LSE CPNSS

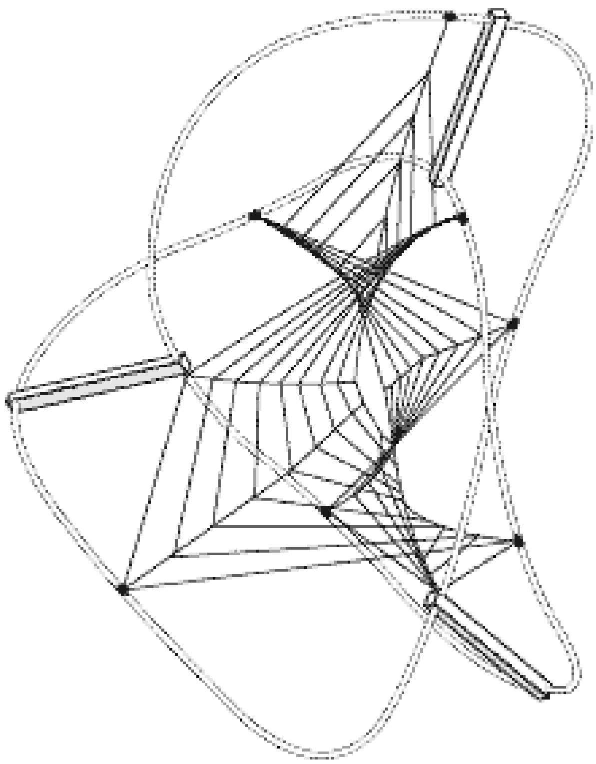
Centre for Philosophy of Natural and Social Science

Causality: Metaphysics and Methods

Technical Report 18/04

Causality and Exogeneity in Non-stationary Economic Time-Series

David F. Hendry



Editor: Julian Reiss

Causality and Exogeneity in Non-stationary Economic Time Series

David F. Hendry*
Economics Department, University of Oxford.
david.hendry@nuffield.ox.ac.uk

June 10, 2004

Abstract

We review the roles of causality and exogeneity in macro-econometric time-series modelling, forecasting and policy; their inter-relationships; problems in establishing them empirically; the importance played by non-stationarity in the data generation process (DGP), and the relation of causality to Granger causality. Although causality and exogeneity need not be invariant features of the DGP, they remain relevant in non-stationary processes, and inferences about them become easier in some ways, though more difficult in others.

Keywords: causality; exogeneity; non-stationarity; Granger causality; forecasting; policy analysis
JEL classifications: E41, C52.

1 Introduction

Scientific knowledge is always and everywhere fallible. The history of Newtonian gravitational theory is the classic example—once deemed unassailable truth and repeatedly confirmed by experiment, it is now seen as an approximate model. Interestingly, that view was in fact first proposed by Adam Smith (1795). 'Causal knowledge' is simply a special case of this general problem, exacerbated by disputable formulations of the concept, doubts about inference procedures, and in economics, by the complexities of data processes. This chapter considers causality in non-stationary macro-economic time-series, where causality is viewed as 'actually determining an aspect of behaviour': exogeneity—the property of being 'determined outside the system under analysis'—is also considered, but primarily in relation to causality.

The literature on both concepts is vast, and thoroughly discussing it is well beyond the scope of this chapter. However, many of the implications of causality and exogeneity were developed in weakly-stationary systems, where it can be difficult to test for their presence or absence. All modern macro-economies are non-stationary from both stochastic and deterministic changes, often modelled as integrated-cointegrated systems prone to intermittent structural breaks—sudden large changes, invariably unanticipated. In such non-stationary data generation processes (DGPs), many results concerning

*Financial support from the UK Economic and Social Research Council under grants RES-015-27-0035 and RES-000-23-0539 is gratefully acknowledged. I am indebted for helpful comments on an earlier draft to Nancy Cartwright, Mike Clements, Clive Granger, Michael Massmann, Grayham Mizon, Adrian Pagan, Franz Palm, Julian Reiss and Elliott Sober; to participants at the 2001 EC^2 meeting on 'Causality and Exogeneity in Econometrics' at Louvain-La-Neuve; and to four anonymous referees.

modelling, forecasting, and policy change substantially relative to the stationary case. The primary difficult becomes that of establishing time invariant relationships, but once that is achieved, the roles of causality and exogeneity are often more easily ascertained.

The precursor to the approach here lies in the theory of forecasting developed in Clements and Hendry (1999, 2002b) and reviewed in Hendry and Clements (2003). Those authors are concerned with the historical prevalence of forecast failure, defined as a significant deterioration in forecast performance relative to the anticipated outcome (usually based on the historical performance of a model): for documentation see (e.g.) Stock and Watson (1996) and Clements and Hendry (2001b). The forecast-error taxonomies established by the latter authors show that the main sources of such forecast failures are changes in coefficients of deterministic terms, or location shifts. Structural breaks affecting mean-zero variables pose fewer difficulties (see e.g. Hendry, 2000b), as do mis-specification and mis-estimation when breaks do not occur. Such results entail reconsidering how causality and exogeneity apply when empirical models are mis-specified for complicated, evolving and non-stationary DGPs. For example, since one cannot prove that causal models need dominate in forecasting, neither forecast success nor failure is informative about causal links. Nevertheless, in other settings, non-constancy can help determine causal direction.

The structure of the chapter is as follows. Section 2 describes the econometric background, noting some of the many salient contributions to the literature. Then section 3 reviews the three well known forms of weak, strong and super exogeneity in relation to inference, forecasting, and policy, and notes their role in non-stationary systems. Analyzing exogeneity first is convenient for section 4, which then provides a similar approach for causality, before briefly discussing Granger causality, following the overview in Hendry and Mizon (1999). Section 5 notes some links between exogeneity and causality. Section 6 examines their role in forecasting, and section 7 discusses the roles of exogeneity and causality in policy analyses. Section 8 concludes.

2 Background

Let $(\Omega, \mathcal{F}, \mathsf{P}(\cdot))$ denote the probability space supporting a vector of m discrete-time, real random variables $\mathbf{w}_t$, with sample space Ω and event space $\mathcal{F}_{t-1}$ at each time $t \in \mathcal{T}$. The joint density $\mathsf{D}_{\mathsf{w}_t}(\mathbf{w}_t|\mathcal{F}_{t-1}, \boldsymbol{\lambda})$ for $\boldsymbol{\lambda} \in \Lambda \subseteq \mathbb{R}^q$ is the data generation process (DGP) of the economy under analysis. The q-dimensional parameter (or 'index') $\boldsymbol{\lambda}$ does not depend on $\mathcal{F}_{t-1}$ at any t, but need not be constant over time since $\boldsymbol{\lambda}$ is determined by economic agents' decisions. The history of the stochastic process $\{\mathbf{w}_t\}$ up to time $(t-1)$ is denoted by $\mathbf{W}_{t-1} = (\mathbf{W}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{t-1}) = (\mathbf{W}_0, \mathbf{W}^1_{t-1})$, where $\mathbf{W}_0$ is the set of initial conditions. For a sample period $t = 1, \ldots, T$, the DGP becomes $\mathsf{D}_{\mathsf{W}}(\mathbf{W}^1_T|\mathbf{W}_0, \boldsymbol{\lambda})$, and is sequentially factorized as:

$$\mathsf{D}_{\mathsf{W}}\left(\mathbf{W}^1_T \mid \mathbf{W}_0, \boldsymbol{\lambda}\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{w}_t}\left(\mathbf{w}_t \mid \mathbf{W}_{t-1}, \boldsymbol{\delta}_t\right) \tag{1}$$

where $\mathbf{g}_\delta(\boldsymbol{\lambda}) = (\boldsymbol{\delta}_1 \ldots \boldsymbol{\delta}_T)$ for a 1–1 function $\mathbf{g}_\delta(\cdot)$. While the notation potentially allows the parameters of (1) to shift every period, we assume that such changes are in practice intermittent. Historical sources of non-constancy are regime shifts (changes in policy), and technological, legislative, behavioural and institutional structural breaks (shifts in the parameters of the system). Since economic processes also evolve for many reasons, the $\{\mathbf{w}_t\}$ are likely to be integrated of at least first order (denoted $\mathsf{I}(1)$). Thus, a key feature of (1) is the non-stationarity of $\{\Delta\mathbf{w}_t\}$ (or of $\left\{\Delta^2\mathbf{w}_t\right\}$ if the level is

$\mathsf{I}(2)$), with changing first and second data moments. Because $\mathsf{D}_{\mathsf{w}_t}(\cdot)$ alters over time, all expectations operators have to be time dated, as in $\mathsf{E}_t[\mathbf{w}_t]$ and $\mathsf{V}_t[\mathbf{w}_t\mathbf{w}_t']$ for unconditional expectations and variances respectively. Such dating is separate from the timing of any conditioning information, such as $\mathbf{W}_{t-1}$. While $\mathsf{D}_{\mathsf{w}_t}(\mathbf{w}_t|\mathbf{W}_{t-1},\boldsymbol{\delta}_t)$ characterizes the generation of the outcomes, measurement errors are bound to add a further layer of complexity to empirical analyses (especially of causality), but as they are not germane to the concepts *per se*, the $\{\mathbf{w}_t\}$ are assumed to be observed.

