bernt Stigum and Pravin Trivedi when the second author presented earlier versions of the paper to seminars at the EUI and in Oslo and Southampton. Jean-Pierre Florens, David Hendry, Søren Johansen, and Jean-François Richard are also thanked for stimulating conversations on the topic. Financial support from the EUI Research Council and the UK ESRC under grant L138251009 is gratefully acknowledged.

1 Introduction

Economists and econometricians who undertake empirical modelling face many problems, not the least of which is how to deal with the fact that, with very rare exceptions, all models analysed are approximations to the actual processes that generate the observed data. The simplest approach to this problem is to ignore it, but potentially this is most likely to lead to invalid and misleading inferences. An alternative approach recognises the problem explicitly by adopting misspecification-robust inference procedures, such as using heteroscedastic-autocorrelation-consistent standard errors (see *inter alia* Andrews, 1991 and Gouriéroux and Monfort, 1995). However, this approach does nothing specific to tackle the problem, rather it is analogous to taking out an insurance policy whilst continuing to engage in hazardous activities - with the insurance eventually becoming more and more expensive, or redundant. Another approach, in addition to recognising the problem explicitly, aims to alleviate it by establishing a statistically and economically well specified framework within which to conduct modelling. This third approach is associated with the LSE methodology (see *inter alia* Hendry, 1995 and Mizon, 1995a) in which the concepts of congruence and encompassing have important roles. Briefly, congruence is the property of a model that has fully exploited all the information implicitly available once an investigator has chosen a set of variables to be used in modelling. Encompassing on the other hand, is the property of a model that can account for the results obtained from rival models, and in that sense makes the rivals inferentially redundant.

The present analysis emphasises the importance of having a congruent model of the joint distribution of all relevant variables as a statistical framework for the evaluation of competing models nested within it. Such evaluation of models done using parsimonious encompassing has many advantages, including: yielding transitive relations amongst the models, and avoiding the anomaly of general models failing to encompass simplifications of themselves. By applying the concept of parsimonious encompassing to the relationship between a model and the corresponding data generation process (DGP) a formal definition of congruence is given. Also it is argued that nesting is more than one model being an algebraic simplification of another, that the congruence of a model is a sufficient condition for it to nest and encompass a simplification (parametric or nonparametric) of itself, and consequently it plays a crucial role in the application of the encompassing principle.

The outline of the chapter is as follows. The next section presents notation and discusses the concepts of the data generation process (DGP), the local data generating process (LDGP), an econometric model, and nesting, all of which have important roles in the rest of the chapter. Section 3 discusses the parametric relationship between the LDGP and an econometric model, prior to defining parametric encompassing and parsimonious encompassing. In section 4 linear and nonlinear parametric, and nonparametric examples of a general model failing to encompass a special case of itself are presented. Section 5 uses the concept of parsimonious encompassing to provide a formal definition of congruence, which is then shown to be a sufficient condition for the anomaly of section 4 not to occur. Since the hypothesis defining congruence is not testable directly, alternative indirect methods for testing congruence are discussed in section 6. This discussion highlights the distinction between misspecification and specification testing of models, and some of the limitations of misspecification hypotheses are analysed. An illustration of an

empirically congruent model that is not the DGP is presented in section 7. Section 8 presents conclusions.

2 Preliminaries

It is assumed that a statistical representation can be given to the process by which observations are made on the phenomena of interest in the economy being studied. This is called the data generation process.

2.1 The Data Generation Process (DGP)

Let the vector of N real random variables \mathbf{w}_t characterize the economy under analysis. At each point in time $t \in \mathcal{T}$, the joint density $D_{\mathbf{w}_t}(\mathbf{w}_t | \mathcal{F}_{t-1})$ of \mathbf{w}_t has a sample space Ω and event space \mathcal{F}_{t-1} . This joint density is a statistical representation of the economy, and as such is the data generation process for \mathbf{w}_t . It is also assumed that this DGP can be represented as a vector stochastic process, with sufficient continuity to make it meaningful to postulate d parameters $\boldsymbol{\delta} \in \mathcal{D}^d \subseteq \mathbb{R}^d$ which do not depend on \mathcal{F}_{t-1} at any t . It is not necessary to restrict the parameters to be constant over time, or to exclude transient parameters. The history of the stochastic process $\{\mathbf{w}_t\}$ up to time $(t-1)$ is denoted by the notation $\mathbf{W}_{t-1} = (\mathbf{W}_0, \mathbf{w}_1, \dots, \mathbf{w}_{t-1}) = (\mathbf{W}_0, \mathbf{W}_{t-1}^1)$, where \mathbf{W}_0 is the set of initial conditions. Then, for a sample period $t = 1, \dots, T$, the DGP is denoted $D_{\mathbf{W}}(\mathbf{W}_T^1 | \mathbf{W}_0, \boldsymbol{\delta}_0)$, and is sequentially factorized as:

$$D_{\mathbf{W}}(\mathbf{W}_T^1 | \mathbf{W}_0, \boldsymbol{\delta}_0) = \prod_{t=1}^T D_{\mathbf{w}}(\mathbf{w}_t | \mathbf{W}_{t-1}, \boldsymbol{\kappa}_t) \quad (1)$$

where $\boldsymbol{\delta}_0$ is a particular point in parameter space, and $\mathbf{g}(\boldsymbol{\delta}_0) = (\boldsymbol{\kappa}_1 \dots \boldsymbol{\kappa}_T)$ for a 1-1 function $\mathbf{g}(\cdot)$, so that the $\boldsymbol{\kappa}_t$ may be non-constant over time due to regime shifts and structural breaks – see Hendry and Mizon (1998) and Hendry and Mizon (2000) for discussion of the distinction between these parameter changes.

2.2 Econometric Models

Even when a sample of T time series observations is available on all N variables in \mathbf{w}_t it is not possible to analyse the inter-relationships between all of them since N is usually large relative to T , and more importantly particular interest may be in a transformed sub-set \mathbf{x}_t of \mathbf{w}_t . Hence econometric modelling will typically employ knowledge from economic theory and the statistical properties of the observed \mathbf{x}_t to guide model specification. An econometric model $f_{\mathbf{x}}(\cdot)$ for $n < N$ variables \mathbf{x}_t implicitly derived from \mathbf{w}_t by marginalization, aggregation and transformation is denoted by:

$$f_{\mathbf{x}}(\mathbf{X}_T^1 | \mathbf{X}_0, \boldsymbol{\xi}) = \prod_{t=1}^T f_{\mathbf{x}}(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\xi}) \quad \text{where } \boldsymbol{\xi} = (\xi_1, \dots, \xi_k)' \in \Xi \subseteq \mathbb{R}^k \quad (2)$$

when $f_x(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\xi})$ is the postulated sequential joint density at time t . It is assumed that $k < d$ and $\boldsymbol{\xi}$ represents the constant parameters postulated by the modeller, so any time-dependent effects have been reparameterized accordingly (as in a ‘structural time-series model’ re-represented as an ARIMA process: see Harvey, 1993).

From the theory of reduction (see, *inter alia*, Hendry, 1995, and Mizon, 1995a), there exists a local DGP (LDGP) for the chosen variables \mathbf{x}_t :

$$D_x(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\phi}) \quad \text{where } \boldsymbol{\phi} = (\phi_1, \dots, \phi_s)' \in \boldsymbol{\Phi} \subseteq \mathbb{R}^s \quad (3)$$

which is derived from $D_w(\mathbf{w}_t | \cdot)$ by reduction. In general, $f_x(\cdot) \neq D_x(\cdot)$, and this divergence has to be taken into account when making inferences about $\boldsymbol{\xi}$ (see *inter alia* Cox (1961) and White (1982)). Indeed, the modeller is free to specify any form of $f_x(\cdot)$ independently of what process has actually generated the data which is almost always unknown. Precisely because the form of the econometric model is under the control of the investigator it is important that criteria are employed to assist in the specification of models. Designing a model to fully exploit the information available to the investigator is requiring the model to be congruent, and ensuring that it is at least as good as its rivals means that the model is encompassing. Both these concepts are analysed in more detail below.

2.3 Nesting

Not only is it usual for an econometric model to differ from the LDGP, but there will often be more than one econometric model to be considered for the analysis of a particular economic phenomenon. For example, one parametric probability model for \mathbf{x}_t might consist of the family of sequential densities indexed by the parameter vector $\boldsymbol{\theta}$:

$$\mathcal{M}_1 = \{g_1(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\beta}_1), \quad \boldsymbol{\beta}_1 \in \mathcal{B}_1 \subseteq \mathbb{R}^{p_1}\} \quad (4)$$

when $\mathbf{X}_{t-1} = (\mathbf{X}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}) = (\mathbf{X}_0, \mathbf{X}_{t-1}^1)$, with \mathbf{X}_0 being initial conditions. Denoting this probability model by \mathcal{M}_1 and considering the alternative probability model \mathcal{M}_2 :

$$\mathcal{M}_2 = \{g_2(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\beta}_2), \quad \boldsymbol{\beta}_2 \in \mathcal{B}_2 \subseteq \mathbb{R}^q\} \quad (5)$$

enables nesting to be defined.

Definition 1 (Nesting). \mathcal{M}_1 is nested within \mathcal{M}_2 if and only if $\mathcal{M}_1 \subseteq \mathcal{M}_2$.

