Model Identification and Non-unique Structure

David F. Hendry
Economics Department, Oxford University, UK

Maozu Lu and Grayham E. Mizon
Economics Department, Southampton University, UK. *

May 2001

Abstract

Identification is an essential attribute of any model's parameters, so we consider its three aspects of 'uniqueness', 'correspondence to reality' and 'interpretability'. Observationally-equivalent over-identified models can co-exist, and are mutually encompassing in the population; correctly-identified models need not correspond to the underlying structure; and may be wrongly interpreted. That a given model is over-identified with all over-identifying restrictions valid (even asymptotically) is insufficient to demonstrate that it is a unique representation. Moreover, structure (as invariance under extended information) need not be identifiable. We consider the role of structural breaks to discriminate between such representations.

1 Introduction

Of the many areas of econometrics to which Manfred Deistler has made important contributions in his distinguished career, *identification* is the one on which we have chosen to concentrate. The literature on the topic is vast, and it may be thought to be definitive: important contributions include Wright (1915), Working (1927), Frisch (1934, 1938), Marschak (1942), Haavelmo (1944), Koopmans (1949), Koopmans and Reiersøl (1950), Fisher (1966), Rothenberg (1971), Bowden (1973), Hatanaka (1975) and Hsiao (1983), with the history documented by Qin (1989) and Aldrich (1994) – also see the critical view in Liu (1960) (echoed by Sims, 1980) and the response by Fisher (1961). An equally large body of work has addressed dynamic systems rather than just simultaneous equations models, and this has been the concern of much of Manfred's work – see, *inter alia*, Deistler (1976), Deistler and Seifert (1978), Deistler, Ploberger and Pötscher (1982) and Hannan and Deistler (1988). Certainly, in terms of technical results, or generalizations thereof, we have no new findings to add. However, we suspect that some interpretations of the available results on identification are less well based than might be believed (see e.g., Faust and Whiteman, 1997), so we seek to clarify what can, and cannot, be deduced from finding that a given model is 'uniquely identified'.

Identification has many meanings: e.g., in the time-series literature, such as Box and Jenkins (1976) and Kalman (1982), it means 'match the model to the evidence' (i.e., discover a representation accurate up to a white-noise error). In the econometric literature, following the pioneering work of the Cowles Commission researchers, it relates to the uniqueness of the parameterization that generated the observed

*Prepared for: *Contributions to Econometrics, Time Series Analysis, and System Identification: A Festschrift in Honour of Manfred Deistler*, *Journal of Econometrics*, edited by Benedikt M. Pötscher and Ingmar R. Prucha. Financial support from the UK Economic and Social Research Council under grant L138251009 is gratefully acknowledged.

data.[1] However, we adopt the approach in Hendry (1995a) who follows the notions first discerned by Wright (1915), and so consider three aspects of identification (see Hendry and Morgan, 1995, p23): 'uniqueness', 'correspondence to the desired entity' and 'satisfying the assumed interpretation (usually of a theory model)'. As an analogy, a regression of quantity on price delivers a unique function of the data second moments, but need not correspond to any underlying economic behavior, and may be incorrectly interpreted as a supply schedule due to a positive sign on price. In practice, the meaning of 'identified' can be ambiguous as in 'Have you identified the parameters of the money demand function?'. The first sense of identification was used by the Cowles Commission (Koopmans, Rubin and Leipnik, 1950) who formalized conditions for the uniqueness of coefficients in simultaneous systems, and this is often the sense intended in econometrics. Conditions for the correct interpretation of parameters in the light of a theory model are not so easily specified in general because they depend on subject-matter considerations. The correspondences between parameters of models and those of the underlying data generation processes (DGPs) are also often hidden, but merit careful appraisal.

Thus, we consider each of these three attributes, and discuss those issues which we do not find fully clear in most presentations. Specifically, we show that uniqueness (as determined by the rank condition, say) holds only within specifications, and that several distinct yet valid over-identified representations can co-exist, each satisfying its own rank condition. Thus, the famous Cowles' Commission rank condition uniquely specifies a model only subject to the given restrictions, and does not preclude other distinct, but conflicting, over-identified models.[2] Secondly, we consider 'correspondence to the desired entity' in a non-stationary world, where models that do not correspond can be eliminated, thereby facilitating unique identification. We also address the identification of 'structure'. Thirdly, we briefly discuss failures of interpretation.

The chapter is organized as follows. Section 2 discusses the concepts of identification and observational equivalence for the DGP and models thereof, and illustrates that a model may be uniquely identified but not correspond to reality, or be interpretable. Section 3 considers observational equivalence and mutual encompassing, and illustrates models being indistinguishable in a sample due to weak evidence. Sections 4 and 5 describe in turn non-unique just- and over- identified representations, relating the former to multivariate cointegration analysis and illustrating the latter by four distinct over-identified simultaneous-equations models that are nevertheless fully consistent with the reduced form. Section 6 investigates the next attribute of identification, namely correspondence to reality, and notes that structure might be inherently unidentifiable in a stationary world. However, section 7 argues that non-stationarities, specifically structural breaks, can help discriminate non-structural from structural representations. Finally, section 8 concludes.

2 Identification

The concepts of identification and observational equivalence apply separately to the DGP and to models thereof: the parameters of the DGP could be identified when those of a model were non-unique; or conversely, the model may have a unique parameterization but the parameters of interest from the DGP may be unobtainable.

Let $\mathbf{x}_t$ be the vector of n variables to be modelled, chosen on the basis of economic considerations related to the phenomena of interest and their statistical properties. From the theory of reduc-

[1] Such problems also arise in the time-series literature in relation to ARMA models, where 'redundant' dynamic common factors can occur.