2.1 Local DGP

Reductions from the DGP for $\{\mathbf{w}_t\}$ to that of the transformed subset of $n < m$ variables $\{\mathbf{x}_t\}$ to be analyzed can radically alter the causality and exogeneity status of variables. In general (see, *inter alia*, Hendry, 1995a, and Mizon, 1995), there exists a 'local DGP' $\mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\gamma}_t)$ with $\boldsymbol{\gamma}_t \in \Gamma \subseteq \mathbb{R}^\ell$ derived from $\mathsf{D}_{\mathsf{w}_t}(\mathbf{w}_t|\cdot)$ by aggregation, transformation, and marginalization with respect to all excluded current-dated and lagged variables. Map from $\mathbf{w}_t$ by an information-preserving transform $\mathbf{h}(\cdot)$ such that $\mathbf{h}(\mathbf{w}_t) = (\mathbf{x}_t', \boldsymbol{\eta}_t')'$. Then $\boldsymbol{\delta}_t$ is transformed to $(\boldsymbol{\rho}_t', \boldsymbol{\gamma}_t')$, such that:

$$\mathsf{D}_{\mathsf{w}_t}(\mathbf{w}_t \mid \mathbf{W}_{t-1},\boldsymbol{\delta}_t) = \mathsf{D}_{\eta_t|\mathsf{x}_t}(\boldsymbol{\eta}_t \mid \mathbf{x}_t, \mathbf{W}_{t-1},\boldsymbol{\rho}_t)\,\mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t \mid \mathbf{W}_{t-1},\boldsymbol{\gamma}_t). \tag{2}$$

To restrict attention to the marginal model, $\mathsf{D}_{\mathsf{x}_t}(\cdot)$, without loss of information requires that:[1]

$$\mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t \mid \mathbf{W}_{t-1},\boldsymbol{\gamma}_t) = \mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t \mid \mathbf{X}_{t-1},\boldsymbol{\gamma}_t), \tag{3}$$

so that lagged $\boldsymbol{\eta}$ must be irrelevant. Finally, the $r \leq \ell$ parameters of interest, $\boldsymbol{\mu}$, must be a function of $\boldsymbol{\gamma}_t$ alone, so $\boldsymbol{\rho}_t$ does not provide information about $\boldsymbol{\gamma}_t$. For example, if important explanatory variables are omitted from (3), and $\boldsymbol{\rho}_t$ is non-constant, so will be the coefficients in (3) even when $\boldsymbol{\gamma}$ in (2) is constant (see e.g., Hendry and Doornik, 1997). Such parametric links are intrinsic to (2), and the usual assumption that the parameters in models thereof are variation free, so $(\boldsymbol{\gamma}_t, \boldsymbol{\rho}_t) \in \Gamma \times \mathcal{R}$, is insufficient. When (3) holds, it is fully informative about the roles of the variables; but in general, there will be a loss of information from the elimination of $\{\boldsymbol{\eta}_t\}$, with a consequential (unintended) transformation of the parameters.

2.2 The econometric model

The econometric model $\mathsf{f}_{\mathsf{x}}(\cdot)$ for $\mathbf{x}_t$ is denoted by:

$$\mathsf{f}_{\mathsf{X}}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = \prod_{t=1}^{T} \mathsf{f}_{\mathsf{x}}(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}) \quad \text{where} \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k \tag{4}$$

when $\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ is the postulated sequential joint density at time t. We assume that $k < \ell$ and that $\boldsymbol{\theta}$ represents the constant (meta)parameters postulated by the model builder. Inevitably, $\mathsf{f}_{\mathsf{x}}(\cdot) \neq \mathsf{D}_{\mathsf{x}_t}(\cdot)$, which may affect inferences about exogeneity or causality. An econometric model is congruent if it matches the data evidence in all relevant directions, and Bontemps and Mizon (2003) show that such a model would encompass the local DGP (denoted LDGP). Further, a successful analysis requires that $\boldsymbol{\mu} = \mathbf{g}_\mu(\boldsymbol{\theta})$, and while achieving that goal is difficult, progressive research can glean important insights even in processes that are not time invariant.

[1] We ignore any complications related to the role of the initial conditions, $\mathbf{W}_0$.

3 Exogeneity

The notion of exogeneity, or synonyms thereof, in relation to econometric modelling dates back to the origins of the discipline (see e.g., Morgan, 1990, and Hendry and Morgan, 1995), including important contributions by Koopmans (1950) and Phillips (1957). Here we follow the approach in Richard (1980), formalized by Engle, Hendry and Richard (1983): Ericsson (1992) provides an exposition.

We assume that $\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ in (4) is a valid reduction to a congruent model of the LDGP, and partition $\mathbf{x}_t' = (\mathbf{y}_t' : \mathbf{z}_t')$ where $\mathbf{y}_t$ is $n_1 \times 1$ and $\mathbf{z}_t$ is $n_2 \times 1$ with $n = n_1 + n_2$, with $\mathbf{X}_{t-1}$ partitioned accordingly as $(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-1})$. To model $\mathbf{y}_t$ conditional on $\mathbf{z}_t$, factorize $\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t|\mathbf{X}_{t-1},\boldsymbol{\theta})$ into a conditional and a marginal density, transforming $\boldsymbol{\theta} \in \Theta$ to $\boldsymbol{\psi} \in \Psi$:

$$\boldsymbol{\psi} = \mathbf{g}_{\psi}(\boldsymbol{\theta}) \quad \text{where} \quad \boldsymbol{\psi} \in \Psi \quad \text{and} \quad \boldsymbol{\theta} \in \Theta, \tag{5}$$

such that $\mathbf{g}_{\psi}(\cdot)$ is a 1–1 reparametrization which sustains the partition $\boldsymbol{\psi}' = (\boldsymbol{\psi}_1' : \boldsymbol{\psi}_2')$, when $\boldsymbol{\psi}_i$ has k_i elements $(k_1 + k_2 = k)$, and:

$$\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t \mid \mathbf{X}_{t-1},\boldsymbol{\theta}) = \mathsf{f}_{\mathsf{y}|\mathsf{z}}(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}, \boldsymbol{\psi}_1)\, \mathsf{f}_{\mathsf{z}}(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\psi}_2). \tag{6}$$

As only $\mathsf{f}_{\mathsf{y}|\mathsf{z}}(\mathbf{y}_t|\mathbf{z}_t,\mathbf{X}_{t-1},\boldsymbol{\psi}_1)$ is to be retained, the analysis of $\boldsymbol{\mu}$ is without loss of information relative to (4) when $\mathbf{z}_t$ is weakly exogenous for $\boldsymbol{\mu}$, namely $\boldsymbol{\mu} = \mathbf{f}(\boldsymbol{\psi}_1)$ alone and $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \in \Psi_1 \times \Psi_2$. Thus exogeneity is not a property of a variable: in (3), all variables are endogenous. Moreover, weak exogeneity is just as relevant to instrumental variables estimation, as the marginal density of $\mathbf{z}_t$ then relates to the distribution of the putative instruments: merely asserting orthogonality is inadequate, as illustrated by counter examples in Hendry (1995a, 1995c).

Strong exogeneity also requires the absence of feed back from lagged $\mathbf{y}_{t-i}$ onto $\mathbf{z}_t$ so that:

$$\prod_{t=1}^{T} \mathsf{f}_{\mathsf{z}}(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\psi}_2) = \mathsf{f}_{\mathsf{Z}}\left(\mathbf{Z}_T^1 \mid \mathbf{X}_0, \boldsymbol{\psi}_2\right). \tag{7}$$

Finally, super exogeneity requires weak exogeneity and the invariance of $\boldsymbol{\mu}$ to changes in $\boldsymbol{\psi}_2$. As is well known, the three concepts respectively sustain conditional inference; conditional multi-step forecasting;[2] and conditional policy analysis (when components of $\mathbf{z}_t$ are policy instruments).