Whenever \mathcal{M}_1 and \mathcal{M}_2 do not satisfy the conditions in this definition they are said to be non-nested. A particular form of nesting that yields an heuristic explanation of nesting arises when $\mathcal{B}_1 \subseteq \mathcal{B}_2$ with $p_1 < p_2$, so that \mathcal{M}_1 is nested within \mathcal{M}_2 as a result of the restrictions on the parameter spaces defining the reduction from \mathcal{B}_2 to \mathcal{B}_1 . In this case both models are purporting to represent the same probability distribution, but \mathcal{M}_1 is a parametric simplification of \mathcal{M}_2 . Note though that one model being a parametric simplification of another does not by itself imply that it is nested in the more general model - see Lu and Mizon (2000) and the discussion below for more details. Throughout the paper \mathcal{M}_1 and \mathcal{M}_2 are used as generic alternative parametric probability models of the same distribution, though the particular distribution will vary from example to example.

When the purpose of modelling is to use econometric models to learn more about the relationships amongst the elements of \mathbf{x}_t in the DGP (as in theory testing and economic policy analysis) there is a premium to having an econometric model that provides a good approximation to the LDGP. Congruence, which is concerned with the relationship between a model f_x and the LDGP D_x , is discussed in detail in section 5. When there is more than one econometric model available it is important to have criteria for assessing their merits, especially if they each appear to be congruent. Encompassing is concerned with the relationships between rival models, and is discussed in detail in the next section.

3 Encompassing

Whether models are nested or non-nested it is important to be able to compare them and evaluate their relative merits, and “The encompassing principle is concerned with the ability of a model to account for the behaviour of others, or less ambitiously, to explain the behaviour of relevant characteristics of other models.” (Mizon (1984)). A model \mathcal{M}_1 encompasses another model \mathcal{M}_2 if \mathcal{M}_1 can account for results obtained from \mathcal{M}_2 . In other words, anything that can be done using \mathcal{M}_2 can be done equally well using \mathcal{M}_1 , and so once \mathcal{M}_1 is available \mathcal{M}_2 has nothing further to offer. These heuristic concepts are made more formal in the next two sub-sections.

3.1 Parametric Encompassing

The concept of encompassing considered in this paper is that of population encompassing in a Neyman-Pearson hypothesis testing framework, which is in accord with the approach in Mizon (1984), Mizon and Richard (1986), and as formally defined in Hendry and Richard (1989). Underlying all empirical econometric analyses is an information set (collection of variables or their sigma field), and a corresponding probability space. This information set has to be sufficiently general to include all the variables thought to be relevant for the empirical implementation of theoretical models in the form of statistical models. It is also important that this information set include the variables needed for all competing models that are to be compared, otherwise there can be non-comparabilities. Let these variables be \mathbf{x}_t and the LDGP that generates these variables be the joint density $D_x(\mathbf{x}_t|\mathbf{X}_{t-1}, \phi)$ as defined in (3), at the particular parameter value $\phi = \phi_0$. Also let a parametric statistical model of the joint distribution be $\mathcal{M}_g = \{f_g(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^k\}$ as defined in (2). Let $\hat{\boldsymbol{\xi}}$ be the maximum likelihood estimator of $\boldsymbol{\xi}$ so that $\hat{\boldsymbol{\xi}} \xrightarrow[\mathcal{M}_g]{P} \boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}} \xrightarrow[\text{LDGP}]{P} \boldsymbol{\xi}(\phi_0) = \boldsymbol{\xi}_0$ which is the pseudo-true value of $\hat{\boldsymbol{\xi}}$. Note that the parameters of a model are not arbitrary in that \mathcal{M}_g and its parameterization $\boldsymbol{\xi}$ are chosen to correspond to phenomena of interest (such as elasticities and partial responses within the chosen probability space), although there may be observationally equivalent parameterizations. Equally, estimation is an issue separate from that of model specification, and in particular an arbitrary estimation method does not define the parameters of a model. Given a particular set of parameters there usually exist good and bad estimation methods for them, some of which are consistent and efficient, whilst others are inconsistent etc. The parameterization is an integral part of the specification of a model, and in many circumstances can be related to the moments

of the variables \mathbf{x}_t . Given the specification of the model \mathcal{M}_g the probability limit under \mathcal{M}_g of the maximum likelihood estimator $\hat{\boldsymbol{\xi}}$ (the optimal estimator under \mathcal{M}_g) defines the parameter $\boldsymbol{\xi}$. However, in general $\boldsymbol{\xi}$ will differ from its pseudo-true value $\boldsymbol{\xi}_0$, as a consequence of inadequate specification of \mathcal{M}_g .

With these definitions it is possible to give a more formal definition of encompassing in the present context.

Definition 2 (Encompassing). \mathcal{M}_1 encompasses \mathcal{M}_2 (denoted $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$) with $\mathcal{M}_1 = \{g_1(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\beta}_1), \boldsymbol{\beta}_1 \in \mathcal{B}_1 \subset \mathbb{R}^{p_1}\}$ and $\mathcal{M}_2 = \{g_2(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\beta}_2), \boldsymbol{\beta}_2 \in \mathcal{B}_2 \subset \mathbb{R}^{p_2}\}$ if and only if $\boldsymbol{\beta}_{20} = \mathbf{b}_{21}(\boldsymbol{\beta}_{10})$ when $\boldsymbol{\beta}_{i0}$ is the pseudo true value of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_i$ of $\boldsymbol{\beta}_i$ $i = 1, 2$, and $\mathbf{b}_{21}(\boldsymbol{\beta}_{10})$ is the binding function given by $\hat{\boldsymbol{\beta}}_2 \xrightarrow[\mathcal{M}_1]{P} \mathbf{b}_{21}(\boldsymbol{\beta}_{10})$. (See Mizon and Richard (1986), Hendry and Richard (1989), and Gouriéroux and Monfort (1995))

Note that this definition of encompassing applies when \mathcal{M}_1 and \mathcal{M}_2 are non-nested as well as nested. However, Hendry and Richard (1989) showed that when \mathcal{M}_1 and \mathcal{M}_2 are non-nested $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ is equivalent to \mathcal{M}_1 being a valid reduction of the minimum completing model $\mathcal{M}_c = \mathcal{M}_1 \cup \mathcal{M}_2^\perp$ (so that $\mathcal{M}_1, \mathcal{M}_2 \subset \mathcal{M}_c$) when \mathcal{M}_2^\perp is the model which represents all aspects of \mathcal{M}_2 that are not contained in \mathcal{M}_1 . When this condition is satisfied \mathcal{M}_1 is said to parsimoniously encompass \mathcal{M}_c a concept to which attention is turned in the next sub-section.

3.2 Parsimonious Encompassing

Let \mathcal{M}_s be a sub-model of \mathcal{M}_g (as defined in section 3.1) which has the form $\mathcal{M}_s = \{f_s(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{R}^n\}$ with $n = k - r$, $r > 0$ being the number of linearly independent restrictions on Ξ which define \mathcal{A} . Note that $\mathcal{M}_s \subseteq \mathcal{M}_g$ as a result of dimension reducing constraints on the parameter space. Further, let the r constraint equations be $\boldsymbol{\psi}(\boldsymbol{\xi}) = \mathbf{0}$ which are such that $\text{rank} \left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\xi}'} \right)_{\boldsymbol{\xi}_0} = r$, and consider the isomorphic reparameterization $\boldsymbol{\lambda}$ of $\boldsymbol{\xi} : \boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\xi}) = (\boldsymbol{\lambda}_1(\boldsymbol{\xi})', \boldsymbol{\lambda}_2(\boldsymbol{\xi})')' = (\boldsymbol{\alpha}', \boldsymbol{\psi}(\boldsymbol{\xi})')'$ which is such that $\text{rank} \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\xi}'} \right)_{\boldsymbol{\xi}_0} = k$ and $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \in \Lambda_1 \times \Lambda_2$ (see Sargan (1988) for discussion of such a reparameterization). Hence $\mathcal{A} = \{\boldsymbol{\alpha} \mid \boldsymbol{\alpha} = \boldsymbol{\lambda}_1(\boldsymbol{\xi}) \text{ such that } \boldsymbol{\psi}(\boldsymbol{\xi}) = \mathbf{0} \forall \boldsymbol{\xi} \in \Xi\}$. With this notation it is possible to define parsimonious encompassing in the following way.

Definition 3 (Parsimonious Encompassing). \mathcal{M}_s parsimoniously encompasses \mathcal{M}_g with respect to $\boldsymbol{\xi}$ (denoted $\mathcal{M}_s \mathcal{E}_p(\boldsymbol{\xi}) \mathcal{M}_g$) if and only if $\boldsymbol{\psi}(\boldsymbol{\xi}) = \mathbf{0}$ and $\mathcal{M}_s \mathcal{E}(\boldsymbol{\xi}) \mathcal{M}_g$.

Hence when $\mathcal{M}_s \mathcal{E}_p(\boldsymbol{\xi}) \mathcal{M}_g$ the former model is a valid reduction of the latter, and so efficient inference on $\boldsymbol{\alpha}$ can be obtained from \mathcal{M}_s . This interpretation of parsimonious encompassing is used in section 5 to provide a formal definition of congruence.