[2] In an earlier discussion related to our approach, Preston (1978) distinguishes between identification of structures and models.

tion (see, *inter alia*, Hendry, 1995a, and Mizon, 1995), there exists a local DGP (LDGP) for these chosen variables $\mathbf{x}_t$ conditional on their history $\mathbf{X}_{t-1} = (\mathbf{X}_0, \mathbf{X}_{t-1}^1)$ when $\mathbf{X}_0$ are initial conditions and $\mathbf{X}_{t-1}^1 = (\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_{t-1})$:

$$\mathsf{D}_\mathsf{x}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}\right) \text{ where } \boldsymbol{\phi} \in \boldsymbol{\Phi} \subseteq \mathbb{R}^s. \tag{1}$$

Let $\boldsymbol{\phi}_1 \in \boldsymbol{\Phi}$ and $\boldsymbol{\phi}_2 \in \boldsymbol{\Phi}$ be two distinct values of the s-dimensional parameter vector $\boldsymbol{\phi}$. When there exist observations $\mathbf{x}_t$ for which $\mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \neq \mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right)$ implies that $\boldsymbol{\phi}_1 \neq \boldsymbol{\phi}_2$, then $\boldsymbol{\phi}$ is a sufficient parameter (see Madansky, 1976). If $\boldsymbol{\phi}_1 \neq \boldsymbol{\phi}_2$ implies that there are observations $\mathbf{x}_t$ for which $\mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \neq \mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right)$, then $\boldsymbol{\phi}$ is (uniquely) identifiable. Thus it is possible to uniquely identify which parameter value generated the data only when different parameter values lead to different event probabilities. This property applies globally, $\forall \boldsymbol{\phi} \in \boldsymbol{\Phi}$. Alternatively, if there exists a neighborhood $\mathcal{N}(\boldsymbol{\phi}_1)$ of $\boldsymbol{\phi}_1$ such that $\boldsymbol{\phi}_2 \neq \boldsymbol{\phi}_1$ implies that $\mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \neq \mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right)$ $\forall \boldsymbol{\phi}_2 \in \mathcal{N}(\boldsymbol{\phi}_1)$ then $\boldsymbol{\phi}_1$ is locally identifiable. When the LDGP is uniquely identified, let the value of $\boldsymbol{\phi}$ that generated the sample data $\mathbf{X}_T^1 = (\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_T)$ be the 'true' value $\boldsymbol{\phi}_0$. Further, any 1–1 transformation of $\boldsymbol{\phi}$, $\boldsymbol{\psi} = \mathbf{f}\left(\boldsymbol{\phi}\right) \in \boldsymbol{\Psi}$, also constitutes a valid parameterization.

In general, the LDGP is unknown and so models thereof are used. Let $\mathcal{M}_1$ be an econometric model of the process generating $\mathbf{x}_t$ denoted by:

$$\mathcal{M}_1 = \{\mathsf{f}_1\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) \text{ for } t = 1, 2, ...T \text{ where } \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p\} \tag{2}$$

when $\mathsf{f}_1(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ is the postulated sequential joint density at time t, and $p < s$ (usually). If $\mathcal{M}_1$ were correctly specified, then the identifiability of $\boldsymbol{\theta}$ could be defined as for that of $\boldsymbol{\phi}$ in the LDGP above. In particular, $\boldsymbol{\theta}$ is uniquely identifiable if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ implies that $\mathsf{f}_1\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_1\right) \neq \mathsf{f}_1\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_2\right)$. However, correct specification is rare and so the identifiability of parameters and the properties of statistics in mis-specified models must be considered. The parameters of a model can be non-unique if the LDGP is not uniquely identified, or if the estimator of $\boldsymbol{\theta}$ does not converge as $T \to \infty$, or both. Confining attention to cases in which the LDGP is uniquely identified, with the 'true' value of $\boldsymbol{\phi}$ for $\mathbf{X}_T^1$ being $\boldsymbol{\phi}_0$, then assuming identifiable uniqueness on $\boldsymbol{\Theta}$ ensures that $\boldsymbol{\theta}$ is uniquely identified (see e.g., Gallant and White, 1988, and White, 1994). Indeed, under these conditions, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_T$ of $\boldsymbol{\theta}$ tends in probability to its pseudo-true value $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{\phi}_\mathbf{0})$ which is given by:

$$\boldsymbol{\theta}(\boldsymbol{\phi}_\mathbf{0}) = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathsf{E}_{LDGP}\left[L_T(\boldsymbol{\theta})\right] \tag{3}$$

when:

$$L_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \mathsf{f}_1\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right). \tag{4}$$

When $L_T(\boldsymbol{\theta})$ has a maximizer $\widehat{\boldsymbol{\theta}}_T \in \boldsymbol{\Theta}$ for each T then, requiring the sequence $\{\widehat{\boldsymbol{\theta}}_T\}$ to be identifiable uniquely on $\boldsymbol{\Theta}$ rules out the possibility that $L_T(\boldsymbol{\theta})$ becomes flatter in a neighborhood of $\widehat{\boldsymbol{\theta}}_T$ as $T \to \infty$, and precludes that there are other sequences of estimators $\{\widetilde{\boldsymbol{\theta}}_T\}$ which are such that $\{L_T(\widetilde{\boldsymbol{\theta}}_T)\}$ approaches arbitrarily closely the almost sure limit of $L_T(\boldsymbol{\theta})$ as $T \to \infty$. Thus the identification of the model parameter $\boldsymbol{\theta}$ is equivalent to the uniqueness of the pseudo-true value $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{\phi}_\mathbf{0})$.

$\mathcal{M}_1$ with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ provides the best approximation to the LDGP $\mathsf{D}_\mathsf{X}\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_0\right)$ in the sense that the Kullback–Leibler information criterion (KLIC):

$$\mathcal{I}\left(\boldsymbol{\phi}_0, \boldsymbol{\theta}\right) = \int \log \frac{\mathsf{D}_x\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_0\right)}{\mathsf{f}_1\left(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}\right)} \mathsf{D}_x\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}_0\right) \mathrm{d}\mathbf{x}_t \tag{5}$$

is minimized by $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In general $\mathcal{I}(\boldsymbol{\phi}_0, \boldsymbol{\theta}) \geqslant 0$, with $\mathcal{I}(\boldsymbol{\phi}_0, \boldsymbol{\theta}) = 0$ if and only if $\mathsf{f}_1(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta}_0) = \mathsf{D}_x(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\phi}_0)$ with probability one (see Kullback and Leibler, 1951). Note that if $\boldsymbol{\theta}_0$ is to be the unique solution to (3), then it is required that:

$$\mathsf{E}_{LDGP}\left[\frac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0} \tag{6}$$

if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\phi}_0)$. Despite being the best KLIC-approximation to the LDGP, $\mathcal{M}_1$ with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ may only be locally identified, as opposed to a correctly specified $\mathcal{M}$ which is globally identified. Equally, a uniquely identified model may not reflect completely the LDGP, or alternatively even if it does reflect the LDGP, it may not be interpretable.

2.1 An example

We first illustrate that a mis-specified model can be uniquely identified but not reflect the LDGP parameters. Section 6 discusses a model's correspondence to reality.