The absence of links between the parameter spaces Ψ_1 and Ψ_2 is not purely a matter of model specification, dependent on the parameters of interest, and at the choice of the investigator. In some settings, as neatly illustrated by Ericsson (1992), changing the parameters of interest can deliver or lose weak exogeneity. But in other settings, the LDGP determines the actual, as opposed to the claimed, exogeneity status. Factorize (3) as:

$$\mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\gamma}_t) = \mathsf{D}_{\mathsf{y}_t|\mathsf{z}_t}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}, \boldsymbol{\phi}_{1,t}\right) \mathsf{D}_{\mathsf{z}_t}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}_{2,t}\right) \tag{8}$$

When $\boldsymbol{\mu}$ enters both $\boldsymbol{\phi}_{1,t}$ and $\boldsymbol{\phi}_{2,t}$ in (8), inference can be distorted if (6) falsely asserts weak exogeneity: see e.g., Phillips and Loretan (1991). The consequences of failures of weak exogeneity can vary from a loss of estimation efficiency through to a loss of parameter constancy, depending on the source of the problem: see Hendry (1995a, ch. 5). We now illustrate both extreme cases.

First, consider an experimental setting where the Gauss–Markov conditions might appear to be satisfied so that:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{with} \quad \boldsymbol{\epsilon} \sim \mathsf{N}_T\left[\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right], \tag{9}$$

[2] Projecting $\mathbf{y}_{T+h}$ onto $\mathbf{x}_{T-i}$ also circumvents any feedback problem: see e.g., Bhansali (2002).

when $\mathbf{Z}' = (\mathbf{z}_1 \dots \mathbf{z}_T)$ is a $T \times k$ matrix, rank$(\mathbf{Z}) = k$, and $\boldsymbol{\epsilon}' = (\epsilon_1 \dots \epsilon_T)$, with:

$$\mathsf{E}[\mathbf{y} \mid \mathbf{Z}] = \mathbf{Z}\boldsymbol{\beta},$$

and hence $\mathsf{E}[\mathbf{Z}'\boldsymbol{\epsilon}] = \mathbf{0}$. Nevertheless, OLS need not be the most efficient unbiased estimator of $\boldsymbol{\beta}$. An explicit weak exogeneity condition is required when $\mathbf{Z}$ is stochastic, such that $\boldsymbol{\beta}$ cannot be learned from its marginal distribution. Otherwise, an unbiased estimator which is an affine function of $\mathbf{y}$ can dominate, possibly dramatically: see Hendry (2003). Secondly, if $\mathbf{z}_t$ is simultaneously determined with $\mathbf{y}_t$, yet experiences a location shift, then a conditional model of $\mathbf{y}_t$ given $\mathbf{z}_t$ (incorrectly treating $\mathbf{z}_t$ as weakly exogenous) will have non-constant parameters even though the DGP equation for $\mathbf{y}_t$ is constant.

Although it can be difficult to test exogeneity claims when unconditional second moments are constant over time, as Engle *et al.* (1983) note, despite only one specification corresponding to reality, both main forms of non-stationarity (integrability and breaks) alter the analysis as we now show.

First, cointegrated systems provide a major forum for testing one aspect of exogeneity: see *inter alia*, Hunter (1992a, 1992b), Urbain (1992), Johansen (1992), Dolado (1992), Boswijk (1992), and Paruolo and Rahbek (1999). Equilibrium-correction mechanisms which cross-link equations violate long-run weak exogeneity, confirming that weak exogeneity cannot necessarily be obtained merely by choosing the ‘parameters of interest’. Conversely, the presence of a given disequilibrium term in more than one equation is testable.

Secondly, processes subject to structural breaks sustain tests for super exogeneity and the Lucas (1976) critique: see e.g., Hendry (1988), Fischer (1989), Favero and Hendry (1992), and Engle and Hendry (1993). When conditional models are constant despite data moments changing considerably, there is *prima facie* evidence of super exogeneity for that model’s parameters; and if the model as formulated does not have constant parameters, resolving that failure ought to take precedence over issues of exogeneity. However, while such tests are powerful for location shifts, changes to ‘reaction parameters’ of mean-zero stochastic variables are difficult to detect: see (e.g.) Hendry (2000b).

Richard (1980) considers changes in causal direction, reconsidered in section 5.1 in relation to determining ‘causal direction’ through a lack of time invariance.

4 Causality

The concept of causality has intrigued philosophers for millennia, so section 4.1 provides a brief incursion into the philosophy of causality. Section 4.2 looks at notions of causality in economic discourse, then sub-section 4.3 considers causal inference in observational disciplines. Both statisticians and economists have attempted to test for its existence, as it is particularly important if one is interested in testing theories or conducting policy analysis: see Simon (1952), Zellner (1979), Hoover (1990) and Hendry (1995a) for general discussions of causality and causal inference in econometrics; and e.g., Holland (1986), Cox (1992) and Lauritzen and Richardson (2002) in statistics. Finally, sub-section 4.4 evaluates the concept of Granger causality, and its empirical implementation. Section 5 considers links between causality and exogeneity.

4.1 Philosophy of causality

As remarked in Hendry (1995a), ‘causality’ is a philosophical mine field – even without the complexities introduced by quantum physics, one-off events, anticipations, ‘simultaneity’, and the issue that (e.g.)

changes may enlarge or restrict the opportunity sets of agents who could choose to do otherwise than just react mechanically. Thus, we will step as lightly as feasible: the increase in the prevalence of footnotes signals the difficulty. Fortunately, Hoover (2001) provides an excellent discussion. In particular, he notes that many analyses conflate the concept with inference about it, which he calls 'the epistemic fallacy', essentially confusing truth with its criterion. We will first consider the concept, briefly comment on what makes something a cause, then turn to how one might ascertain causal relationships.[3]

What makes a cause just that? A cause is a quantitative process that induces changes over time, mediated within a structure. In turn, a structure is an entity that remains invariant to interventions 'and directly characterizes the relationships involved' (i.e., corresponds to reality): see Hendry (1995a, ch. 2).[4] The relation between cause and effect is asymmetric: the latter does not induce the former. The 'causal field' or structure is central, as changes to that structure can alter what causes what.

This notion of cause is consistent with at least some earlier formulations, such as Simon (1957, chs. 1, 3) who formalizes causal order as an asymmetric relation invariant under interventions to the 'basic' parameters of the system; Cartwright (1989) who discusses 'causal capacities', such that causes out need causes in (i.e., we need to postulate that causal links already exist); Hoover (1990) who argues for invariance under interventions; and Hoover (2001) where cause is an asymmetric relationship with 'unconnected parameters' within a causal structure that is a feature of reality.[5]

Our everyday thinking is replete with causal assertions: the car stopped because the driver braked; the atom disintegrated because it was hit by the proton; the pressure of the gas rose because heat was applied; output rose because interest rates fell;... At one level, such statements seem unexceptional. Consider the first example: the brakes were first applied; the braking system remained intact; the harder the brakes were applied, the more the car slowed; the brakes stopped the car (not the car slowing 'caused' the brakes to come on). If the brake cable were cut (as in some murder stories), or the brake fluid was too low (as happens in reality) and so on, pressing the brake pedal would achieve little: that system is not invariant to such 'interventions'. More fancifully, the car might be attached to a cable, the tightening of which 'actually causes' it to slow, so the braking is incidental – causal discussion is often based on thought experiments involving counterfactuals. More realistically, it could be argued that the driver caused the car to stop, or even that the trigger for the driver pressing the brakes was the cause... 'Causal chains' may have many steps, with the 'ultimate causes' possibly hidden from human knowledge.

Consequently, it is difficult to independently specify what makes something a cause. Given the key role of the 'causal structure', causal connections need be neither necessary nor sufficient. The former (necessity) implies that if A is to cause B, then not-A entails not-B; the latter (sufficiency) that not-B entails not-A. Counter examples to necessity arise when there are multiple possible causes, as in medicine – smoking causes lung cancer, but so might environmental pollution. Equally, sufficiency

[3]This organization partially follows Hoover's discussion of Hume, who distinguished between the conceptual, ontological, epistemological, and pragmatic aspects of causality. Hoover cites Hume as complaining that his efforts at clarification 'heated [his] brain', a view with which I sympathize.