Note that whenever $\mathcal{M}_1 \subset \mathcal{M}_2$ and $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ then $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_2$. Also whenever \mathcal{M}_1 and \mathcal{M}_2 are non-nested, but $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$, then $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_c$ with $\mathcal{M}_c = \mathcal{M}_1 \cup \mathcal{M}_2^\perp$. Hence parsimonious encompassing plays an important role in the comparison of alternative models whether they are nested or non-nested. Therefore it is relevant to ask whether there is a separate role for encompassing as opposed to parsimonious encompassing: that there is can be seen from the following argument. Parsimonious encompassing is the property that a model is a valid reduction

of a more general model. Noting that in practice, particularly if a general-to-simple modelling strategy is adopted, the general model will embed the different econometric models implementing rival economic theories for the phenomenon of interest, searching for the model that parsimoniously encompasses the general model is specification searching amongst models nested within a common general model. If after completing a modelling exercise and having determined which model is preferred, further information becomes available, this raises the question as to whether the existing results are robust to the extension of the information set. The further information may take the form of new a priori theories which imply an extension of the information set to embrace the empirical models implementing these new theories. Testing the preferred model's ability to encompass the newly available rival models (which can be done using non-nested test statistics) is a form of misspecification testing. Indeed it can be interpreted as testing the adequacy of the information set originally chosen for the modelling, against extensions suggested by rival theories. However, given that one degree of freedom non-nested test statistics (such as those of Cox, 1961 and Davidson and MacKinnon, 1981) have low power against a wide range of alternatives (see e.g., Mizon and Richard, 1986) there will always be an argument for re-commencing the modelling with a more general information set.

The above analysis has indicated that underlying all encompassing comparisons there is a general model which has the rival models as simplifications of it. This general model might be one of the rival models which nests all its rivals, or it might be a completing model providing the statistical framework for model comparison. The next section discusses the relationship between encompassing and general models.

4 Encompassing and General Models

The theory of reduction implies that for the variables in \mathbf{x}_t there is a local DGP $D_{\mathbf{x}}(\mathbf{x}_t | \mathbf{X}_{t-1}, \phi)$, which then by definition encompasses all econometric models of the same distribution having the form $f_{\mathbf{x}}(\mathbf{x}_t | \mathbf{X}_{t-1}, \xi)$. Therefore, it might be expected heuristically that a general model will encompass simplifications of itself. Indeed Hendry (1995) states:

“A general model must encompass a special case thereof: indeed, misspecification analysis could not work without this result.”

However, that this need not always be the case was pointed out by Gouriéroux and Monfort (1995) who state:

“A model \mathcal{M}_2 nesting a model \mathcal{M}_1 does not necessarily encompass \mathcal{M}_1 .”

In the following sub-sections four examples are presented of general models failing to encompass algebraic simplifications of themselves. Possible reasons for this phenomenon are then discussed.

4.1 A linear parametric example

Consider the two linear models \mathcal{M}_1 and \mathcal{M}_2 of the conditional distribution $y_t | z_t, u_t$ defined by:

$$\begin{aligned} \mathcal{M}_1 & : y_t = \beta z_t + \zeta_{1t} & \zeta_{1t} & \sim \mathcal{N}(0, \sigma_{\zeta_1}^2) \\ \mathcal{M}_2 & : y_t = \gamma z_t + \delta u_t + \varepsilon_{1t} & \varepsilon_{1t} & \sim \mathcal{N}(0, \sigma_{\varepsilon_1}^2) \end{aligned}$$

where (z_t, u_t) are independent and identically distributed variables, with ζ_{1t} and ε_{1t} asserted to be zero mean, constant variance Gaussian white noise processes, and the parameters β and $\pi' = (\gamma, \delta)$ belong respectively to \mathbb{R}^+ and $\mathbb{R}^+ \times \mathbb{R}^+$. Clearly, \mathcal{M}_1 seems to be nested within \mathcal{M}_2 in the sense that \mathcal{M}_1 corresponds to the special case of \mathcal{M}_2 in which $\delta = 0$, and the parameter space for β is included in that of π . Then, \mathcal{M}_2 should be able to explain the results of its own sub-model, and so encompass \mathcal{M}_1 . However, in general, this is not the case.

In the framework of Gouriéroux and Monfort (1995), who proposed this example, whether or not \mathcal{M}_2 encompasses \mathcal{M}_1 depends on the DGP. For example, if the DGP were non-linear of the form:

$$\mathcal{P}_0 \quad : \quad y_t = m_0(z_t, u_t) + v_t \quad v_t \sim \mathcal{N}(0, \sigma_0^2) \quad (6)$$

then \mathcal{M}_1 and \mathcal{M}_2 are misspecified, and the pseudo-true value β_0 of β under \mathcal{P}_0 is the L^2 -projection of $m_0(z_t, u_t)$ onto the half line A_x defined by $\{\beta z_t, \beta \geq 0\}$, and π_0 , the pseudo-true value of π is the corresponding projection onto the cone C_{zu} defined by $\{\gamma z_t + \delta u_t, \gamma, \delta \geq 0\}$ (see figure 1 of Gouriéroux and Monfort (1995)). The \mathcal{M}_2 -pseudo-true value $\beta_2 = \underset{\mathcal{M}_2}{p \lim} (\hat{\beta}) = \beta(\pi_0)$ of β is in general different from β_0 (i.e. $\beta_2 = \beta(\pi_0) \neq \beta_0$), so that \mathcal{M}_2 does not encompass \mathcal{M}_1 in general. Gouriéroux and Monfort (1995) further argue that $\mathcal{M}_2 \mathcal{E} \mathcal{M}_1$ will occur though for those particular forms of the DGP in which $\beta_2 = \beta(\pi_0) = \beta_0$.

One possible interpretation of this result is that whether or not a model encompasses a simplification of itself depends fortuitously on the nature of the unknown DGP. In this vein Gouriéroux and Monfort (1995) argue that whenever the unknown DGP is nonlinear then the law of iterated projections will not apply in general, and so a larger model may not encompass an algebraic simplification of itself, as in the above example. However, if in the above example the parameter spaces are unrestricted so that $\beta \in \mathbb{R}$ and $\pi \in \mathbb{R}^2$ and the assumption that z_t and u_t are identically independently distributed is dropped, but \mathcal{M}_2 now includes the false assertion that z_t and u_t are orthogonal, then even if the DGP were linear \mathcal{M}_2 would fail to encompass \mathcal{M}_1 . This example, was used by Govaerts, Hendry and Richard (1994) and Hendry (1995) to discuss the apparent anomaly of some models failing to encompass algebraic simplifications of themselves. The conclusion drawn by these authors was that as a result of the false assertion that z_t and u_t are orthogonal, \mathcal{M}_1 and \mathcal{M}_2 are non-nested - the orthogonality hypothesis having no role in \mathcal{M}_1 despite it being a part of the specification of \mathcal{M}_2 . Hence the problem is not simply associated with nonlinear DGPs. In particular, by reconsidering the models from the perspective of the joint density of (y_t, z_t, u_t) , Hendry (1995) argues that the conditional model associated with \mathcal{M}_1 is not nested within the one associated with \mathcal{M}_2 .

The anomaly for the linear example presented in this sub-section has a nonlinear counterpart, as shown in the next sub-section.

4.2 A Nonlinear Parametric Example

Let the two non-linear conditional models \mathcal{M}_1 and \mathcal{M}_2 be defined as:

$$\begin{aligned} \mathcal{M}_1 \quad : \quad & y_t = f_2(z_t, \beta) + \zeta_{2t} \\ \mathcal{M}_2 \quad : \quad & y_t = f_2(z_t, \gamma) + g_2(u_t, \delta) + \varepsilon_{2t} \end{aligned}$$

where f_2 and g_2 are two regular functions defined on $\mathbb{R} \times \mathbb{R}$. Since \mathcal{M}_1 is a special case ($g_2 = 0$) of \mathcal{M}_2 it would appear to be nested in it, but it is not in general encompassed by \mathcal{M}_2 . The restrictions on the form of \mathcal{M}_2 , namely that the conditional mean of y_t is additively separable in z_t and u_t , will result in \mathcal{M}_2 failing to encompass \mathcal{M}_1 whenever the DGP does not have this property. For example, were the DGP to have the form of (6) then $\mathcal{M}_2 \mathcal{E} \mathcal{M}_1$ will not hold, because the pseudo-true values of β under \mathcal{P}_0 and under \mathcal{M}_2 differ, and so $f_2(z_t, \beta_0) \neq f_2(z_t, \beta_2)$. Here again, the situation in terms of pseudo-true values, may be viewed as a sequence of projections which lead to different values of f_2 . Hence the failure of $\mathcal{M}_2 \mathcal{E} \mathcal{M}_1$ to hold arises because of a false auxiliary hypothesis involved in \mathcal{M}_2 , which can be interpreted as rendering \mathcal{M}_2 and \mathcal{M}_1 non-nested.