Suppose all *a priori* information suggests that the following model provides the best description of the data $(x_1, x_2, ...x_T)$:

$$f(\theta) = \begin{cases} \theta x^{\theta-1} & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}, \quad \theta > 0$$

when in fact the LDGP is a uniform distribution on $[0,\ 1+\delta]$, $\delta > 0$. Then, the expectation of the log-density $\log f(\theta)$ under the LDGP is given by:

$$\begin{aligned} \mathsf{E}_{LDGP}[\log\ f(\theta)] &= \frac{1}{1+\delta}\int_0^{1+\delta} \log f(\theta)\ dx \\ = & \frac{1}{1+\delta}\left\{\int_0^1 [\log\theta + (\theta-1)\log x]\, dx + \int_1^\delta 0 dx\right\} \\ = & \frac{1}{1+\delta}\{\log\theta - (\theta-1)\}, \end{aligned}$$

which attains its maximum at:

$$\theta^*(\delta) = \underset{\theta>0}{\operatorname{argmax}}\left\{\frac{1}{1+\delta}[\log\theta - (\theta-1)]\right\} = 1.$$

Hence, the pseudo-true value $\theta^*(\delta) = 1$ is uniquely determined, implying that θ is identified. However, $\theta^*(\delta)$ does not depend on the LDGP parameter δ, and so any change in δ (with $\delta > 0$) will leave $\theta^*(\delta)$ unaffected at unity. In particular, though $f(\theta^*(\delta))$ is a uniform distribution, it is defined on a different interval from that of the LDGP: observations outside the interval $[0,1]$ would of course reveal that mis-specification. Equally, each quasi log-likelihood function:

$$\log L_T(\theta) = T\log\theta + (\theta-1)\sum_{t=1}^{T}\log x_t,$$

has a well defined maximum at:

$$\theta^*_T = \frac{-T}{\sum_{t=1}^T \log x_t},$$

the existence and uniqueness of which does not depend on δ in the LDGP.

We next illustrate that a model which does reflect the LDGP parameters may nonetheless be a worse description thereof. Let $\delta < 0$ in the LDGP, hence:

$$\begin{aligned} \int \log\ f(\theta)\, h(\delta) = \frac{1}{1+\delta}\int_0^{1+\delta} \log f(\theta)\ dx \\ = \frac{1}{1+\delta}\left\{\int_0^{1+\delta}\left[\log\theta + (\theta-1)\log x\right]dx\right\} \\ = \log\theta + (\theta-1)\left\{\log(1+\delta)-1\right\}, \end{aligned}$$

implies that the pseudo-true value is given by:

$$\begin{aligned} \theta^*(\delta) &= \underset{\theta>0}{\operatorname{argmax}}\left\{\log\theta + (\theta-1)\left[\log(1+\delta)-1\right]\right\} \\ &= \frac{1}{1-\log(1+\delta)}. \end{aligned} \tag{7}$$

Monotonicity of the logarithmic function then guarantees that $\theta^*(\delta)$ is uniquely determined for $-1 < \delta < 0$. Thus the model parameter θ is uniquely identified and does reflect the LDGP, but the model is no longer a uniform distribution.

Finally, it is possible for a model to be uniquely identified and reflect the LDGP parameters, but not be interpretable – also see section 4. The 'classic' example is regression estimation of an unidentifiable supply-demand model in price (p_t) and quantity (q_t) which nevertheless delivers a unique function of the LDGP parameters and the error (co)variances:

$$\begin{aligned} p_t &= \mu_{11} q_t + v_{1,t} \\ q_t &= \mu_{21} p_t + v_{2,t} \end{aligned}$$

with:

$$\begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \sim \mathsf{IN}_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right],$$

then the OLS estimator of μ_{11} converges to:

$$\underset{T\to\infty}{\operatorname{plim}}\ \widehat{\mu}_{11} = \frac{(1+\mu_{11}\mu_{21})\,\sigma_{12} + \mu_{11}\sigma_{22} + \mu_{21}\sigma_{11}}{\sigma_{22} + 2\mu_{21}\sigma_{12} + \mu_{21}^2\sigma_{11}},$$

which could have either sign (but the same sign as $\widehat{\mu}_{21}$).

3 Observational equivalence and mutual encompassing

When two models always generate identical outcomes, they are observationally equivalent and data alone cannot distinguish between them. A sufficient condition is that all their parameters be unidentifiable, but this is not necessary, and identified models can be observationally equivalent. Observational equivalence arises whenever there is an unidentified model, and there is an equivalence set of models that impose just-identifying restrictions. As an example consider the following bivariate regression model:

$$y_t = \alpha + \beta x_t + u_t \ \text{ with } \ u_t \sim \mathsf{IN}\left[\mu, \sigma^2\right] \tag{8}$$

in which μ and α are not uniquely identified. However, the set of models that imposes a single restriction on μ and α (e.g., $\mu = 0$ or $\alpha = \alpha^*$) forms a set of models that cannot be distinguished on the basis of observations. More generally at the level of the LDGP, since $\mathsf{D_X}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\phi})$ is unchanged under 1–1 transformations of the parameter $\boldsymbol{\phi}$ to $\boldsymbol{\psi}=\boldsymbol{\psi}(\boldsymbol{\phi})\in\boldsymbol{\Psi}$ then $\mathsf{D_X}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\phi})$ and $\mathsf{D_X}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\psi})$ are observationally equivalent and hence isomorphic. If $\boldsymbol{\psi}=\boldsymbol{\psi}(\boldsymbol{\phi})$ but is not 1–1 (e.g., due to imposing some irrelevant parameters at their population values of zero), the processes are said to be equivalent.

Observationally-equivalent models are KLIC-equivalent, in that the relevant version of the criterion in (5) will be zero. Equally, since encompassing is the ability of one model to account for the salient features of another model (see Mizon, 1984, Mizon and Richard, 1986, and Hendry, 1995a), observationally-equivalent models will encompass each other, that is, be mutually encompassing. In analyzing the relationships between observational equivalence, KLIC equivalence, and encompassing, Lu and Mizon (1999) showed that models are KLIC-equivalent if and only if they are mutually encompassing with respect to their complete parameter vectors (complete parametric encompassing) and their log sequential densities (Cox encompassing). Further, Bontemps and Mizon (2001) defined a congruent model to be one that parsimoniously encompasses the LDGP, and showed that congruence of a nesting model is sufficient for it to encompass models nested within it. Therefore, an example of mutual encompassing arises whenever a nesting model is both congruent and parsimoniously encompassed by a nested model.

A distinction can be drawn between population and sample mutual encompassing. Mutual encompassing in the population is observational equivalence. For example, there might exist an equivalence set of representations of the LDGP, in which the representations are usually re-parameterizations of each other, though not necessarily having parameter spaces of the same dimension. However, mutual encompassing in the sample can arise from observational equivalence, or from weak evidence resulting in the models being indistinguishable on the basis of the available information.