[4]The perceptive reader will not miss the problem: it is impossible to define systems that are invariant to all interventions – else nothing would change. While that is a view of the world which Parmenides might have endorsed, Heraclitus would have disagreed, since he regarded flux as so intense one cannot even step into the same river twice (on both views, see Gottlieb, 2000). Neither of theses settings is conducive to causal inference – but then neither seems realistic. However, Nancy Cartwright has correctly noted that my definition is in the nature of a sufficient condition to discern a cause, rather than defining cause *per se*. For example, a rope can cause an object to which it is attached not to move.

[5]Conversely, the notion of a 'causal chain', in (say) Strotz and Wold (1960), was shown to be insufficient to characterize exogeneity by Engle *et al.* (1983) precisely because it failed to specify the absence of links between parameters.

crumbles if (say) whether or not lung cancer occurs in any given smoker may depend on the quality of their 'DNA repair kit': a really good 'kit' may continually repair the incipient damage so the disease never appears. Such examples reinforce that the 'causal field' characterizing the system and its background conditions are central: economists are anyway used to the idea that non-stationarities and non-linearities may mean that a cause sometimes has an effect, and sometimes does not, depending on thresholds. But these are difficulties for inference about causal links, not about causality as a concept.

It can be no surprise that ascertaining actual causes is difficult. My countryman, David Hume (1758) tends to be the source of much thinking about the topic.[6] Hume asserted that we cannot *know* necessary connections in reality (my italics). The key word is know, for therein lies the difficulty: causal inference is uncertain (although, as Hoover, 2001, notes, this did not stop Hume from arguing in 'causal' terms about economics). Just as we cannot have a criterion for truth, yet science seeks for 'truth', so with causal models, we cannot have a criterion for knowing that the necessary connections have been ascertained. Hoover himself remarks (p24) that "our knowledge of reality consists of empirically based conjectures and is necessarily corrigible". Even greater problems arise when inferring anything from probability relations. Scientists – and economists – often formulate theory models with theory causal connections between variables, test those models empirically, and if they are not rejected, use the theory connections as if they held in reality: the 'causes in' come from the theory, and the 'causes out' from the evidence not rejecting that theory—rather a weak basis. Policy changes often provide the backdrop for such testing, and play an important role below, so we turn to economics.

4.2 Notions of causality in economics

Economists often conceptualize cause as a well-defined dependence in a theoretical model, such as $y = f(z)$, implying that a change in z will induce a change in y: most recently, see Heckman (2000). That approach suffers from four drawbacks. First, it is essentially circular, since a model would presumably not be 'well defined' if the relation in question did not sustain the conclusion. Secondly, some form of asymmetry needs to be independently added so that (e.g.) when $f(\cdot)$ is invertible, y does not then cause z. Thirdly, a theory can be well defined but irrelevant to the behaviour of the actual economy, whereas the objective of most economic analyses is doubtless to infer some properties of reality. Finally, a theory-dependent definition is insufficient to characterize all the important examples of causal links: for example, 'aspirin' clearly caused the removal of headaches even before the key ingredient (acetylsalicylic acid) was isolated from willow-tree bark, yet there was no 'well-defined model' of how it worked till very recently (see Weissmann, 1991). In any case, the 'superlative' theories of the natural sciences represent the outcome of close interaction between successive theories and accumulating evidence, sometimes reinterpreted retrospectively (see e.g., Penrose, 1989), not pure 'a priori' postulates as to causal links. Thus, a more general formulation is needed.

The key point is that causality is a property of reality not of a model: either z does or does not cause y, so causality is not usefully defined purely in a theory context. Conversely, commencing from observational evidence alone raises that frightening spectre of 'inferring causes from correlations' which worries so many economists, and is the subject of section 4.3. Rather, in economics, theoretical implications are tested against the observed correlations. Although corroboration is clearly insufficient, and rejection is not definitive, knowledge about possible causal links can accumulate by progressive

[6] Bibliographical references to him are also a bit of minefield, as Hume often reissued essays or parts thereof under different titles and editions.

research, as has happened in other sciences.

An issue that is not necessarily problematic for the conceptualization of causality, but is for practical inference, is that action in anticipation of a change can mask the 'true' effect. For example, if the Bank of England is expected to cut its interest rate, futures markets may adjust in anticipation, so nothing happens when the cut actually occurs. Indeed, some 'causality' tests might infer that the market 'caused' the Bank to react. Care is required here: it is the anticipation which does the causing, not the later outcome that was anticipated. The timing effect is correct but hidden, and in (say) tests of Granger causality (see section 4.4) might incorrectly suggest the reversal of the actual causal direction.

Potentially immense problems and benefits are raised by non-stationarity. At one extreme, causal relations themselves may be transient and render inference completely unreliable. Alternatively, by 'excitation' of the data space, considerable advantages could accrue, by refuting false attributions and providing powerful evidence in favour of valid theories. We resume this line in sections 5.1 and 7.1.

4.3 Causal inference in observational sciences

It may be thought that non-experimental research is at a severe disadvantage for causal inference compared to experimental (see Blalock, 1961, and Wold, 1969, on causal inference in non-experimental research). While it is undoubtedly true in many cases that an efficacious, well-designed and carefully controlled experiment can deliver important insights, experimental evidence may also mislead (see Cartwright, 1983, for a general critical appraisal of physical 'laws' mainly derived from experiments). Doll (2001) recounts the fascinating tale of how cigarette smoking was implicated in lung cancer. Initial suggestions in the 1920s of a possible connection (based on the steady rise in both) prompted a series of laboratory 'tests' on animals, which failed to find any link. When Doll commenced his own observational study, he suspected that the widespread use of 'tar' as a road surfacing material might be to blame, and focused on testing that idea. Fortunately, he also collected data on other aspects of the sample of hospital patients investigated. Tabulating their smoking prevalence against lung cancer incidence revealed a dramatically strong connection. Later experiments confirmed his finding. In retrospect, we can understand why the early experiments had not found the link: there is a long latency between smoking and contracting the disease, and the first experiments had not persevered for sufficiently long, whereas later investigators were more persistent. Doll provides several other examples where theoretical or experimental analyses failed to suggest links whereas careful study of the observational data revealed the causal connections.[7]

Economists, of course, have long been aware of both spurious (see Yule, 1897) and nonsense (see Yule, 1926) correlations. Indeed, the claim that causal inference is hazardous from observational data alone is buttressed by issues of observational equivalence (see e.g., Basmann, 1988), lack of identification, and model mis-specification (see e.g., Lutkepohl, 1982). However, the potential dangers for empirical research from these sources may have been over-emphasized: it is here that non-stationarities due to DGPs suffering structural breaks begin to offer dividends, and not just complications. In a world subject to intermittent large shifts, observational equivalence and lack of unique identification seem most unlikely: the problem is finding any autonomous relations, not a plethora of them. Model mis-specification is manifestly ubiquitous, and poses problems for all aspects of inference, not just about

[7]Doll also highlights the possible dangers in such inferences, citing the close observational link between cirrhosis of the liver and smoking, which we now know is intermediated by alcohol (a higher proportion of smokers drink heavily than non-smokers).

causality. However, enhanced data variation can help to diagnose mis-specification in a progressive research strategy, namely an approach in which knowledge is gradually accumulated as codified, reproducible, information about the world without needing complete prior information as to its nature (see e.g., Hendry, 1995a, ch. 8, and Hendry, 2000a, ch. 8). For example, location shifts in policy variables that do not induce forecast failure in relationships linking them to targets provide strong evidence of a causal link: see section 7.

4.4 Granger causality

Most economic relations are inexact, so stochastic formulations are inevitable, leading to both statistical testing and, more fundamentally, to the idea of evaluating causality in terms of changes to the joint distributions of the observables. Granger (1969, 1980, 1988b) has been the most forceful advocate of such an approach: also see Chamberlain (1982), Newbold (1982), Geweke (1984), Phillips (1988), Florens and Mouchart (1982, 1985), Mosconi and Giannini (1992), and Toda and Phillips (1993, 1994). Granger (1969) provided a precise definition of his concept, although most later investigators followed a different route. The fundamental basis for such tests is that causes contain 'special information' about effects not contained elsewhere in the information set; and that 'time's arrow' is unidirectional – only the past can cause the present, the future cannot. Anticipations of future events can influence present outcomes, but absent a crystal ball, those must be functions of available information. Thus, 'We inhabit a world in which the future promises endless possibilities and the past lies irretrievably behind us.' (Coveney and Highfield, 1990, p297, who base the direction of time on increasing entropy).