4.3 Comfac Example

Mizon (1993) in the process of illustrating the importance of modelling in a framework which has a congruent general model, produced the following illustration of a model failing to encompass an algebraic simplification of itself (the example is explored in more detail in Mizon (1995b)):

$$\begin{aligned}\mathcal{M}_1 : y_t &= \delta z_t + \zeta_{3t} \\ \mathcal{M}_2 : y_t &= \rho y_{t-1} + \gamma z_t - \rho \gamma z_{t-1} + \varepsilon_{3t}\end{aligned}$$

in which z_t is a white noise process with constant variance σ_z^2 distributed independently of ζ_{3t} and ε_{3t} . In this example it appears that \mathcal{M}_1 is nested within \mathcal{M}_2 with the restriction $\rho = 0$ rendering the models equivalent. However, when the DGP lies in the class of densities characterized by the stationary ($|\alpha| < 1$) linear partial adjustment model:

$$\mathcal{P}_0 : y_t = \alpha y_{t-1} + \beta z_t + v_t$$

the pseudo-true value of $\hat{\delta}$ is $\delta_0 = \beta$, whereas the pseudo-true values ρ_0 and γ_0 of $\hat{\rho}$ and $\hat{\gamma}$ respectively are given as the solutions of the equations:

$$\begin{aligned}\rho_0 - \alpha &= -\alpha \gamma_0 (\gamma_0 - \beta) / [\gamma_0^2 - 2\beta \gamma_0 + \sigma_0^2 / (1 - \alpha^2) \sigma_z^2] \\ \gamma_0 - \beta &= -\alpha \beta \rho_0 / (1 + \rho_0^2)\end{aligned}$$

The solution for γ_0 is:

$$\gamma_0 = \beta [(1 + \rho_0^2) - \alpha \rho_0] / [1 + \rho_0^2]$$

when ρ_0 is a root of the fifth order polynomial:

$$aZ^5 + bZ^4 + cZ^3 + dZ^2 + eZ + f = 0$$

with:

$$\begin{aligned}a &= \sigma_0^2 + \beta^2 \sigma_z^2 \alpha^2 - \beta^2 \sigma_z^2 & b &= \beta^2 \sigma_z^2 \alpha - \alpha \sigma_0^2 - \beta^2 \sigma_z^2 \alpha^3 \\ c &= -2\beta^2 \sigma_z^2 + 2\beta^2 \sigma_z^2 \alpha^2 + 2\sigma_0^2 & d &= -2\alpha \sigma_0^2 - 2\beta^2 \sigma_z^2 \alpha^3 + 2\beta^2 \sigma_z^2 \alpha \\ e &= -\beta^2 \sigma_z^2 + \beta^2 \sigma_z^2 \alpha^4 + \sigma_0^2 & f &= \beta^2 \sigma_z^2 \alpha - \alpha \sigma_0^2 - \beta^2 \sigma_z^2 \alpha^3\end{aligned}$$

Further, the \mathcal{M}_2 -pseudo-true value of $\hat{\delta}$ is $\delta_2 = \gamma_0$, but $\gamma_0 = \beta$ if and only if $\alpha = 0$ in which case $\rho_0 = 0, \pm i, \pm i$ thus implying that $\mathcal{M}_2 \mathcal{E} \mathcal{M}_1$ if and only if the DGP lies in the class of static

densities - the four roots on the unit circle are ruled out by the stationarity condition. Hence in general $\delta_2 = \gamma_0 \neq \beta$ and so \mathcal{M}_2 does not encompass \mathcal{M}_1 despite the latter being an algebraic simplification of the former. Mizon (1995b) attributed this failure of \mathcal{M}_2 to encompass \mathcal{M}_1 to the false common factor auxiliary hypothesis implicit in \mathcal{M}_2 rendering it non-nested relative to \mathcal{M}_1 , since the latter model does not involve the common factor hypothesis.

Similar anomalies can arise in a nonparametric context, as the next sub-section shows for two nonparametric regression models.

4.4 A Nonparametric Example

Let \mathcal{M}_1 and \mathcal{M}_2 be two conditional models defined with respect to the conditional distribution $y_t | \sigma(z_t, u_t)$ but without any specified parametric form, when $\sigma(z_t, u_t)$ is the sigma field generated by z_t and u_t . \mathcal{M}_1 hypothesises that the conditional mean only includes z_t , whilst \mathcal{M}_2 excludes the variable z_t :

$$\begin{aligned} \mathcal{M}_1 & : \quad \mathbf{E} [y_t | \sigma(z_t, u_t)] = \mathbf{E} [y_t | \sigma(z_t)] \\ \mathcal{M}_2 & : \quad \mathbf{E} [y_t | \sigma(z_t, u_t)] = \mathbf{E} [y_t | \sigma(u_t)] \end{aligned}$$

Defining the functions f , and g to be the following conditional expectations :

$$\begin{aligned} f(\cdot) & = \mathbf{E} [y_t | z_t = \cdot] \\ g(\cdot) & = \mathbf{E} [y_t | u_t = \cdot] \end{aligned}$$

yields the nonparametric non-nested regression models :

$$\begin{aligned} \mathcal{M}_1 & : \quad y_t = f(z_t) + \zeta_{4t} \\ \mathcal{M}_2 & : \quad y_t = g(u_t) + \varepsilon_{4t} \end{aligned}$$

In this context a natural nesting model is:

$$\mathcal{M} : \quad y_t = m(z_t, u_t) + \eta_{4t}$$

where

$$m(z_t, u_t) = \mathbf{E} [y_t | \sigma(z_t, u_t)]$$

which as a result of unrestrictedly conditioning on both z_t and u_t nests both \mathcal{M}_1 and \mathcal{M}_2 . Indeed, \mathcal{M} is conditioned on the sigma-field $\sigma(z_t, u_t)$, generated by the variables (z_t, u_t) , and so will encompass all nonparametric sub-models defined by hypotheses relating $\sigma(z_t, u_t)$ to sigma-fields included within it. Consequently, \mathcal{M} will encompass \mathcal{M}_1 and \mathcal{M}_2 whatever the DGP is, provided that the conditioning on $\sigma(z_t, u_t)$ is valid.

A situation similar to the previous parametric ones, arises in considering the alternative ‘nesting’ model \mathcal{M}' :

$$\mathcal{M}' : \quad y_t = f(z_t) + g(u_t) + \nu_t$$

In this case, \mathcal{M}_1 corresponds to the special case where $g \equiv 0$, and \mathcal{M}_2 corresponds to the special case where $f \equiv 0$, and so they appear to be nested in \mathcal{M}' . However, as a result of incorporating

the restriction of additive separability in its specification \mathcal{M}' does not nest either \mathcal{M}_1 or \mathcal{M}_2 , both of which are defined with respect to the distribution $y_t | \sigma(z_t, u_t)$ with $\sigma(z_t, u_t)$ unrestricted. So \mathcal{M}' will not encompass the sub-models \mathcal{M}_1 and \mathcal{M}_2 . This problem can be described alternatively as one in which \mathcal{M}' is based on S' , the union of the sigma fields $\sigma(z_t)$ and $\sigma(u_t)$, and as such is not a sigma field itself without the restriction of additive separability.

The next section provides a formal definition of congruence, which is then shown to be a sufficient condition for the anomaly discussed in this section not to occur.

5 Congruence

Congruence has been discussed in numerous places such as Hendry (1985), Hendry (1987), Hendry and Mizon (1990), and Mizon (1995a). From Hendry (1995) it can be defined in the following way.

Models are said to be congruent when they have:

- homoscedastic, innovation errors;
- weakly exogenous conditioning variables for the parameters of interest;
- constant, invariant parameters of interest;
- theory consistent, identifiable structures;
- data admissible formulations on accurate observations.

Thus, for example, if $\mathcal{M}_g = \{f_g(\mathbf{x}_t | \mathbf{X}_{t-1}, \boldsymbol{\xi}) \quad \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^k\}$ is congruent it will be theory consistent (coherent with *a priori* theory), data coherent (coherent with observed sample information), and data admissible (coherent with the properties of the measurement system). Detailed explanations of each of these conditions, including discussion of how they might be tested, can be found in Hendry (1995) and Mizon (1995a). Theory consistency does not require that the model conform in all aspects to a very detailed statement of a theory, such as one associated with an inter-temporal optimization problem. Rather it requires that \mathbf{x}_t incorporate the relevant set of variables (transformed if necessary) to enable alternative densities to be specified that include the parameters of interest. Alternatively stated, a congruent model will be coherent with low level theory, and will provide a framework within which the hypotheses of high level theory can be tested. The requirement that the model errors are innovations with respect to \mathbf{X}_{t-1} ensures that all the information contained in linear functions of \mathbf{X}_{t-1} has been fully exploited. If this were not the case, then there is information already available that could improve the performance of the model, that is currently not used. Note that this requirement does not rule out the possibility of specifying the model error to be generated by a moving average process, such as $\eta_t = \epsilon_t + \theta\epsilon_{t-1}$ for which η_t is not an innovation. What is required in such a case is that the white noise error ϵ_t be an innovation with respect to \mathbf{X}_{t-1} .

In each of the examples discussed in section 4 \mathcal{M}_2 is not coherent with the ‘observed sample information’ and so is not congruent. If the feature of each of these \mathcal{M}_2 models that leads to the lack of congruence were removed, then they would each encompass the corresponding \mathcal{M}_1 model. Accordingly it would appear that a general model \mathcal{M}_2 which is congruent will encompass a simple model \mathcal{M}_1 that it nests. This surmise will be proved below as a corollary to the following formal definition of congruence.

The interpretation of parsimonious encompassing presented in section 3.2 immediately suggests the following definition of congruence.