3.1 An example

Consider the congruent representation of the LDGP for y_t given in $\mathcal{M}_2$:

$$\mathcal{M}_2\colon y_t=\mu+\epsilon_t+\theta\epsilon_{t-1} \tag{9}$$

when $\epsilon_t\sim\mathsf{IN}[0,1]$ and $|\theta|<1$. An alternative, and observationally-equivalent, representation of the LDGP is given by $\mathcal{M}_3$:

$$\mathcal{M}_3\colon y_t=\frac{\mu}{1+\theta}+\sum_{i=1}^{\infty}(-\theta)^i y_{t-i}+\epsilon_t. \tag{10}$$

In this simple example, $\mathcal{M}_2$ and $\mathcal{M}_3$ are mutually encompassing, both congruent, and observationally equivalent. However, $\mathcal{M}_3$ is only relevant in the population, since only a finite-order autoregression can be estimated using sample data. In fact, since $|\theta|<1$, a finite-order autoregression:

$$\mathcal{M}_4\colon y_t=\alpha+\sum_{i=1}^{m}\beta_i y_{t-i}+u_t \tag{11}$$

will give a good approximation to (10) when only sample data is available. Indeed, $\mathcal{M}_4$ with $m=3$ is likely to be indistinguishable from $\mathcal{M}_2$ empirically, so both $\mathcal{M}_2$ and $\mathcal{M}_4$ would be empirically congruent even though only the former is congruent. Bontemps and Mizon (2001) contains a more detailed analysis of a related example.

4 Non-unique just-identified representations

A well-known example of identified models forming an equivalence set arises in the just-identified simultaneous equations model (SEM). Consider a closed vector autoregression (VAR):

$$\mathbf{x}_t = \sum_{i=1}^{k} \mathbf{D}_i \mathbf{x}_{t-i} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t \quad \text{with } \boldsymbol{\varepsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}\right] \tag{12}$$

which can be written alternatively as a vector equilibrium-correction model (VEqCM):

$$\Delta\mathbf{x}_t = \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\pi}\mathbf{x}_{t-1} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t \quad \text{with } \boldsymbol{\varepsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}\right]. \tag{13}$$

When $\boldsymbol{\pi}$ has full rank n, the variables $\mathbf{x}_t$ are $\mathsf{I}(0)$, and the parameters $(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_{k-1}, \boldsymbol{\pi}, \boldsymbol{\delta}, \boldsymbol{\Omega})$ or $(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_k, \boldsymbol{\delta}, \boldsymbol{\Omega})$ are all identified, in that their maximum likelihood estimators are unique, and obtained by multivariate least squares. Indeed, the VAR and the VEqCM are both just-identified and observationally-equivalent models. The set of observationally-equivalent just-identified models, though, includes far more than these two models. The parameters of interest for many investigators are those of a SEM such as:

$$\mathbf{A}_0\mathbf{x}_t = \sum_{i=1}^{k} \mathbf{A}_i \mathbf{x}_{t-i} + \mathbf{c} + \mathbf{v}_t \quad \text{with } \mathbf{v}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Sigma}\right] \tag{14}$$

rather than the VAR or the VEqCM. Without further information, the parameters of (14) are unidentified as is well known, and this leads to the traditional analysis of identification in simultaneous equations models – see *inter alia* Spanos (1986) and Greene (2000). All the SEMs resulting from sets of *a priori* restrictions on $(\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_k, \mathbf{c}, \boldsymbol{\Sigma})$ that achieve just-identification are observationally equivalent, and thus observationally equivalent to the VAR and the VEqCM in (12) and (13) respectively.

A further identification issue arises when rank$(\boldsymbol{\pi}) = r < n$, in which case $\mathbf{x}_t \backsim \mathsf{I}(1)$, but there are r cointegrating vectors $\boldsymbol{\beta}'\mathbf{x}_t \backsim \mathsf{I}(0)$. In this case, (13) becomes:

$$\Delta\mathbf{x}_t = \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t \quad \text{with } \boldsymbol{\varepsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}\right],$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r. It is well known that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identified without further restrictions. Nevertheless, the Johansen procedure (see e.g., Johansen, 1995) for empirically determining the value of r, produces unique estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as a result of requiring $\boldsymbol{\beta}$ to be orthogonal and normalized (see e.g., Johansen and Juselius, 1994). This estimate of $\boldsymbol{\beta}$, for a given value of r, spans the space of just-identified cointegrating vectors, and is observationally equivalent to any other just-identified estimate of $\boldsymbol{\beta}$. The fact that the just-identified estimate of $\boldsymbol{\beta}$ provided by the Johansen procedure may not have an economic interpretation is usually unimportant, since this estimate is only used to provide a value for the unrestricted log likelihood function to be compared with the value of the log-likelihood function corresponding to sets of over-identifying restrictions on $\boldsymbol{\beta}$ which do have an economic interpretation. When $r = 1$, of course, $\boldsymbol{\beta}$ should have an economic interpretation, perhaps subject to eliminating irrelevant coefficients; and may do for $r > 1$.

In both cases, mis-interpretation of $\boldsymbol{\beta}$ can occur. One illustration is when an equation normalized on (say) money, is interpreted as a 'long-run money-demand relation' because $\gamma > 0$:

$$m - p - y = -\gamma\left(R_l - R_s\right), \tag{15}$$

(where m is nominal money, p is the price deflator of real income y, and R_l and R_s are long- and short-term interest rates), but actually is an 'interest-rate spread' equation, as in:

$$R_l = R_s + \gamma^{-1} v,$$

where $v = p + y - m$ is the velocity of circulation. The values of the α_{ij} can help discrimination (see e.g., Hendry and Juselius, 2001, for an exposition): if there was no feedback from (15) onto money (interest rates), one might question the first (second) interpretation respectively.

5 Non-unique over-identified representations

A related class of model where the common interpretation of the available results on identification may not always be well founded is that of over-identified simultaneous equations models. Though this analysis applies for closed versions of linear dynamic models, such as the VAR and VEqCM in (12) and (13), we use a notation that usually is associated with static SEMs:[3]

$$\mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{z}_t = \mathbf{u}_t \text{ with } \mathbf{u}_t \sim \mathsf{IN}_q \left[\mathbf{0}, \boldsymbol{\Sigma}\right]. \tag{16}$$

Several models of $(\mathbf{y}_t, \ \mathbf{z}_t)$ can be over-identified, satisfy the rank condition, and not fail over-identification tests empirically, even when such models conflict theoretically (see Hendry and Mizon, 1993). Thus, the Cowles' rank condition is insufficient for the three attributes, although it is sufficient for uniqueness within theories, thus achieving uniqueness for a given interpretation as we now show.