Specifically, Granger (1969) proposed that if in the universe of information, deleting the history of one set of variables does not alter the joint distribution of any other variables, then the omitted variables do not to cause the others: we refer to this property as Granger non-causality (denoted GNC). Conversely, knowing the cause should help forecast the future.

(3) above required that $\boldsymbol{\eta}$ did not Granger cause $\mathbf{x}$. There, $\mathsf{D}_{\mathsf{x}_t}\left(\mathbf{x}_t \mid \mathbf{W}_{t-1}, \boldsymbol{\gamma}_t\right)$ defined the 'causal structure' which directly characterizes the relationships (i.e., it is the DGP of $\mathbf{x}_t$); the causality takes place in time, and is asymmetric (without precluding that the opposite direction may also hold); and the past induces changes in other variables. Granger causality, therefore, has many attributes in common with the earlier characterization of cause. However, the issue of invariance is not resolved, Granger's definition of causality does not explicitly involve parameters (see e.g., Engle *et al.*, 1983, Buiter, 1984, and Hoover, 2001), and contemporaneous links are eschewed. Consequently, Granger causality does not seem to completely characterize the notion of 'cause'. The 'anticipations' example in section 4.2 is sometimes cited as a counter example, but does not in fact violate the concept, and would anyway confuse inference procedures for most definitions of causality unless the entire mechanism was known.

4.4.1 Empirical Granger non-causality

Perhaps most importantly, the definition of GNC is non-operational as it relates to the *universe* of information. Empirical tests for Granger causality are usually based on reductions within $\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t|\mathbf{X}_{t-1}, \cdot)$, often without testing its congruence. Hendry and Mizon (1999) defined empirical Granger non-causality (EGNC) with respect to the information set generated by $\{\mathbf{x}_t\}$ as follows. If the density $\mathsf{f}_{\mathsf{z}}\left(\cdot\right)$ does not depend on $\mathbf{Y}_{t-1}$, so that:

$$\mathsf{f}_{\mathsf{z}}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}, \cdot\right) = \mathsf{f}_{\mathsf{z}}\left(\mathbf{z}_t \mid \mathbf{Z}_{t-1}, \mathbf{X}_0, \cdot\right), \tag{10}$$

then $\mathbf{y}$ does not empirically Granger-cause $\mathbf{z}$. Since causality in the LDGP need not entail that in the DGP, and vice versa, it cannot be a surprise that the same is true of EGNC. Among the drawbacks of EGNC, Hendry and Mizon (1999) list:

- the presence of EGNC in a model does not entail the existence of GNC in the DGP;
- the existence of GNC in the DGP need not entail the presence of EGNC in a model (see e.g., Hendry, 1997);
- the existence of GNC is specific to each point in time, but detecting EGNC requires extended sample periods.

Any or all of these drawbacks can seriously confound empirical tests. The first two are important results for interpreting empirical modelling, and are almost certainly difficult to disentangle in stationary processes. However, concerning the third, if parameter change is a feature of reality, that is hardly the fault of Granger's conceptualization. In any case, location shifts in subsets of the variables should help clarify what the genuine links are – bringing Granger causality close to the concept in section 4.1. Hendry and Mizon (1999) also show that EGNC plays a pervasive role in econometrics, irrespective of whether or not there exist 'genuine DGP causes'. They illustrate their claim for ten areas of econometric modelling, namely marginalizing; conditioning; distributions of estimators and tests; inference via simulation; cointegration; encompassing; forecasting; policy analysis; dynamic simulation; and impulse response analysis. As their paper is recent, we skip the details, although several issues recur below.

5 Links of causality and exogeneity

Exogeneity is neither necessary nor sufficient for causality, even in the DGP. A variable can be exogenous for the parameters of interest in a given system, but irrelevant – i.e., have a zero coefficient – so sufficiency fails. Equally, a causal variable can operate through parameters linked such that exogeneity is violated, so necessity fails as well. Engle *et al.* (1983) noted that variables may be absent from the conditional density, yet not be weakly exogenous, and also showed that Granger non-causality is neither necessary nor sufficient for weak exogeneity, although some authors have mistakenly viewed GNC as sufficient for variables to be 'exogenous' (see e.g., Sims, 1972, and Geweke, 1984). Empirically, matters are even less clear, since variables may act as proxies, induced by invalid reductions, rather than genuine 'forces'.

None of these results entails that causality and exogeneity are unconnected as shown by their interacting roles in strong and super exogeneity respectively. The former was discussed above, but we have one aspect of the latter still to explore.

5.1 Super exogeneity and causality

Super exogeneity augments weak exogeneity by the requirement that the parameters of interest be invariant to shifts in the parameters of the marginal distribution. Such a condition is far stronger than 'variation freeness', and is more like the condition of 'independent parameters' used in the causality formulation of Hoover (2001). If causality corresponded to a regular response that was invariant under interventions, then super exogeneity would embody several of the key elements of causality. Simon (1952, 1953) used the invariance of a relationship under interventions to an input variable as an operational notion of cause, as did Hoover (1990).

Consider the conditional model:

$$\mathsf{D}_{\mathsf{y}_t|\mathsf{z}_t}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}, \boldsymbol{\psi}_1\right) \tag{11}$$

when there is a non-zero dependence of $\mathbf{y}_t$ on $\mathbf{z}_t$:

$$\frac{\partial \mathbf{y}_t}{\partial \mathbf{z}_t'} \neq \mathbf{0}, \tag{12}$$

and the parameters of interest are $\boldsymbol{\mu} = \mathbf{f}_\mu\left(\boldsymbol{\psi}_1\right)$. When $\mathbf{z}_t$ is super exogenous for $\boldsymbol{\mu}$, so $\boldsymbol{\psi}_2$ in (7) can change without altering the conditional relationship between $\mathbf{y}_t$ and $\mathbf{z}_t$, then many of the ingredients for a cause are satisfied: an asymmetric quantitative process, (12), that induces changes over time, mediated within the structure $\mathsf{D}_{\mathsf{y}_t|\mathsf{z}_t}(\cdot)$. Thus, $\mathbf{z}_t$ causes the resultant change in $\mathbf{y}_t$, and the response of $\mathbf{y}_t$ to $\mathbf{z}_t$ remains the same for different sequences $\{\mathbf{z}_t\}$. Our ability to detect changes in $\boldsymbol{\psi}_1$ in response to shifts in $\boldsymbol{\psi}_2$ depends on the magnitude of the changes in the latter: large or frequent changes in $\boldsymbol{\psi}_2$ that left (11) invariant would provide strong evidence of a causal link which could sustain policy changes when (e.g.) $\mathbf{z}_t$ was under government control.

However, economic policy usually depends on disequilibria in the rest of the economy (such as excess demands) which would appear to interlink private sector and policy parameters, thereby violating weak exogeneity of $\mathbf{z}_t$ for $\boldsymbol{\psi}_1$ as required for super exogeneity. Nevertheless, as Hendry and Mizon (2000) argue, reliable estimates of policy responses can be obtained when $\mathbf{z}_t$ is not weakly exogenous provided the parameters to be shifted by the policy agency are not functions of $\boldsymbol{\psi}_1$: for example, cointegration relations are often the basis of estimated disequilibria, and are usually established at the level of the complete system, whereas their parameters are not generally subject to policy interventions (which might explain the problems arising in cases when they are, as – say – in the transition of economies from controlled to free markets).

A similar argument would seem to hold when nature creates the ‘experimental design’. For example, Hume’s ‘causal’ analysis of inflation in response to a major gold discovery assumes invariant relations to propagate the shock.

5.1.1 Determining causal direction

The appendix to Engle and Hendry (1993) in Ericsson and Irons (1994) shows that if a given conditional model has both invariant parameters and invariant error variances across regimes, whereas the joint process varies across those regimes, then the reverse regression cannot have invariant parameters. Thus, such a reverse conditioning should fail either or both constancy and invariance tests, precluding its interpretation as a causal link, and leaving at most one possibility. This notion is echoed in the analysis of interventions in relation to interpreting causality in directed acyclic graphs by Lauritzen and Richardson (2002). Importantly, a break in a model for a subset of variables will persist even under extensions of the information set: for example, if an unmodelled jump in CPI inflation is later ‘explained’ by a jump in oil prices, the latter now reflects the break in the model: see e.g. Hendry (1988).[8]

[8]Of course, apparent breaks may be the result of an unmodelled non-linearity: slipping down a gentle bank then over a cliff is a non-linear effect, but the problem is the break at the end...