Definition 4 (Congruence). *A theory consistent, data admissible model*

$\mathcal{M}_g = \{f_g(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\xi}), \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^k\}$ is congruent if and only if $\mathcal{M}_g \mathcal{E}_p(\phi)$ LDGP at the point $\phi = \phi_0$ when LDGP = $\{D_x(\mathbf{x}_t|\mathbf{X}_{t-1}, \phi), \phi \in \Phi \subset \mathbb{R}^s\}$, which means that \mathcal{M}_g is a valid reduction of the local DGP. Equivalently \mathcal{M}_g is congruent if and only if $\boldsymbol{\mu}_2(\phi_0) = \mathbf{0}$ when $\boldsymbol{\mu} = \boldsymbol{\mu}(\phi) = (\boldsymbol{\mu}_1(\phi)', \boldsymbol{\mu}_2(\phi)') = (\boldsymbol{\xi}', \boldsymbol{\mu}_2(\phi)')$ is an isomorphic reparameterization of ϕ with $\left(\frac{\partial \boldsymbol{\mu}}{\partial \phi'}\right)_{\phi_0}$ having full rank.

Congruence therefore requires a model \mathcal{M}_g , in addition to being data coherent ($\boldsymbol{\mu}_2(\phi_0) = \mathbf{0}$), to be defined with respect to a probability space that has a density function $f_x(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\xi})$, a set of variables \mathbf{x}_t , and parameterization $\boldsymbol{\xi}$, that are capable of being (low level) theory consistent and data admissible. Though many models that are commonly used in econometrics are linear in variables and parameters, it is conceivable that the unknown LDGPs are nonlinear. Whilst finding a congruent nonlinear model would be invaluable, in its absence linear approximations may capture the main features of the data sufficiently well for there to be no evidence of non-congruence. Indeed such linear econometric models can be highly effective vehicles for economic analysis using available information, despite not being observationally equivalent to the LDGP. Further positive consequences of a model being congruent are considered in the next two sub-sections.

5.1 General Models, Encompassing and Nesting

From the example of Gouriéroux and Monfort (1995) discussed in section 4 it might appear that whether a parametrically more general model will encompass a simplification of itself depends fortuitously on the nature of unknown LDGP. However, this ignores the possibility that models can be designed to ensure that they encompass simple cases of themselves. Although Gouriéroux and Monfort argued that the LDGP can be such that \mathcal{M}_2 does not encompass its sub-model \mathcal{M}_1 , they also provided a zone which defines a class of LDGPs which will enable \mathcal{M}_2 to encompass \mathcal{M}_1 . Since the LDGP is unknown and cannot be designed (except in situations like Monte Carlo simulation experiments), whereas models are artifacts that can be designed to have particular characteristics, it seems natural to consider what design features are required for models to encompass simplifications of themselves. Heuristically a congruent nesting model can be expected to encompass all sub-models of itself. Hence if a general model fails to encompass a simplification of itself it is possible that this has arisen as a result of the general model not being congruent or not nesting the simplification, or both. The examples in section 4 suggest that one model will encompass an algebraic simplification of itself if and only if there are no false auxiliary hypotheses implicit in the nesting model, which are not also an integral part of the ‘nested’ model. When a general model \mathcal{M}_g is data coherent (i.e., $\boldsymbol{\mu}_2(\phi_0) = \mathbf{0}$), it is a valid reduction of the LDGP and so will be able to explain the properties of all models nested within it, as the following lemma records.

Lemma 5. *Since a data coherent model is a valid reduction of the LDGP, and by definition the LDGP encompasses all models defined on the same probability space, a data coherent model will encompass all models nested within it.*

A direct consequence of this lemma is the following corollary.

Corollary 6. *Since a congruent model is data coherent, it too will encompass all models nested within it. (i.e. if $\mu_2(\phi_0) = \mathbf{0}$ then $\mathcal{M}_g \mathcal{E} \mathcal{M}_i \quad \forall \mathcal{M}_i \subseteq \mathcal{M}_g$)*

Hence congruence is a sufficient condition for a model to encompass models nested within it, and so provides a sufficient condition for the absence of the anomaly pointed out by Gouriéroux and Monfort (1995). The fact that it is not necessary is illustrated by the following example. Let the set of variables agreed to be relevant for a particular problem be (y_t, x_t, z_t, u_t) , and for the parameters of interest let (x_t, z_t, u_t) be weakly exogenous variables,¹ so that the LDGP lies in $\mathbf{D}(y_t \mid x_t, z_t, u_t)$. Consider the models:

$$\begin{aligned} \mathcal{M}_1 & : y_t = \beta x_t + \zeta_{1t} & \zeta_{1t} & \sim \mathcal{N}(0, \sigma_{\zeta_1}^2) \\ \mathcal{M}_2 & : y_t = \gamma x_t + \delta z_t + \varepsilon_{1t} & \varepsilon_{1t} & \sim \mathcal{N}(0, \sigma_{\varepsilon_1}^2) \end{aligned}$$

and note that in general \mathcal{M}_2 is not congruent since it wrongly excludes u_t , and so is not a valid reduction of the LDGP. However, if it were the case that u_t were orthogonal to both x_t and z_t , then \mathcal{M}_2 would encompass \mathcal{M}_1 despite \mathcal{M}_2 not being congruent.

Further note that since congruence is a sufficient condition for the absence of the anomaly it will be possible for a non-congruent nesting model to encompass some, though not all, nested models. In fact, given that congruence cannot be directly established (see the discussion of testing congruence below) it is fortunate that congruence is a sufficient, but not necessary, condition for $\mathcal{M}_g \mathcal{E} \mathcal{M}_s$ when $\mathcal{M}_s \subseteq \mathcal{M}_g$. Equally, although congruence and nesting ensure that the nesting model encompasses the nested model, nesting alone does not ensure encompassing.

5.2 General-to-specific Modelling

In addition to providing a sufficient condition for the anomaly noted by Gouriéroux and Monfort (1995) not to occur, congruence provides justification for the modelling strategy which seeks to evaluate alternative models of the phenomenon of interest in the framework of a congruent nesting model - see Hendry (1995), Mizon (1977b), Mizon (1977a) and Mizon (1995a). The above definitions, lemma, and corollary indicate that a good strategy to avoid misleading inferences is to test hypotheses such as $\mathcal{M}_s \mathcal{E}_p(\boldsymbol{\theta}) \mathcal{M}_g$ only when \mathcal{M}_g is congruent. Equally, they make it clear that the equivalence $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2 \iff \mathcal{M}_1 \mathcal{E}_p \mathcal{M}_c$ requires \mathcal{M}_c to be congruent, or at least data coherent (cf. Hendry and Richard (1989) and Hendry (1995)). Also note that since $\boldsymbol{\mu}(\phi)$ is an isomorphic reparameterization of ϕ it follows that $\boldsymbol{\xi}$ and $\boldsymbol{\mu}_2$ are variation free, which is a convenient property for a model like \mathcal{M}_g to have given that the LDGP is in general unknown. However, despite $\boldsymbol{\xi}$ and $\boldsymbol{\mu}_2$ being variation free, fully efficient inference on $\boldsymbol{\xi}$ can only be obtained from \mathcal{M}_g if it is congruent, that is when $\boldsymbol{\mu}_2 = \mathbf{0}$. It is therefore important to test models for congruence.

6 Testing Congruence

Unfortunately, the definition of congruence is non-operational since the LDGP (and hence its parameterization) are unknown, so that a test of the hypothesis $\boldsymbol{\mu}_2(\phi_0) = \mathbf{0}$ is not feasible. To

¹This is not an essential feature of the example, but simplifies the presentation.

the extent that part of μ_2 is known it would be possible to conduct a partial test of congruence, and this would be in the spirit of misspecification testing. Indeed, the fact that only a subset of μ_2 might be known, and hence part of the condition for congruence testable, reflects the fact that at best it is only possible to test necessary conditions for congruence. A commonly adopted strategy for testing congruence is to test the adequacy of a model against misspecification in particular directions, such as having residuals that are serially correlated and heteroscedastic, there being invalid conditioning, and parameter non-constancies. Indeed this is the basis of the definitions of congruence given in, for example, Hendry (1995) and Mizon (1995a). In the present context misspecification testing applied to \mathcal{M}_g in order to assess its congruence can be characterized as testing the m hypotheses $\rho_i = 0$ for $\mathcal{M}_g \mathcal{E}_p \mathcal{M}_g^i$ when the augmented models are given by $\mathcal{M}_g^i = \{f_g^i(\mathbf{x}_t | \mathbf{X}_{t-1}; \xi, \rho_i), \xi \in \Xi, \rho_i \in \mathbb{R}\}$, with $f_g^i(\mathbf{x}_t | \mathbf{X}_{t-1}; \xi, 0) = f_g(\mathbf{x}_t | \mathbf{X}_{t-1}; \xi) \quad \forall i$ and the ρ_i 's might be residual serial correlation coefficients, parameter changes, or heteroscedasticity parameters. Note that in cases were $\mu_2(\phi_0) \neq \mathbf{0}$ the non-congruence of \mathcal{M}_g can exhibit itself in many different ρ_i 's being non-zero. For example, a shift in a long run equilibrium of a system often results in serially correlated residuals. This illustrates the well known fact that it is often (usually) inappropriate when an hypothesis $\rho_i = 0$ (such as zero residual serial correlation) is rejected, to modify the tested model \mathcal{M}_g in the direction of the alternative hypothesis for the misspecification test (e.g. by introducing a serially correlated error process). See inter alia Mizon (1995b) for further discussion of this point. In the present context notice also that for a test of misspecification not to yield potentially misleading information it is necessary that the completing model \mathcal{M}_g^i encompass \mathcal{M}_g . This accords with the quotation from Hendry (1995) given in section 4, but note that $\mathcal{M}_g \subseteq \mathcal{M}_g^i$ does not ensure that $\mathcal{M}_g^i \mathcal{E} \mathcal{M}_g$. To the extent that the misspecification hypotheses $\rho_i = 0$ for $i = 1, 2, \dots, m$ are not orthogonal to each other this poses a problem. Were each \mathcal{M}_g^i congruent the problem would disappear, but fortunately congruence of \mathcal{M}_g^i is not essential for misspecification testing of \mathcal{M}_g provided that the investigator does not adopt \mathcal{M}_g^i as the preferred model if $\rho_i = 0$ is rejected.