Consider all linear transforms $\mathbf{R}$ of $(\mathbf{B} : \mathbf{C})$ to establish whether the uniquely admissible $\mathbf{R}$ is $\mathbf{R} = \mathbf{I}_q$. When $(\mathbf{B} : \mathbf{C})$ are unconstrained, $(\mathbf{RB} : \mathbf{RC})$ comprise all linear systems. However, when $(\mathbf{B} : \mathbf{C})$ are restricted, admissible $\mathbf{R}$s are only relative to the restrictions on the given choice, so no longer span all relevant linear models. Thus:

$$\mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{z}_t = \mathbf{u}_t \text{ and } \mathbf{B}^*\mathbf{y}_t + \mathbf{C}^*\mathbf{z}_t = \mathbf{u}_t^*,$$

can generate the same $\{\mathbf{y}_t\}$, so long as:

$$\mathbf{B}^{-1}\mathbf{C} = \left(\mathbf{B}^*\right)^{-1}\mathbf{C}^* = -\boldsymbol{\Pi},$$

and:

$$\mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1\prime} = \left(\mathbf{B}^*\right)^{-1}\boldsymbol{\Sigma}^*(\mathbf{B}^*)^{-1\prime}.$$

The equivalence class is:

$$\mathbf{C}^* = \mathbf{B}^*\mathbf{B}^{-1}\mathbf{C}, \tag{17}$$

or any $\mathbf{S}$ such that $\mathbf{B}^* = \mathbf{SB}$ at the same time as $\mathbf{C}^* = \mathbf{SC}$, even though within their own restriction sets, $(\mathbf{B} : \mathbf{C})$ and $(\mathbf{B}^* : \mathbf{C}^*)$ are both uniquely identified. This matches Hsiao (1983), who proves that observational equivalence requires such an $\mathbf{S}$ since, from (17) we then have:

$$\mathbf{C}^* = \mathbf{B}^*\mathbf{B}^{-1}\mathbf{C} = \mathbf{SBB}^{-1}\mathbf{C} = \mathbf{SC},$$

consistent with his claim. Therefore, all examples must satisfy this restriction.

How does this relate to the analysis of the Cowles Commission researchers? They sought $\mathbf{R}$ such that for *restricted* $\mathbf{B}$ and $\mathbf{C}$, the only admissible $\mathbf{R}$ is $\mathbf{R} = \mathbf{I}_q$ where:

$$\left(\mathbf{RB} : \mathbf{RC}\right) = \left(\mathbf{B}^* : \mathbf{C}^*\right).$$

[3] This is not a limitation since the $k \times 1$ vector $\mathbf{z}_t$ in (16) could be defined to include lagged values of the $q \times 1$ vector $\mathbf{y}_t$.

Such an $\mathbf{R}$ must satisfy the *a priori* constraints on the $(\mathbf{B}:\mathbf{C})$ matrix. It is clear that the (unrestricted) $\mathbf{\Pi}$ matrix in:

$$\mathbf{y}_t + \mathbf{\Pi}\mathbf{z}_t = \mathbf{v}_t \text{ with } \mathbf{v}_t \sim \mathsf{IN}_q\left[\mathbf{0}, \mathbf{\Omega}\right],$$

is always identified because:

$$\left(\mathbf{DI}_q : \mathbf{D\Pi}\right) = \left(\mathbf{I}_q : \mathbf{\Pi}\right),$$

enforces $\mathbf{D} = \mathbf{I}_q$: that result holds true independently of the correctness (or otherwise) of the model specification, and the interpretability of its coefficients. The Cowles' rank condition ensures the same for $(\mathbf{B}:\mathbf{C})$ – but it does not preclude the possibility of a differently restricted $(\mathbf{B}^*:\mathbf{C}^*)$ that also satisfies the rank condition, such that:

$$\left(\mathbf{B}^* : \mathbf{C}^*\right) = \left(\mathbf{SB} : \mathbf{SC}\right),$$

which is thus a member of the equivalence set.

5.1 An example

We consider an example in which there are four alternative observationally-equivalent representations, each of which is over-identified and has (potentially at least) an economic interpretation. The system ('reduced form') itself is restricted, as in the following LDGP for a set of two endogenous variables conditional on four strongly exogenous regressors $(y_{1,t}, y_{2,t}, z_{1,t}, z_{2,t},\ z_{3,t},\ z_{4,t})$:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & 0 \\ \pi_{21} & 0 & \pi_{23} & \pi_{24} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \tag{18}$$

The reduced-form coefficient matrix $\mathbf{\Pi}$ in this case has six free elements, and imposes two restrictions. Notice that any linear restrictions can be re-parameterized to zero restrictions. Consequently, one of the over-identified representations is (18). Two other over-identified representations follow.

5.1.1 Simultaneous representation 1

Consider the simultaneous-equations representation given in (19):

$$\begin{pmatrix} 1 & b_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & 0 & c_{14} \\ c_{21} & 0 & c_{23} & c_{24} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}, \tag{19}$$

where:

$$\begin{aligned} \mathbf{\Pi} &= \begin{pmatrix} 1 & b_{12} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} c_{11} & c_{12} & 0 & c_{14} \\ c_{21} & 0 & c_{23} & c_{24} \end{pmatrix} \\ &= \begin{pmatrix} c_{11} - b_{12}c_{21} & c_{12} & -b_{12}c_{23} & c_{14} - b_{12}c_{24} \\ c_{21} & 0 & c_{23} & c_{24} \end{pmatrix} \\ &= \begin{pmatrix} c_{11} - b_{12}c_{21} & c_{12} & -b_{12}c_{23} & 0 \\ c_{21} & 0 & c_{23} & c_{24} \end{pmatrix}. \end{aligned}$$

Comparison with (18) shows that the population values must satisfy $c_{14} = b_{12}c_{24}$: clearly, it is necessary that $c_{24} \neq 0$ otherwise $z_{4,t}$ becomes an irrelevant variable. This representation will be valid if and only if $b_{12} = -\pi_{13}/\pi_{23}$ with $\pi_{23} \neq 0$, which defines b_{12}. The second equation is over-identified, and hence so is the system although the first is just identified (imposing $c_{14} = b_{12}c_{24}$ would ensure both equations were over-identified).

5.1.2 Simultaneous representation 2

A second over-identified formulation consistent with (18) is given by:

$$\begin{pmatrix} 1 & 0 \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} & d_{13} & 0 \\ 0 & d_{22} & d_{23} & d_{24} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \qquad (20)$$

where now:

$$\mathbf{\Pi} = \begin{pmatrix} d_{11} & d_{12} & d_{13} & 0 \\ -b_{21}d_{11} & 0 & d_{23} - b_{21}d_{13} & d_{24} \end{pmatrix},$$

so this system requires $d_{22} = b_{21}d_{12}$ and imposes $b_{21} = -\pi_{21}/\pi_{11}$ which defines b_{21}.