6 Causality implications in forecasting

Both senses of causality (i.e., from sections 4.1 and 4.4) play a role in forecasting, but perhaps not in ways that might be expected.

First, in DGPs subject to location shifts, one cannot prove the dominance of causal variables in forecasting: counter examples are provided in (e.g.) Clements and Hendry (1999). Consequently, models based on variables that do not enter the DGP can outperform in forecasting over models using every variable that was causal in the DGP prior to the forecast period. More generally, the forecasting theory underlying that result has achieved a range of successful explanations of otherwise puzzling features of the empirical forecasting literature, including (see Hendry and Clements, 2003):

- why intercept corrections have value added;
- why extrapolative devices can win forecasting competitions;
- why simplicity does not account for that outcome, but robustness to location shifts does;
- why forecast pooling can dominate the best individual forecasting device; and:
- why forecast failure occurs, but is not informative about the presence or absence of causal links.

Fildes and Makridakis (1995) and Makridakis and Hibon (2000) stress the discrepancy between theory and empirical findings in forecasting in general. Clements and Hendry (2001a) show how their theory can help close that gap. Moreover, they also account for the ‘principles’ based on empirical econometric forecasting performance enunciated by Allen and Fildes (2001), who find that admissible reductions of VARs with relatively generous lag specifications, estimated by least squares, and tested for constant parameters do best on average, even though congruent models need not outperform non-congruent. Such findings stand in marked contrast to what can be established when the economy is representable as a stationary stochastic process (with unconditional moments which are constant over time): well-tested, causally-relevant congruent models which embodied valid restrictions would both fit best, and by encompassing, dominate in forecasting on average. Unfortunately, new unpredictable events continue to occur, so any operational theory of economic forecasting must allow for data moments altering: see Stock and Watson (1996) and Clements and Hendry (2001b) on the prominence of structural change in macroeconomic time-series.

Thus, for forecasting *per se*, correctly established causality in congruent models does not ensure success. However, care is essential in how that result is interpreted. Location shifts (or mimics thereof) are sufficiently common to have left a less than impressive track record of macro-econometric forecasting. No one can forecast the unpredictable, and all devices will fail for events that were *ex ante* unpredictable. But some devices do not adjust after breaks and so suffer systematic forecast failure: equilibrium-correction mechanisms based on cointegration relationships whose mean has changed – but that is not known – are a potentially disastrous example. Nevertheless, causal-based modelling should not be abandoned as a basis for forecasting, particularly in a policy context (see section 7): rather, the solution lies in formulating variants of models that are robust to such shifts. Intercept corrections and additional differencing are simple possibilities, but hopefully better ones will be developed now that an explanation exists for the historical outcomes (see e.g., Hendry, 2004).

However, it is the converse implication that is crucial: causal links are not sensibly tested by forecast evaluation, since neither success nor failure entails correct or incorrect attribution of causality. Forecast failure could, but need not, imply inappropriately attributed causal links. Indeed, all four possibilities can occur without logical contradiction: correct causality followed by forecast failure; incorrect causal-

ity followed by forecast failure; correct causality followed by forecast ‘success’; incorrect causality followed by forecast ‘success’. Worse still, ‘tricks’ exist that can help avoid forecast failure independently of the validity of any causal attributions by the model in use: intercept corrections are a well-known device with such properties.

Despite such a fundamental result, many economists seem to persist in the view that ‘forecasting is the ultimate test of a model’. Three comments can be made about such a view: see Clements and Hendry (2003). First, *ex post* parameter-constancy tests and *ex ante* forecast evaluations have very different properties, with the latter susceptible to many additional problems such as (increased) data inaccuracy at the forecast-origin and over the forecast horizon. Secondly, non-constancy in coefficients of mean-zero variables has a much smaller impact on forecast accuracy than changes in location components. Thirdly, when the future can differ substantively from the past, ‘because of the things we don’t know we don’t know’ (see Singer, 1997), forecast failure is not so much a diagnostic of a model as an indication of an event unrelated to existing information: thus, it potentially provides new knowledge. Consequently, if the basis for their view is that new evidence is needed for ‘independent’ checking of inferences, when data moments are non-constant, then great care is needed in using such information correctly, and *ex ante* forecast evaluation is unlikely to be a reliable approach to doing so.

Turning to the role of Granger causality, for multi-step conditional forecasts, EGNC of the conditioning variables is crucial for valid inferences: neglected feedbacks would otherwise violate the conditioning assumptions. On the other hand, EGNC does not matter in closed models, where every variable is jointly forecast, nor for 1-step ahead forecasts even in open models, nor if multi-step estimation is used.

Surprisingly, therefore, in the forecasting arena, neither sense of causality can be accorded an important role.

7 Exogeneity and causality in policy analysis

Ericsson (1992), Ericsson, Hendry and Mizon (1998), and Hendry and Mizon (1998) consider the roles of exogeneity, causality, and co-breaking in policy analysis, so again we merely summarize the discussion, but return to co-breaking in section 7.1. Also, Granger (1988a) discusses the role of Granger causality in policy effectiveness (but see Buiter, 1984), and Granger and Deutsch (1992) investigate the evaluation of policy models.

The use of models for economic policy analysis is the arena where causality plays a fundamentally important role. If a variable changed by an economic policy is causal, but omitted from a model, or is not causal but is deemed such in a model, then incorrect predictions of its effects are virtually bound to occur (although luck can never be excluded). Accurate predictions of the effects of a change in a policy variable will only result if that variable can be changed, is causal in a model that treats it as such, when the transmission of the effect is invariant, and has been appropriately estimated. Thus, stringent conditions must be satisfied for successful policy analysis.

Based on the results reviewed in section 6, Hendry and Mizon (2000) show that the ‘best’ forecasting model may be useless for policy. Conversely, and perhaps surprisingly in view of the Lucas (1976) critique, forecast failure *per se* is not sufficient to reject a policy model. The rationale for the first step of their analysis is that extrapolative devices need not even involve policy variables; and for the second, that forecast failure primarily derives from unanticipated location shifts which need not (but of course, could)

affect policy conclusions. Moreover, since extrapolative devices rarely include the policy instruments, shifts in those instruments then act as post-forecasting breaks to such devices, inducing a failure not present in an econometric model which is well-specified for the policy effects. Consequently, compared to in-sample structural breaks, the situation reverses for these two model types. Hence, pooling of both forms of model will be needed in the face of location structural breaks and policy regime shifts.

Policy outcomes depend on reaction parameters connecting target variables with instruments. As shown in Hendry (2000b), structural breaks which do not alter the unconditional expectations of the $\mathsf{I}(0)$ transforms of variables are not easily detected by conventional constancy tests, so rarely induce forecast failure. This has adverse implications for impulse-response analyses as discussed in section 7.3 below. However, most policy changes entail location shifts in variables (as against, e.g., mean-preserving spreads), and hence provide a crucial step in a progressive research strategy: if causal attribution is incorrect, then forecast failure should result from the policy, allowing substantive learning of causal connections, providing these do not themselves change too often (which seems unlikely). Thus, research effort into establishing which forecast failures resulted from policy change,s and which from other sources of location shifts, would seem merited.

7.1 Co-breaking

Co-breaking is the property that when variables shift, there exist linear combinations of variables which do not shift, and so are independent of the breaks (see Clements and Hendry, 1999, ch. 9, Krolzig and Toro, 2002, and Massmann, 2001). Co-breaking is analogous to cointegration, where a linear combination of variables is stationary even though all the component series are integrated. Whenever there is co-breaking between the instruments of economic policy and the target variables, changes in the former will produce consistent changes in the latter, so constant co-breaking implements a causal relation. The existence of co-breaking between the means of the policy instruments and those of the targets is testable: a policy shift that induced forecast failure would be strong evidence that the causal links were incorrectly specified. As section 7.3 shows, co-breaking is also necessary to justify impulse-response analysis.