Since the LDGP is not known thus implying that $\mu_2(\phi)$ is unknowable, whereas ξ can be estimated, the following restricted definition of parsimonious encompassing might be considered as the basis for an alternative way to test congruence.

Definition 7 (Restricted Parsimonious Encompassing). *Model \mathcal{M}_s parsimoniously encompasses another model \mathcal{M}_g with respect to α , $(\mathcal{M}_s \mathcal{E}_p(\alpha) \mathcal{M}_g)$ if and only if :*

- $\mathcal{M}_s \subseteq \mathcal{M}_g$
- $\mathcal{M}_s \mathcal{E}(\alpha) \mathcal{M}_g$

Note that the encompassing comparison is made with respect to α , the parameter of the nested model \mathcal{M}_s , rather than ξ of the nesting model \mathcal{M}_g . However, the implicit null hypothesis (see Mizon and Richard, 1986) associated with this definition has the same basic form as that of a Hausman specification test statistic Hausman (1978), namely the hypothesis that the contrast between the pseudo true value of α in \mathcal{M}_s and that of α in \mathcal{M}_g (when ξ is reparameterized as $\xi' = (\alpha', \psi')$) is zero. Hence this hypothesis does not define a valid reduction from \mathcal{M}_g . Indeed, as shown by Holly (1982), the implicit null hypothesis holds if either (i) $\psi(\xi) = \mathbf{0}$ or (ii) $\text{plim}_{T \rightarrow \infty} \left(T^{-1} \frac{\partial^2 L_g}{\partial \alpha \partial \psi'} \right) = 0$. The fact that (ii) does not ensure that \mathcal{M}_s contains all information relevant for inference on α highlights the limitation of this approach. This is there-

fore not a promising route for testing for valid reductions from a fully specified alternative (e.g. $\mathcal{M}_s \mathcal{E}_p(\boldsymbol{\xi}) \mathcal{M}_g$), and thus not for specification testing. On the other hand when \mathcal{M}_g is not a serious alternative to \mathcal{M}_s but simply a vehicle for testing the adequacy of \mathcal{M}_s , then testing the hypothesis $\mathcal{M}_s \mathcal{E}_p(\boldsymbol{\alpha}) \mathcal{M}_g$ can be useful as a misspecification test. Hence the role of such restricted parsimonious encompassing hypotheses seems to lie in misspecification testing rather than specification testing (see Mizon, 1977a for discussion of this distinction). Thus there is a potential role for testing restricted parsimonious encompassing hypotheses of the type $\mathcal{M}_g \mathcal{E}_p(\boldsymbol{\xi}) \mathcal{M}_g^i$ when \mathcal{M}_g^i is not a serious alternative to \mathcal{M}_g but simply a vehicle for indirectly testing the congruence of \mathcal{M}_g .

7 Empirical Illustration

As a means of illustrating many of the concepts discussed in this chapter this section contains analysis of data generated from an artificial LDGP. Using artificial data, rather than observed macro time series data, has the advantage of allowing the LDGP to be designed precisely to illustrate the chosen concepts. In particular, the LDGP is known, as is the equivalence set of representations for it, and thus for this special circumstance they provide clear benchmarks against which empirical models can be compared. For the particular sample of data generated, the alternative representations of the LDGP which are known to be observationally equivalent can be estimated. In addition, there are other empirical models, which despite not being observationally equivalent to the LDGP, display no evidence of misspecification, and so cannot be rejected as congruent models.

Consider a situation in which the variable to be modelled y_t , and an unrelated (and thus non-causal) variable z_t , are generated by:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} + \theta \epsilon_{1,t-1} \\ \epsilon_{2,t} \end{pmatrix} \quad (7)$$

when $(\epsilon_{1,t}, \epsilon_{2,t})' = \boldsymbol{\epsilon}_t \sim \text{NI}(\mathbf{0}, \mathbf{I}_2)$. Hence y_t is generated from an $ARMA(1, 1)$, and z_t from an independent white noise process. An observationally equivalent representation of this LDGP is given with the first equation of (7) replaced by:

$$y_t = \sum_{i=0}^{\infty} \phi_i y_{t-1-i} + \epsilon_{1,t}. \quad (8)$$

with $\phi_i = (-\theta)^i [\alpha + \theta]$. Note that $\phi_i \rightarrow 0$ as $i \rightarrow \infty$ for $|\alpha| < 1$ and $|\theta| < 1$ so that a finite order autoregression will give a good approximation to (8) under these conditions. Further, although this particular LDGP has two observationally equivalent representations, the one involving the moving average error format in (7) might be preferred on grounds of parsimony **if the LDGP were known**. In fact, this $ARMA(1, 1)$ format was used to generate the sample of $T = 100$ observations on y_t and z_t (with $\alpha = 0.9$ and $\theta = 0.5$), rather than the less tractable $AR(\infty)$ representation in (8). However, in practice the LDGP is not known and the best that can be done is to compare the relative merits of alternative empirical models.

The following analysis illustrates how it is possible to find more than one empirical model for which in the sample there is no evidence of misspecification so that the hypothesis of congruence is not rejected, even though all but one of the models differ in the population from the LDGP. Hence, although in the population a congruent model is observationally equivalent to the LDGP, a model for which the hypothesis of congruence has not been rejected need not be equivalent to the LDGP. However, an important feature of models that are empirically congruent is that they are indistinguishable from the LDGP on the basis of sample information. Further empirically congruent models have properties in the sample similar to those of the LDGP in the population, namely: (i) they will encompass models nested within them; (ii) they provide a valid basis against which to test reduction or simplification hypotheses; and (iii) they are able to successfully predict the properties of other models (see Hendry, 1995 for a discussion of misspecification encompassing).

Consider the following classes of empirical model:

$$\begin{aligned}\mathcal{M}_1 : y_t &= \delta_1 + \beta_1 y_{t-1} + u_{1,t} + \gamma_1 u_{1,t-1} \\ \mathcal{M}_2 : y_t &= \delta_2 + \sum_{i=1}^4 \beta_{2,i} y_{t-i} + \sum_{i=0}^4 \lambda_{2,i} z_{t-i} + u_{2,t} \\ \mathcal{M}_3 : y_t &= \delta_3 + \sum_{i=1}^5 \beta_{3,i} y_{t-i} + u_{3,t}\end{aligned}\tag{9}$$

\mathcal{M}_1 is an $ARMA(1, 1)$ and includes the LDGP at $\delta_1 = 0$, $\beta_1 = 0.9$ and $\gamma_1 = 0.5$, plus $E(u_{1,t}) = 0$, $\text{Cov}(u_{1,t}, u_{1,s}) = 0$ $t \neq s$, and $V(u_{1,t}) = 1$. \mathcal{M}_2 is an autoregressive-distributed lag model $AD(4, 4)$, and although it does not include the LDGP so it is not congruent in the population, the hypothesis of congruence is not rejected in sample. \mathcal{M}_3 is a fifth order autoregression $AR(5)$, which neither includes nor is equivalent to the LDGP, but is a finite order truncation of (8). As a preliminary to considering simplifications of them, these three models are now estimated and diagnostic statistics used to assess the evidence for their misspecification.

First, the LDGP, \mathcal{M}_1 , was estimated and the particular sample evidence for misspecification assessed. The parameter point estimates and standard errors and the residuals $\hat{u}_{1,t}$ were calculated using the exact maximum likelihood estimation option for linear regression models with moving average errors in *Microfit* (see Pesaran and Pesaran, 1997). The diagnostic statistics were calculated in *PcGive* (see Doornik and Hendry, 1996), which was used together with *PcGets* (see Krolzig and Hendry, 2000) for the other results reported in this section. The diagnostic statistics reported are: single equation residual standard deviations $\hat{\sigma}$, $F_{ar}(p, \cdot)$ a Lagrange multiplier test of p^{th} order residual serial correlation, F_{het} a test of residual heteroscedasticity, $F_{arch}(q, \cdot)$ a test of q^{th} order residual autoregressive heteroscedasticity, $\chi_{norm}^2(2)$ a test of residual normality; AIC the Akaike information criterion, and SC the Schwarz information criterion (see Hendry and Doornik, 1996 for more details).

$$y_t = -\underset{(0.164)}{0.1522} + \underset{(0.041)}{0.9188} y_{t-1} + \hat{u}_{1,t} + \underset{(0.093)}{0.4483} \hat{u}_{1,t-1}$$

$$R^2 = 0.919, s = 1.0357, T = 99, AIC = 0.100, SC = 0.179\tag{10}$$

$$F_{ar}(5, 93) = 0.24[0.94], F_{arch}(4, 90) = 0.70[0.59]$$

$$F_{het}(2, 94) = 0.29[0.75], \chi_{norm}^2(2) = 0.21[0.90]$$

The point estimates of δ , α and θ are close to, with none of them significantly different from, their population values, and there is no evidence of misspecification in the model. Hence this

particular sample of data is capable of providing accurate estimates of the LDGP, and does not give any indication of non-congruence via the reported diagnostic statistics.