Thus, all three over-identified models are observationally equivalent. To satisfy that requirement, we must have:

$$\mathbf{\Pi} = \begin{pmatrix} \pi_{11} & \pi_{12} & -b_{12}\pi_{23} & 0 \\ -b_{21}\pi_{11} & 0 & \pi_{23} & \pi_{24} \end{pmatrix},$$

where b_{12} and b_{21} are defined above. Such a $\mathbf{\Pi}$ has six free elements as required, and satisfies the two restrictions in (18). The matrix $\mathbf{S}$ above that links the two simultaneous-equations representations is:

$$\mathbf{S} = \begin{pmatrix} 1 - b_{12}b_{21} & b_{12} \\ -b_{21} & 1 \end{pmatrix},$$

where $|\mathbf{S}| = 1$.

5.1.3 Simultaneous representation 3

The final over-identified formulation consistent with (18) combines the two ‘simultaneous’ relations:

$$\begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & 0 & f_{14} \\ 0 & f_{22} & f_{23} & f_{24} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}, \qquad (21)$$

where $b_{12}b_{21} \neq 1$ and:

$$\mathbf{\Pi} = \frac{1}{1 - b_{12}b_{21}} \begin{pmatrix} f_{11} & (1 - b_{12}b_{21})\, f_{12} & -b_{12}f_{23} & 0 \\ -b_{21}f_{11} & 0 & f_{23} & (1 - b_{21}b_{12})\, f_{24} \end{pmatrix},$$

which implies that $b_{21} = -\pi_{21}/\pi_{11}$ and $b_{12} = -\pi_{13}/\pi_{23}$ (both scaled by the determinant $\pi_{23}\pi_{11}/\left(\pi_{23}\pi_{11} - \pi_{13}\pi_{21}\right)$): enforcing $f_{22} = b_{21}f_{12}$ and $f_{14} = b_{12}f_{24}$ would ensure both equations were over-identified.

Consequently, demonstrating that a given model is over-identified and that all the over-identifying restrictions are valid (even asymptotically) is insufficient to demonstrate that it is a unique representation.

6 Identifying structure

We turn to the next attribute of identification, namely 'correspondence to reality'. This notion shares features with the time-series concept of identification. Historically, Frisch (1938) thought that structure was inherently unidentifiable (see the commentary offered in Hendry and Morgan, 1995), and it is easy to construct examples of unidentified structures, where a non-structural sub-system is identified. Consider the system:

$$m_t - \rho y_t - \lambda p_t = \boldsymbol{\delta}_1' \mathbf{z}_t + v_{1,t} \tag{22}$$

$$y_t - \phi p_t = \boldsymbol{\delta}_2' \mathbf{z}_t + v_{2,t} \tag{23}$$

$$p_t = \boldsymbol{\delta}_3' \mathbf{z}_t + v_{3,t}. \tag{24}$$

Here the $\mathbf{z}_t$ are strongly exogenous, and the $v_{i,t}$ are iid errors. When there are no restrictions on $\boldsymbol{\delta}_1$, then (22) is not identifiable.

However, consider a setting where y_t is both unobserved, and its relevance is not realized, so analyses only consider the non-structural sub-system:

$$m_t - (\lambda + \rho\phi)\, p_t = \left(\boldsymbol{\delta}_1' + \rho\boldsymbol{\delta}_2'\right) \mathbf{z}_t + v_{1,t} + \rho v_{2,t} \tag{25}$$

$$p_t = \boldsymbol{\delta}_3' \mathbf{z}_t + v_{3,t}. \tag{26}$$

When a theory correctly specifies that sufficient elements of $\boldsymbol{\delta}_1' + \rho\boldsymbol{\delta}_2'$ are zero, (25) can be identifiable on the conventional rank condition in this bivariate process. Equation (25) may even be interpretable (e.g., as a money-demand equation when $\lambda + \rho\phi = 1$), but it obviously does not correspond to the structure.

Thus, conventional notions of identification are indeed limited to 'uniqueness', despite the sometimes ambiguous use of the phrase noted in the introduction. In the next section, we consider whether 'structural change' in an economy can help to discriminate between structural and non-structural representations of the LDGP.

7 Structural change

We define structure as the set of invariant features of the economic mechanism (see Frisch, 1934, Haavelmo, 1944, Wold and Juréen, 1953, and Hendry, 1995b), or more precisely, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ defines a structure if $\boldsymbol{\theta}$ directly characterizes the relations of the economy and is invariant over time, and to extensions of the information set and policy interventions. The last three attributes are empirically testable, although the first is not, as with the corresponding attribute of identification. Consequently, although all four representations considered in the example of section 5.1 are well defined and observationally equivalent in a constant-parameter world, at most one representation of each equation can be invariant to changes and hence be structural (see Hendry and Mizon, 1993, and Hendry, 1995a).

Clearly, no parameterization $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ can be invariant to all change, so structure is relative – atomic war would radically alter the economies of all participants. Thus, the class of 'admissible' interventions must be delineated. In their analyses of the sources of forecast errors, Clements and Hendry (1998, 1999) find that shifts in deterministic terms (intercepts and deterministic trends etc.) are the primary cause of forecast failure, a result corroborated by the Monte Carlo results in Hendry (2000) where other forms of structural break (e.g., in other parameters) did not induce forecast failure. Consequently, invariance to deterministic shifts in unmodeled variables seems one essential (but insufficient) requirement of structure. Indeed, the concept of identification in Working (1927) and Frisch (1938) is close to using 'shifts' to isolate the invariant structure.

7.1 External change

In terms of the preceding example, all four representations are equally structural to external changes, such as shifts in the distributions of the $z_{k,t}$, precisely because they all correspond to the same system (reduced form) which is itself invariant to such shifts.

Nevertheless, consider a model of the first equation that also – or incorrectly – included $z_{4,t}$ when its long-run mean (i.e., $\mathsf{E}[z_{4,t}] = \mu_4$ say) changed: then that equation would forecast a shift in $y_{1,t}$ which did not materialize, hence fail to be structural on the grounds of a lack of invariance. As argued in Hendry and Mizon (1993), such 'spurious' structures can often be detected using structural breaks induced by natural experiments and policy changes. Some of the simulation experiments in Clements and Hendry (2001) illustrate this situation. Thus, mis-specification (or serious mis-estimation) can be revealed by external structural change, but alternative, correctly-specified, mutually-encompassing, representations all survive.