7.2 Control

The status of variables in a system can be altered by deliberate changes in governmental control procedures. For example, a variable (such as an interest rate) that was weakly exogenous for the parameters of (say) an inflation equation, can cease to be so after a control rule is introduced (see e.g., Johansen and Juselius, 2000), yet the VAR involved need not suffer from forecast failure. Of course, the control rule can be effective only if there is an already existing causal link between the instrument and the target, so the new feedback rule can exploit that link to achieve its objective: control rules cannot create policy-target links by specifying target-instrument links.

7.3 Impulse response analyses

Although impulse response analyses are widespread (see e.g., Lütkepohl, 1991, Runkle, 1987, and Sims, 1980), they suffer from many drawbacks: see Banerjee, Hendry and Mizon (1996), Ericsson *et al.* (1998), and Hendry and Mizon (1998).[9] Here, we focus on those problems which are germane to a

[9] The literature on 'structural VARs' (see, e.g., Bernanke, 1986, and Blanchard and Quah, 1989) faces similar difficulties.

discussion of exogeneity and causality in non-stationary processes.

First, weak exogeneity is crucial for impulse-response analyses since the same size of 'shock' to the error and to the intercept are indistinguishable in a model, but the reaction in the economy will be as anticipated only if the means and variances are linked in the same way – which is (e.g.) the weak exogeneity condition for a conditional equation in a VAR. Specifying a variable to be weakly or strongly exogenous alters the calculated impulse responses, irrespective of whether or not that variable actually is exogenous. Moreover, results are dependent on the ordering of variables – but 'orthogonalized impulses' violate weak exogeneity for most selections, so may lose invariance to the very shocks to be analyzed.

Secondly, in closed systems – where policy variables are 'modelled' – impulse-response analyses assume that the process remains constant under the 'shock', so actually requires super exogeneity of the appropriate policy conditioning variables. Alternatively expressed, unless there is mean co-breaking between the shocked variable and the target, the actual responses in the economy will not match those derived from the model. Impulse-response approaches to evaluating the policy implications of models are also dependent on the absence of 'undetected breaks', so can be mis-leading in both sign and magnitude when shifts have occurred in zero-mean variables, even when models are rigorously tested (and certainly so when no such testing has occurred): see Hendry and Mizon (2000).

Thirdly, even when a model is rigorously derived from a consistent theory, is congruent and encompasses rival models, and its policy implications are invariant to extensions of the information used – so it embodies structure (see Hendry, 1995b) – that does not imply that its residuals are structural. Residuals can only be invariant to extensions of information over time, regimes and variables if the model coincides with the DGP, and not just the LDGP. Error terms in equations induce changes in the values taken by the dependent variable only by an assumption of causality so, for example:

$$y_t = \boldsymbol{\beta}'\mathbf{z}_{t-1} + \epsilon_t, \tag{13}$$

is usually written on the assumption that:

$$\frac{\partial y_t}{\partial \epsilon_t} = 1. \tag{14}$$

This is far from certain in a non-experimental discipline, where (13) could simply be a decomposition, so $\epsilon_t = y_t - \boldsymbol{\beta}'\mathbf{z}_{t-1}$, and not a causal model. Conversely, (14) does hold if all errors are due to mis-measurement of y_t, but then is uninformative about causal structure. Further, residuals are inevitably derived representations of our ignorance: indeed, residuals may not even contain all the actual 'DGP shocks', because incorrectly included variables may have 'absorbed' some of those. The confusion of residuals with errors is an egregious mistake, as impulse responses are meaningless in a world of derived models.

Finally, Granger non-causality may be sufficient for the equivalence of standard-error based impulse responses from systems and conditional models, but does not ensure efficient or valid inferences unless the conditioning variables are weakly exogenous.

The implication of this analysis is that unless causality has been independently established between an input variable and an output, impulse responses are not a reliable inference procedure.

8 Conclusion

Both exogeneity and causality play different roles in modelling, forecasting and policy. This is well known for exogeneity, where different concepts have been explicitly defined in Engle *et al.* (1983), but

seems less well established for causality. A cause was viewed as an asymmetrical process inducing change over time in a structure. The perceptive reader will not have missed the close connections to dynamic econometric systems which are invariant to extensions of information (over time, interventions, and additional variables). When causality is defined within a theory model, the correspondence of the model to reality becomes the key link. Thus, we conclude that causality is important in modelling; and manifestly crucial in policy. However, causality cannot be proved to be a necessary property of variables in dominating forecasting models. When location shifts occur, robust forecasting methods can outperform 'causal models'. Nevertheless, by itself, such a finding – or even forecast failure – is insufficient to preclude the use of the causal model for policy: the reaction parameters of interest to policy could have remained constant, perhaps due to co-breaking.

Both concepts (exogeneity and causality) seem robust to extensions to non-stationary systems, but their implications are sometimes less clear cut, and inferences about them can be more hazardous. In particular, pre-existing exogeneity, or causal direction, can alter over time. Conversely, it can be difficult in weakly stationary systems to correctly ascertain exogeneity or causality since the systems in question do not change enough. Thus, the news is not all bad. For example, it is well known that cointegrated relations which remove unit roots provide a basis for testing long-run weak exogeneity; and policy-induced location shifts can highlight the presence or absence of causal links.

Granger causality shares many of the characteristics of the general definition, together with the resultant inferential difficulties. The requirement that causality be judged against the universe of available information renders it non-operational; but attempts to infer GNC from empirical models become prone to serious errors unless a congruent, encompassing and invariant system is used.

The strength of evidence about causality depends on the magnitudes of changes in inputs which nevertheless produce consistent output responses. Co-breaking with causal links is needed to sustain economic policy, since few policy changes are of the 'mean-preserving spread' form, and most involve location shifts. Although the latter are the main problem for forecasting, in a progressive research strategy, the resultant forecast failure can be of benefit for modelling, and so later policy. Conversely, shifts in mean-zero parameters are difficult to detect in forecasting, but can seriously distort impulse-response based policy analyses.

Finally, without mean co-breaking, or causal links, impulse responses need not deliver useful information about reactions in the economy. There seems no alternative to modelling the exogeneity and causality structure of the economy if reliable policy inferences are desired.

References

Allen, P. G., and Fildes, R. A. (2001). Econometric forecasting strategies and techniques. In Armstrong, J. S. (ed.), *Principles of Forecasting*, pp. 303–362. Boston: Kluwer Academic Publishers.

Banerjee, A., Hendry, D. F., and Mizon, G. E. (1996). The econometric analysis of economic policy. *Oxford Bulletin of Economics and Statistics*, **58**, 573–600.

Basmann, R. L. (1988). Causality tests and observationally equivalent representations of econometric models. *Journal of Econometrics*, **39**, 69–104.

Bernanke, B. S. (1986). Alternative explorations of the money-income correlation. In Brunner, K., and Meltzer, A. H. (eds.), *Real Business Cycles, Real Exchange Rates, and Actual Policies*, Vol. 25

of *Carnegie-Rochester Conferences on Public Policy*, pp. 49–99. Amsterdam: North-Holland Publishing Company.

Bhansali, R. J. (2002). Multi-step forecasting. in Clements, and Hendry (2002a), pp. 206–221.

Blalock, H. M. J. (1961). *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Blanchard, O., and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, **79**, 655–673.

Bontemps, C., and Mizon, G. E. (2003). Congruence and encompassing. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 354–378. Princeton: Princeton University Press.

Boswijk, H. P. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.

Buiter, W. H. (1984). Granger causality and policy effectiveness. *Economica*, **51**, 151–162.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.

Chamberlain, G. (1982). The general equivalence of Granger and Sims causality. *Econometrica*, **50**, 569–582.

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. (2001a). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Clements, M. P., and Hendry, D. F. (2001b). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

Clements, M. P., and Hendry, D. F. (eds.)(2002a). *A Companion to Economic Forecasting*. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2002b). Explaining forecast failure in macroeconomics. in *A Companion to Economic Forecasting* (2002a), pp. 539–571.

Clements, M. P., and Hendry, D. F. (2003). Evaluating a model by forecast performance. Unpublished paper, Economics Department, University of Warwick.

Coveney, P., and Highfield, R. (1990). *The Arrow of Time*. New York: Fawcett Columbine.

Cox, D. R. (1992). Causality: Some statistical aspects. *Journal of the Royal Statistical Society, A*, **155**, 291–301.

Dolado, J. J. (1992). A note on weak exogeneity in VAR cointegrated systems. *Economics Letters*, **38**, 139–143.