The estimates for \mathcal{M}_2 are:

$$\begin{aligned} \hat{y}_t = & -0.075 + 1.40 y_{t-1} - 0.59 y_{t-2} + 0.10 y_{t-3} + 0.03 y_{t-4} \\ & \quad \quad \quad (0.12) \quad (0.11) \quad (0.19) \quad (0.19) \quad (0.11) \\ & + 0.06 z_t - 0.16 z_{t-1} + 0.01 z_{t-2} + 0.002 z_{t-3} + 0.04 z_{t-4} \\ & \quad \quad \quad (0.11) \quad (0.11) \quad (0.11) \quad (0.11) \quad (0.11) \end{aligned} \quad (11)$$

$$\begin{aligned} R^2 = 0.924, \quad s = 1.0313, \quad T = 96, \quad AIC = 0.160, \quad SC = 0.427 \\ F_{ar}(5, 81) = 0.30 [0.91], \quad F_{arch}(4, 78) = 0.92 [0.46] \\ F_{het}(18, 67) = 0.88 [0.61], \quad \chi_{norm}^2(2) = 0.17 [0.92] \end{aligned}$$

Again there is no evidence of misspecification in this estimated model, so \mathcal{M}_2 also provides a valid basis from which to test the validity of reductions. Note in particular that \mathcal{M}_2 includes the irrelevant non-causal variables $z_t, z_{t-1}, \dots, z_{t-4}$, none of whose estimated coefficients is significantly different from zero. In addition, the estimated coefficients of y_{t-1} and y_{t-2} are close to $\phi_1 = 1.4$ and $\phi_2 = -0.7$, whilst those of y_{t-3} and y_{t-4} corresponding to $\phi_3 = 0.35$ and $\phi_4 = -0.175$ are not significantly different from these values or zero. The fact that none of the diagnostic statistics indicates non-congruence implies that the fourth order autoregression within \mathcal{M}_2 provides a good approximation to LDGP.

The results from estimating \mathcal{M}_3 also yield a model with no evidence of misspecification, and suggest that an $AR(2)$ provides a good representation of these sample data.

$$\begin{aligned} \hat{y}_t = & -0.08 + 1.39 y_{t-1} - 0.61 y_{t-2} \\ & \quad \quad \quad (0.12) \quad (0.11) \quad (0.19) \\ & + 0.14 y_{t-3} - 0.01 y_{t-4} + 0.02 y_{t-5} \\ & \quad \quad \quad (0.20) \quad (0.19) \quad (0.11) \end{aligned} \quad (12)$$

$$\begin{aligned} R^2 = 0.920, \quad s = 1.0278, \quad T = 95, \quad AIC = 0.116, \quad SC = 0.277 \\ F_{ar}(5, 84) = 1.53 [0.19], \quad F_{arch}(4, 81) = 1.102 [0.36] \\ F_{het}(10, 78) = 0.82 [0.61], \quad \chi_{norm}^2(2) = 0.29 [0.86] \end{aligned}$$

On the basis of results so far it is concluded that:

- (a) estimating 'general' models involving the available data for y_t and z_t yields three models for which the hypothesis of congruence is not rejected, only one of which is equivalent to the LDGP;
- (b) the irrelevant non-causal variables $z_t, z_{t-1}, \dots, z_{t-4}$ are not falsely indicated to be relevant;
- (c) the informal general-to-specific modelling within the autoregressive-distributed lag class of model reported above works well;
- (d) estimating \mathcal{M}_1 which includes the LDGP works best in terms of point estimates and goodness of fit adjusted for loss of degrees of freedom in higher dimensional parameterizations - see the AIC and SC values.

Although the use of information criteria to choose amongst congruent models as referred to in (d) leads to the choice of \mathcal{M}_1 , it is clear that both \mathcal{M}_2 and \mathcal{M}_3 can be simplified and so

further evaluation of their merits is in order. Note that \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 are non-nested models and so strictly are only alternatives to each other in the context of an embedding model, such as the minimum completing model:

$$\mathcal{M}_4 : y_t = \delta_4 + \sum_{i=1}^5 \beta_{4,i} y_{t-i} + \sum_{i=0}^4 \lambda_{4,i} z_{t-i} + u_{4,t} + \gamma_4 u_{4,t-1} \quad (13)$$

which is an $AD(5, 4)MA(1)$. When estimated using *Microfit*, this model too exhibited no evidence of misspecification as the following results show.

$$\begin{aligned} y_t = & -\frac{0.12}{(0.19)} + \frac{0.77}{(0.19)} y_{t-1} + \frac{0.31}{(0.23)} y_{t-2} - \frac{0.33}{(0.16)} y_{t-3} + \frac{0.13}{(0.13)} y_{t-4} + \frac{0.02}{(0.11)} y_{t-5} \\ & + \frac{0.06}{(0.11)} z_t - \frac{0.13}{(0.13)} z_{t-1} - \frac{0.06}{(0.13)} z_{t-2} + \frac{0.03}{(0.13)} z_{t-3} + \frac{0.03}{(0.10)} z_{t-4} + \hat{u}_{4,t} + \frac{0.65}{(0.17)} \hat{u}_{4,t-1} \\ R^2 = & 0.924, s = 1.0362, T = 95, AIC = 0.188, SC = 0.511 \\ & F_{ar}(5, 78) = 0.99 [0.43], F_{arch}(4, 75) = 1.33 [0.27] \\ & F_{het}(20, 62) = 0.73 [0.78], \chi_{norm}^2(2) = 0.16 [0.93] \end{aligned} \quad (14)$$

It is important to note that typically models with moving average errors will only be considered if a general model with such an error is specified. In fact, regarding \mathcal{M}_2 and \mathcal{M}_3 as alternative general and unrestricted models (GUMs), it is seen that it is possible to get GUMs which exhibit no evidence of non-congruence without specifying moving average errors. However, for these sample data once the $MA(1)$ error specification is entertained the estimated moving average coefficient remains significant in all simplifications that have no evidence of misspecification. Indeed, applying a simplification strategy of sequentially deleting the least significant variable until all remaining variables are significant, results in \mathcal{M}_1 being selected. If the $MA(1)$ error specification is not considered then the minimum completing model for \mathcal{M}_2 and \mathcal{M}_3 is an $AD(5, 4)$, which is (13) with $\gamma_4 = 0$. Denoting this model as \mathcal{M}_5 and estimating it yields the following results, which also reveal no evidence of misspecification:

$$\begin{aligned} \hat{y}_t = & -\frac{0.08}{(0.13)} + \frac{1.40}{(0.11)} y_{t-1} - \frac{0.58}{(0.20)} y_{t-2} + \frac{0.07}{(0.21)} y_{t-3} + \frac{0.06}{(0.20)} y_{t-4} - \frac{0.01}{(0.11)} y_{t-5} \\ & + \frac{0.06}{(0.11)} z_t - \frac{0.16}{(0.11)} z_{t-1} + \frac{0.02}{(0.11)} z_{t-2} + \frac{0.01}{(0.11)} z_{t-3} + \frac{0.03}{(0.11)} z_{t-4} \end{aligned} \quad (15)$$

$$\begin{aligned} R^2 = & 0.922, s = 1.0416, T = 95, AIC = 0.190, SC = 0.486 \\ & F_{ar}(5, 79) = 1.22 [0.31], F_{arch}(4, 76) = 0.91 [0.47] \\ & F_{het}(65, 18) = 0.75 [0.76], \chi_{norm}^2(2) = 0.21 [0.90] \end{aligned}$$

For models in the autoregressive-distributed class a computer-automated general-to-specific modelling strategy has been implemented in *PcGets* (see Krolzig and Hendry, 2000), and this was applied to \mathcal{M}_5 . The computer program *PcGets* first tests a GUM for congruence, then conducts variable elimination tests for ‘highly irrelevant’ variables at a loose significance level (25% or 50%, say) in order to simplify the GUM. The program then explores the many paths that can be used to simplify the GUM by eliminating statistically-insignificant variables on F- and t-tests, and applying diagnostic tests to ensure that only valid reductions are entertained.

This ensures that the preferred model in each path is congruent. Encompassing procedures and information criteria are then used to select a model from this set of congruent models. As a final check of model adequacy, the constancy of the chosen model across sub-samples is tested. Using \mathcal{M}_5 as a congruent GUM, and applying *PcGets* finds the following $AR(2)$ (denoted by \mathcal{M}_6 below) as a unique valid reduction:

$$\hat{y}_t = -0.11 + 1.34 y_{t-1} - 0.41 y_{t-2}$$

(0.12) (0.10) (0.10)

$$R^2 = 0.918, s = 1.0230, T = 95, AIC = 0.077, SC = 0.157 \quad (16)$$

$$F_{ar}(5, 87) = 0.76 [0.58], F_{arch}(4, 84) = 1.31 [0.27]$$

$$F_{het}(4, 87) = 0.23 [0.92], \chi_{norm}^2(2) = 0.67 [0.72]$$

Hence the application of *PcGets* to these data selects \mathcal{M}_6 as a congruent reduction of \mathcal{M}_5 . This is in conformity with the results in Hendry and Trivedi (1972) and Hendry (1977) which note that an $MA(p)$ is well approximated in terms of goodness of fit by an $AR(p)$, so that an $ARMA(p, q)$ will be well approximated by an $AR(p + q)$. In the case of \mathcal{M}_1 $p = q = 1$ so that the $AR(2)$ in \mathcal{M}_6 is expected to provide a good approximation to \mathcal{M}_1 .