However, the issue is complicated by the possibility of 'extended constancy' (see Hendry, 1996, and Ericsson, Hendry and Prestwich, 1998): a model may fail on forecasts because of a deterministic shift, be extended to incorporate the variable that changed, reparameterized to link that variable into an already included regressor (as in an interest-rate differential, say) and finish with precisely the same parameters on the same number of variables as initially, but with one variable redefined (or, more precisely, re-measured). The example in section 7.3 illustrates this phenomenon.

7.2 Internal change

Now we allow for one of the π_{ij} in each equation to shift, so neither equation in the 'reduced form' (18) can be structural, nor can either of the corresponding equations in (19) or (20). However, for shifts of the form $\pi_{13} = b_{21}\pi_{23}$ then b_{12} remains constant, as do all the other parameters of the first equation, so it is structural in (19) and (21); equivalently for the second equation when $\pi_{21} = b_{12}\pi_{11}$ (say). However, if both π_{11} and π_{23} shift, neither equation in (21) remains constant, and the matching equation in (19) or (20) only remains constant if other parameters also shift (e.g., π_{21} offsets the change in π_{11} for (19) to remain constant).

The first example can be made a special case of the second by endogenizing all the variables, in which case, shifts in some variables' distributions preclude them from being structural, albeit that the same shifts might highlight the structurality of other equations.

An interesting re-interpretation of identification follows when the internal break is discovered and correctly modelled. Consider (19) when $\pi_{13} = b_{12}\pi_{23}$ but π_{23} shifts at time T_1 to $\pi_{23} + \eta$. Let $1_{\{t \geq T_1\}}$ denote the indicator for this event, and consider the augmented system:

$$\begin{pmatrix} 1 & b_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & 0 & c_{14} & 0 \\ c_{21} & 0 & c_{23} & c_{24} & \eta \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \\ 1_{\{t \geq T_1\}} z_{3,t} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}. \quad (27)$$

Then (27) has constant parameters, but the first equation is more strongly over-identified than that in (19), confirming the identifying role of 'shifts' first discussed by Wright (1915).

7.3 An example

The Banking Act of 1984 in the UK, which permitted interest payments on current accounts in exchange for all interest payments being after the deduction of 'standard rate' tax, provided a natural experiment

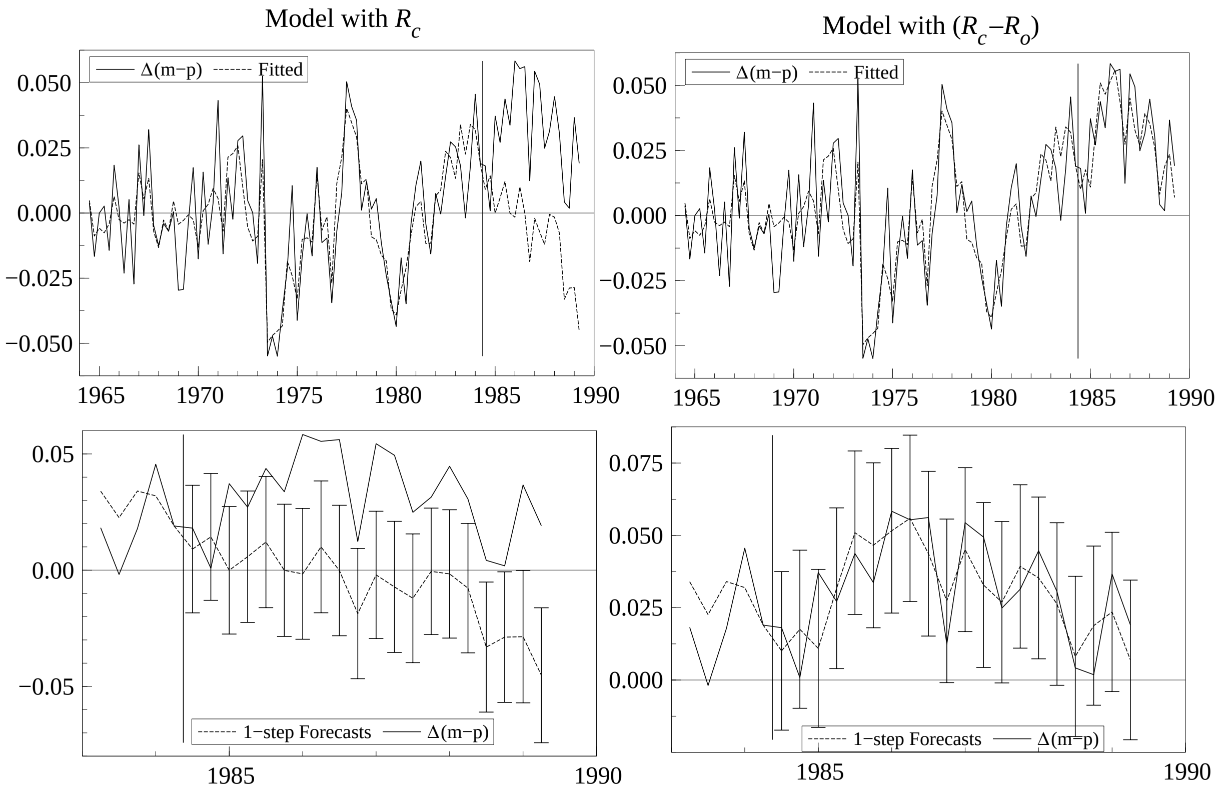

Figure 1 The effects of a change in the opportunity cost of holding money.

that illustrates the role of structural breaks in isolating structure. Following this legislative change, previously-estimated models of the demand for narrow money (M1) (such as Hendry and Mizon, 1993) suffered serious forecast failure. This is shown in the first column of graphs in figure 1 for which the model used the competitive rate of interest R_c as a measure of the opportunity cost of holding money. In fact, the own rate of interest (R_o) changed from zero to near the value of the competitive rate (R_c: about 12 per cent per annum at the time) within 18 months, inducing very large inflows to M1. The effect was a large shift in the opportunity cost of holding money, namely a deterministic shift from R_c to $(R_c - R_o)$. Models that correctly re-measured the opportunity cost by $(R_c - R_o)$ continued to forecast well, once the break was observed – see the second column of graphs in figure 1. Moreover, these models had the same estimated parameter values after the break as before. Thus, the forecast failure of models using R_c as a proxy for the opportunity cost was instrumental in the recognition that, once a more appropriate measure of opportunity cost was used, there was no change in structure for the money demand equation, although there clearly was for the opportunity cost equation (or both equations for R_c and R_o).