Doll, R. (2001). Proof of causality: Deductions from epidemiological evidence. Fisher memorial lecture, University of Oxford, Oxford.

Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publish-

ers, 1993, and Oxford University Press, 2000; and in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

Ericsson, N. R. (1992). Cointegration, exogeneity and policy analysis: An overview. *Journal of Policy Modeling*, **14**, 251–280.

Ericsson, N. R., Hendry, D. F., and Mizon, G. E. (1998). Exogeneity, cointegration and economic policy analysis. *Journal of Business and Economic Statistics*, **16**, 370–387.

Ericsson, N. R., and Irons, J. S. (1994). *Testing Exogeneity*. Oxford: Oxford University Press.

Favero, C., and Hendry, D. F. (1992). Testing the Lucas critique: A review. *Econometric Reviews*, **11**, 265–306.

Fildes, R. A., and Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, **63**, 289–308.

Fischer, A. M. (1989). Policy regime changes and monetary expectations: Testing for super exogeneity. *Journal of Monetary Economics*, **24**, 423–436.

Florens, J.-P., and Mouchart, M. (1982). A note on non-causality. *Econometrica*, **50**, 583–592.

Florens, J.-P., and Mouchart, M. (1985). A linear theory for noncausality. *Econometrica*, **53**, 157–175.

Geweke, J. B. (1984). Inference and causality in economic time series models. In Griliches, Z., and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 2, Ch. 19. Amsterdam: North-Holland.

Gottlieb, A. (2000). *The Dream of Reason*. London: The Penguin Press.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.

Granger, C. W. J. (1980). Testing for causality – A personal viewpoint. *Journal of Economic Dynamics and Control*, **2**, 329–352.

Granger, C. W. J. (1988a). Causality, cointegration, and control. *Journal of Economic Dynamics and Control*, **12**, 551–559.

Granger, C. W. J. (1988b). Some recent developments in the concept of causality. *Journal of Econometrics*, **39**, 199–211.

Granger, C. W. J., and Deutsch, M. (1992). Comments on the evaluation of policy models. *Journal of Policy Modeling*, **14**, 497–516.

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics*, **115**, 45–97.

Hendry, D. F. (1988). The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers*, **40**, 132–149. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

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. (1995c). On the interactions of unit roots and exogeneity. *Econometric Reviews*, **14**, 383–419.

Hendry, D. F. (1997). The econometrics of macroeconomic forecasting. *Economic Journal*, **107**, 1330–1357. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Hendry, D. F. (2000a). *Econometrics: Alchemy or Science?* Oxford: Oxford University Press. New

Edition.

Hendry, D. F. (2000b). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in *The Economics of Structural Change*, Hagemann, H. Landesman, M. and Scazzieri (eds.), Edward Elgar, Cheltenham, 2002.

Hendry, D. F. (2003). A modified Gauss–Markov theorem for stochastic regressors. Unpublished paper, Economics Department, Oxford University.

Hendry, D. F. (2004). Robustifying forecasts from equilibrium-correction models. Unpublished paper, Economics Department, University of Oxford.

Hendry, D. F., and Clements, M. P. (2003). Economic forecasting: Some lessons from recent research. *Economic Modelling*, **20**, 301–329. European Central Bank, Working Paper 82.

Hendry, D. F., and Doornik, J. A. (1997). The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy*, **44**, 437–461. Special Issue.

Hendry, D. F., and 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., and Mizon, G. E. (1999). The pervasiveness of Granger causality in econometrics. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*. Oxford: Oxford University Press.

Hendry, D. F., and Mizon, G. E. (2000). Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy*, **16**, 138–159.

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

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960 and 968–970.

Hoover, K. D. (1990). The logic of causal inference: Econometrics and the conditional analysis of causation. *Economics and Philosophy*, **6**, 207–234.

Hoover, K. D. (2001). *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

Hume, D. (1758). *An Enquiry Concerning Human Understanding, (1927 ed.)*. Chicago: Open Court Publishing Co.

Hunter, J. (1992a). Cointegrating exogeneity. *Economics Letters*, **34**, 33–35.

Hunter, J. (1992b). Tests of cointegrating exogeneity for PPP and uncovered interest rate parity in the United Kingdom. *Journal of Policy Modeling*, **14**, 453–463.

Johansen, S. (1992). Testing weak exogeneity and the order of cointegration in UK money demand. *Journal of Policy Modeling*, **14**, 313–334.

Johansen, S., and Juselius, K. (2000). How to control a target variable in the VAR model. Mimeo, European University of Institute, Florence.

Koopmans, T. C. (1950). When is an equation system complete for statistical purposes?. In Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models*, No. 10 in Cowles Commission Monograph, Ch. 17. New York: John Wiley & Sons.

Krolzig, H.-M., and Toro, J. (2002). Testing for super-exogeneity in the presence of common deterministic shifts. *Annales d'Économie et de Statistique*, **67/68**, 41–71.

Lauritzen, S. L., and Richardson, T. S. (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society, B*, **64**, 1–28.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Brunner, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Lutkepohl, H. (1982). Non-causality due to omitted variables. *Journal of Econometrics*, **19**, 367–378.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.

Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

Massmann, M. (2001). Co-breaking in macroeconomic time series. Unpublished paper, Economics Department, Oxford University.

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.

Morgan, M. S. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Mosconi, R., and Giannini, C. (1992). Non-causality in cointegrated systems: Representation, estimation and testing. *Oxford Bulletin of Economics and Statistics*, **54**, 399–417.

Newbold, P. (1982). Causality testing in economics. In Anderson, O. (ed.), *Time Series Analysis: Theory and Practice 1*, pp. 701–716. Amsterdam, The Netherlands: North Holland.

Paruolo, P., and Rahbek, A. (1999). Weak exogeneity in I(2) systems. *Journal of Econometrics*, **93**, 281–308.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Phillips, A. W. H. (1957). Stabilization policy and the time form of lagged response. *Economic Journal*, **67**, 265–277. Reprinted in Leeson, R. (ed.) *A. W. H. Phillips: Collected Works in Contemporary Perspective*, Cambridge: Cambridge University Press, 2000.

Phillips, P. C. B. (1988). Reflections on econometric methodology. *Economic Record*, **64**, 344–359.

Phillips, P. C. B., and Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies*, **58**, 407–436.

Richard, J.-F. (1980). Models with several regimes and changes in exogeneity. *Review of Economic Studies*, **47**, 1–20.

Runkle, D. E. (1987). Vector autoregressions and reality. *Journal of Business and Economic Statistics*, **5**, 437–442.

Simon, H. A. (1952). On the definition of causal relations. *Journal of Philosophy*, **49**, 517–527.

Simon, H. A. (1953). Causal ordering and identifiability. In Hood, W. C., and Koopmans, T. C. (eds.), *Studies in Econometric Method*, No. 14 in Cowles Commission Monograph, Ch. 3. New York: John Wiley & Sons.

Simon, H. A. (1957). *Models of Man*. New York: John Wiley & Sons.

Sims, C. A. (1972). Money, income and causality. *American Economic Review*, **62**, 540–552.

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.

Singer, M. (1997). Thoughts of a nonmillenarian. *Bulletin of the American Academy of Arts and Sciences*, *51*(2), 36–51.

Smith, A. (1795). The history of astronomy.. pp. 33–105. Edinburgh: W. Creech. Liberty Classics edition, by I. S. Ross, 1982.

Stock, J. H., and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.

Strotz, R. H., and Wold, H. O. A. (1960). Recursive versus non-recursive systems: An attempt at a synthesis. *Econometrica*, **28**, 417–421.

Toda, H. Y., and Phillips, P. C. B. (1993). Vector autoregressions and causality. *Econometrica*, **61**, 1367–1393.

Toda, H. Y., and Phillips, P. C. B. (1994). Vector autoregressions and causality: A theoretical overview and simulation study. *Econometric Reviews*, **13**, 259–285.

Urbain, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics*, **54**, 187–207.

Weissmann, G. (1991). Asprin. *Scientific American*, 58–64.

Wold, H. O. A. (1969). Econometrics as pioneering in non-experimental model building. *Econometrica*, **37**, 369–381.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, **60**, 812–838.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society*, **89**, 1–64. Reprinted in Hendry, D. F. and Morgan, M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Zellner, A. (1979). Causality and econometrics. In Brunner, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, pp. 9–54. Amsterdam: North-Holland Publishing Company.