Given that $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_6$ are all empirically congruent it is interesting to compare their goodness of fit and the values of the *AIC* and *SC* information criteria. From inspection of Table 1 the best fit, as judged by the residual sum of squares (*RSS*), is obtained as to be expected by the most profligate parameterization, namely \mathcal{M}_4 . The other over-parameterized models \mathcal{M}_5 , \mathcal{M}_2 and \mathcal{M}_3 also perform well in goodness of fit. However, these four models are strongly dominated in terms of information criteria by \mathcal{M}_1 and \mathcal{M}_6 , with \mathcal{M}_1 being preferred to \mathcal{M}_6 despite each model have three parameters plus the residual variance to be estimated.

	Table 1			
	<i>AIC</i>	<i>SC</i>	<i>RSS</i>	<i>Rank</i>
\mathcal{M}_1	0.04266	0.09643	95.0531	1
\mathcal{M}_2	0.159973	0.42709	91.46786	4
\mathcal{M}_3	0.115911	0.27721	94.01662	3
\mathcal{M}_4	0.18867	0.51127	89.1141	5/6/6
\mathcal{M}_5	0.19012	0.48583	91.14155	6/5/5
\mathcal{M}_6	0.05630	0.10973	97.41528	2

The results for these estimated models can also be used to illustrate that a congruent model will be able to predict the properties of other models, particularly those nested within it. For example, \mathcal{M}_6 (an $AR(2)$) which is empirically congruent and a valid reduction of \mathcal{M}_5 , predicts that the errors of the following parsimonious models $AR(1)$, $AD(1, 1)$, $AD(1, 0)$ and $AD(0, 0)$

will all be serially correlated. This is indeed the case as the results in the Table 2 show:

	$F_{ar}(5, \cdot)$	$F_{arch}(4, \cdot)$	$\chi_{norm}^2(2)$	$F_{het}(\cdot, \cdot)$	RSS	SC
$AD(1, 1)$	3.58 [0.01]**	0.24 [0.92]	0.05 [0.98]	0.89 [0.51]	119.717	0.376
$AD(1, 0)$	3.68 [0.00]**	0.44 [0.78]	0.09 [0.96]	0.80 [0.53]	120.514	0.336
$AR(1)$	3.55 [0.01]**	0.43 [0.78]	0.02 [0.99]	0.08 [0.92]	121.599	0.298
$AD(0, 0)$	198.0 [0.00]**	73.53 [0.00]**	4.80 [0.09]	1.54 [0.22]	1286.19	2.646

Finally, note that using the SC to select amongst the models in Table 2 would result in the $AR(1)$ model being selected, despite its non-congruence. Indeed, using the SC criterion results in the $AR(2)$ model being preferred to all models other than the \mathcal{M}_1 , \mathcal{M}_6 and \mathcal{M}_3 . Hence selecting models solely on the basis of information criteria such as the SC , and paying no attention to the congruence of the alternative models, can result in a model being selected that is parsimoniously parameterized but misspecified. Noting that inferences based on misspecified models can be very misleading, are not exploiting relevant information that is already available, illustrates the value of seeking congruent and encompassing models and of adopting a general-to-specific modelling strategy.

8 Conclusions

In this chapter the relationship between alternative econometric models and the LDGP has been considered. In particular, an econometric model has been defined to be congruent if and only if it parsimoniously encompasses the LDGP. In the population this implies that a congruent model is observationally equivalent to the LDGP, and that they mutually encompass each other. Hence by definition a congruent model contains the same information content as the LDGP. The fact that alternative parameterizations for a given density can exist means that the LDGP can be a member of an equivalence set of models. The principle of, *ceteris paribus*, preferring parsimonious to profligate parameterizations can be applied to select the simplest representation of the LDGP (e.g., preferring the $ARMA(1, 1)$ representation over the $AR(\infty)$ in section 7). However, congruence so defined is not directly testable in practice, and as a result is tested indirectly via tests for evidence of misspecification. Consequently, when the hypothesis of congruence is not rejected using sample data, this means that the present use that has been made of the available information has been unable to distinguish between the model and the LDGP. In section 7 this was illustrated by the $AR(2)$ model providing an excellent approximation to the LDGP - an $ARMA(1, 1)$. Another feature of empirically congruent models is that they mimic the properties of the LDGP: they can accurately predict the misspecifications of non-congruent models; they can encompass models nested within them; and they provide a valid statistical framework for testing alternative simplifications of them. The fact that there can be more than one empirically congruent model means that it is important to evaluate their relative merits, and this can be done via encompassing comparisons. A very powerful way to find congruent and encompassing models in practice is to use a general-to-specific modelling strategy, beginning from a

congruent GUM. Finally, the arguments in this chapter apply equally to econometric systems and single equation models. The empirical analysis in section 7 only considers single equation models to keep the illustration simple.

References

- Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.
- Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 105–123 Berkeley: University of California Press.
- Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, **49**, 781–793.
- Doornik, J. A., & Hendry, D. F. (1996). *GiveWin: An Interactive Empirical Modelling Program*. London: Timberlake Consultants Press.
- Gouriéroux, C., & Monfort, A. (1995). Testing, encompassing, and simulating dynamic econometric models. *Econometric Theory*, **11**, 195–228.
- Govaerts, B., Hendry, D. F., & Richard, J.-F. (1994). Encompassing in stationary linear dynamic models. *Journal of Econometrics*, **63**, 245–270.
- Harvey, A. C. (1993). *Time Series Models*, 2nd edn. Hemel Hempstead: Harvester Wheatsheaf.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, **46**, 1251–1271.
- Hendry, D. F. (1977). On the time series approach to econometric model building. In Sims, C. A. (ed.), *New Methods in Business Cycle Research*, pp. 183–202. Minneapolis: Federal Reserve Bank of Minneapolis. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.
- Hendry, D. F. (1985). Monetary economic myth and econometric reality. *Oxford Review of Economic Policy*, **1**, 72–84. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.
- Hendry, D. F. (1987). Econometric methodology: A personal perspective. In Bewley, T. F. (ed.), *Advances in Econometrics*, Ch. 10. Cambridge: Cambridge University Press.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F., & Doornik, J. A. (1996). *Empirical Econometric Modelling using PcGive for Windows*. London: Timberlake Consultants Press.
- Hendry, D. F., & Mizon, G. E. (1990). Procrustean econometrics: or stretching and squeezing data. In Granger, C. W. J. (ed.), *Modelling Economic Series*, pp. 121–136. Oxford: Clarendon Press.
- Hendry, D. F., & Mizon, G. E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics*, **23**, 267–294.

- Hendry, D. F., & Mizon, G. E. (2000). On selecting policy analysis models by forecast accuracy. In Atkinson, A. B., Glennerster, H., & Stern, N. (eds.), *Putting Economics to Work: Volume in Honour of Michio Morishima*, pp. 71–113. London School of Economics: STICERD.
- Hendry, D. F., & Richard, J.-F. (1989). Recent developments in the theory of encompassing. In Cornet, B., & Tulkens, H. (eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press.
- Hendry, D. F., & Trivedi, P. K. (1972). Maximum likelihood estimation of difference equations with moving-average errors: A simulation study. *Review of Economic Studies*, **32**, 117–145.
- Holly, A. (1982). A remark on hausman's specification test. *Econometrica*, **50**, 749–760.
- Krolzig, H.-M., & Hendry, D. F. (2000). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*. forthcoming.
- Lu, M., & Mizon, G. E. (2000). Nested models, orthogonal projection, and encompassing. In Marriott, P., & Salmon, M. (eds.), *Applications of Differential Geometry to Econometrics*, pp. 64–84. Cambridge: Cambridge University Press. ISBN 0 521 651166.
- Mizon, G. E. (1977a). Inferential procedures in nonlinear models: An application in a UK industrial cross section study of factor substitution and returns to scale. *Econometrica*, **45**, 1221–1242.
- Mizon, G. E. (1977b). Model selection procedures. In Artis, M. J., & Nobay, A. R. (eds.), *Studies in Modern Economic Analysis*, pp. 97–120. Oxford: Basil Blackwell.
- Mizon, G. E. (1984). The encompassing approach in econometrics. In Hendry, D. F., & Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, pp. 135–172. Oxford: Basil Blackwell.
- Mizon, G. E. (1993). Empirical analysis of time series: Illustrations with simulated data. In de Zeeuw, A. (ed.), *Advanced Lectures in Quantitative Economics*, Vol. II, pp. 184–205. New York.: Academic Press.
- Mizon, G. E. (1995a). Progressive modelling of macroeconomic time series: the LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, pp. 107–169. Dordrecht: Kluwer Academic Press.
- Mizon, G. E. (1995b). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, **69**, 267–288.
- Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.
- Pesaran, M. H., & Pesaran, B. (1997). *Microfit for Windows. Version 4*. Oxford: Oxford University Press.
- Sargan, J. D. (1988). *Lectures on Advanced Econometric Theory*. Oxford: Basil Blackwell.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–26.