8 Conclusions

If identification were no more than the uniqueness of a parameterization, it could be achieved by imposing sufficient arbitrary restrictions. We suspect that was not what Cowles Commission researchers envisaged, nor does it correspond to the normal use of language: few would accept the 'identification' of an approaching 'Mini' as a Rolls-Royce by false claims as to its size, shape, composition and form. Rather, the attributes of correct interpretation and correspondence to an actual entity are also important. Thus, we re-considered these three attributes, as well as the potential identification of 'structure', defined as invariance under extensions of the information set.

First, for uniqueness, both local and global identification of the parameters of the local data generation process (LDGP) and of a model were considered. The converse of uniqueness is observational equivalence, which arises whenever, in the population, models are mutually encompassing. In sample, though, mutual encompassing can arise from the available information being unable to discriminate between distinct models.

Next, we showed that Cowles Commission rank conditions for simultaneous equations models (SEMs) only ensure uniqueness within a theory – there can be other over-identified (and interpretable) models that are observationally equivalent under constant parameters. Forecast failure and structural change can be valuable in discriminating non-structural (but uniquely over-identified) representations, from those which potentially contain structure. However, there is no guarantee that structure can be identified. Finally, interpretation remains in the eye of the beholder, usually dependent on a theoretical framework. However, rejection of the relevant theory-based identifying restrictions, or violation of theory-derived constancy requirements, would preclude such interpretations.

References

Aldrich, J. (1994). Haavelmo's identification theory. *Econometric Theory*, **10**, 198–219.

Bontemps, C., and Mizon, G. E. (2001). Congruence and encompassing. In Stigum, B. (ed.), *Studies in Economic Methodology*. Cambridge, MA: MIT Press.

Bowden, R. (1973). The theory of parametric identification. *Econometrica*, **41**, 1069–1074.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day. First published, 1970.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2001). Modelling methodology and forecast failure. *Econometrics Journal*, **Forthcoming**, –.

Deistler, M. (1976). The identifiability of linear econometric models with autocorrelated errors. *International Economic Review*, **17**, 26–45.

Deistler, M., Ploberger, W., and Pötscher, B. M. (1982). Identifiability and inference in ARMA systems. In Anderson, O. (ed.), *Time Series Analysis: Theory and Practice 2*, pp. ??–?? Amsterdam: North-Holland.

Deistler, M., and Seifert, H. (1978). Identifiability and consistent estimability in econometric models. *Econometrica*, **46**, 969–980.

Ericsson, N. R., Hendry, D. F., and Prestwich, K. M. (1998). The demand for broad money in the United Kingdom, 1878–1993. *Scandinavian Journal of Economics*, **100**, 289–324.

Faust, J., and Whiteman, C. H. (1997). General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model of the DGP: A translation and critique. *Carnegie–Rochester Conference Series on Public Policy*, **47**, 121–161.

Fisher, F. M. (1961). On the cost of approximate specification in simultaneous equation estimation. *Econometrica*, **29**, 139–170.

Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York: McGraw Hill.

Frisch, R. (1934). *Statistical Confluence Analysis by means of Complete Regression Systems*. Oslo: University Institute of Economics.

Frisch, R. (1938). Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in Hendry D. F. and Morgan M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Gallant, A. R., and White, H. (1988). *A unified theory of estimation and inference for nonlinear dynamic models*. Oxford: Blackwell.

Greene, W. H. (2000). *Econometric Analysis*. New Jersey: Prentice-Hall. fourth edition.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, **12**, 1–118. Supplement.

Hannan, E. J., and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. John Wiley & Sons: New York.

Hatanaka, M. (1975). On the global identification of the dynamic simultaneous equations model with stationary disturbances. *International Economic Review*, **16**, 545–554.

Hendry, D. F. (1995a). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1995b). Econometrics and business cycle empirics. *Economic Journal*, **105**, 1622–1636.

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.

Hendry, D. F., and Juselius, K. (2001). Explaining cointegration analysis: Part II. *Energy Journal*, **22**, 75–120.

Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.

Hendry, D. F., and Morgan, M. S. (1995). *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Hsiao, C. (1983). Identification. In Griliches, Z., and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 1, Ch. 4. Amsterdam: North-Holland.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S., and Juselius, K. (1994). Identification of the long-run and the short-run structure: An application to the ISLM model. *Journal of Econometrics*, **63**, 7–36.

Kalman, R. E. (1982). Identification from real data. In Hazewinkel, M., and Rinrooy Kan, A. H. G. (eds.), *Current Developments in the Interface: Economics, Econometrics, Mathematics*, pp. 161–196. Dordrecht: D. Reidel.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica*, **17**, 125–144. Reprinted with minor revisions in Hood, W. C. and Koopmans, T. C. (eds.) (1953), *Studies in Econometric Method*. Cowles Commission Monograph 14, New York: John Wiley & Sons.

Koopmans, T. C., and Reiersøl, O. (1950). The identification of structural characteristics. *The Annals* *of Mathematical Statistics*, **21**, 165–181.

Koopmans, T. C., Rubin, H., and Leipnik, R. B. (1950). Measuring the equation systems of dynamic economics. In Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models*, No. 10 in Cowles Commission Monograph, Ch. 2. New York: John Wiley & Sons.

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Liu, T. C. (1960). Underidentification, structural estimation, and forecasting. *Econometrica*, **28**, 855–865.

Lu, M., and Mizon, G. E. (1999). Mutual encompassing and model equivalence. mimeo, Economics Department, University of Southampton.

Madansky, A. (1976). *Foundations of Econometrics*. Amsterdam: North-Holland.

Marschak, J. (1942). Economic interdependence and statistical analysis. In Lange, I., *et al.* (eds.), *Studies in Mathematical Economics and Econometrics – In Memory of Henry Schultz*, pp. 125–150. Chicago: University of Chicago Press.

Mizon, G. E. (1984). The encompassing approach in econometrics. In Hendry, D. F., and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, pp. 135–172. Oxford: Basil Blackwell.

Mizon, G. E. (1995). 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., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.

Preston, A. J. (1978). Concepts of structure and model identifiability for econometric systems. In Bergstrom, A. R., Catt, A. J. L., Peston, M. H., and Silverstone, B. D. J. (eds.), *Stability and Inflation*, Ch. 16. New York: Wiley.

Qin, D. (1989). Formalisation of identification theory. *Oxford Economic Papers*, **41**, 73–79.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, **39**, 577–592.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

Wold, H. O. A., and Juréen, L. (1953). *Demand Analysis: A Study in Econometrics*, 2nd edn. New York: John Wiley.

Working, E. J. (1927). What do statistical demand curves show?. *Quarterly Journal of Economics*, **41**, 212–235.

Wright, P. G. (1915). Review of Moore, 'economic cycles' (1915). *Quarterly Journal of Economics*, **29**. Reprinted in Hendry, D. F. and Morgan M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.