EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 10, No. 1, 2017, 12-29
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

Sir Clive W.J. Granger Memorial Special Issue on Econometrics

Granger Causality

David F. Hendry

Economics Department and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK.

Abstract. Despite an extraordinary level of debate about the concept of 'causality' and establishing causal links, Granger (1969) [18] proposed an approach based on the arrow of time and the effects of eliminating the history of some variables from the joint distribution of all variables. There was no Granger causality from the eliminated variables if the conditional and marginal distributions of the remaining variables were the same. In practice, the non-operational nature of his definition was finessed by testing whether dropping a subset of variables from a larger set affected the goodness of fit of models of the remaining variables. This paper notes the drawbacks that arise from such a route to making his concept of causality operational, but also emphasises its pervasive role in econometric modelling of time series.

2010 Mathematics Subject Classifications: 91G70, 91B84, 62P20, 62M10

Key Words and Phrases: Clive Granger, Granger Causality, Empirical Granger Non-causality, Testing, Encompassing, Simulating, Forecasting

---

1. Introduction

Scientific knowledge is always and everywhere fallible. Corroboration of a theory is insufficient justification, because false theories—such as the Ptolemic model of the solar system—can find supporting evidence. Equally, apparent rejection is often not definitive as it may be due to mistakes in the test, as with Pasteur's germ theory—seeking to replace Aristotelian notions of 'spontaneous generation' of life—which was apparently refuted initially by what later became known as thermophiles. Newtonian gravitational theory is an interesting example. Believed by many at the time to be 'truth', its verisimilitude was questioned by Adam Smith (1795) [50], who claimed it was just a 'model' of the

Email address: `david.hendry@nuffield.ox.ac.uk` (D. F. Hendry)

solar system, and would be replaced in due course by a better model: "... philosophical systems are mere inventions of the imagination...". That replacement has of course actually occurred, but unless relativistic effects matter, as with clocks on orbiting satellites, Newtonian calculations are still used.

Despite the absence of certainty in science, huge advances have occurred in understanding our physical and biological world, manifested in the advanced technologies that are part of everyday life. The close interactions between ever better established empirical evidence and increasingly general theories thereof have sustained those advances, so certainty does not seem necessary to the scientific endeavour. Conversely, there remain known gaps in our understanding in most scientific fields, of which explaining galaxies' rates of rotation, currently attributed to unknowns called 'dark matter' and 'dark energy', are among the most salient in physics.

'Causal knowledge' is just a special case of scientific knowledge, but with the additional problems of being a disputable concept, doubts about inference procedures to determine causality, and occasional failures to distinguish proof from evidence. David Hume (1758) [36] set the scene for grave doubts about establishing causality by asserting that we cannot *know* necessary connections in reality: they must always be based on induction, so must be tenuous and open to failure. The modern refrain is that of 'confusing correlations with causes', but it may be easier to remember that a relation holding for many observations cannot *prove* it will hold for all, even when supported by accurate predictions of new phenomena. Nevertheless, Hume himself argued in 'causal' terms when it came to his analyses in economics: for example, that the influx of gold and silver from South America had caused inflation in Spain. Hume knew he seemed inconsistent to others by doing so, but claimed he was not: 'My practice, you say, refutes my doubts. But you mistake the purport of my question'. In other words, while we cannot *know for certain* that the observed increase in precious metals was the 'true cause' of the recorded inflation, it was the most likely culprit.

In many ways, 'causality' remains a philosophical minefield, exacerbated by its apparent absence at quantum levels, the symmetry of time in many physical theories notwithstanding that the second law of thermodynamics provides an 'arrow of time', the existence of one-off events, the role of anticipations in human behaviour, closely related concepts like 'simultaneity' being ambiguous in the general theory of relativity, the lack of either necessity or sufficiency in causal links in multivariate evolving worlds, processes potentially 'enabling' rather than causing outcomes, and in economics the added complexities—and potential benefits—of its processes being wide-sense non-stationary. Yet everyday thinking is replete with causal assertions: the car stopped because the driver braked, or output rose because interest rates fell. However, 'causal chains' may have many steps, and 'ultimate causes' may be hidden: as we know from crime novels, if the brake cable is cut, or brake fluid is too low, braking can fail, so the 'stopping system' is not invariant to such 'interventions'—a notion that will recur below.

Debates about causal connections and causal transmission mechanisms have indeed long been an issue in economics, noting that [49] was entitled *An Inquiry into the Nature* *and **Causes** of the Wealth of Nations*, even though Smith was a close friend of Hume.* Hume's views, however well based they may seem, have not stopped others from investigating 'causality', and how to determine it empirically: for general overviews see among many others, Herbert Simon [46], Rom Harré and E.H. Madden (1975) [22], John Hicks [34], John Mackie [39] and Nancy Cartwright [4]. Statistical aspects are considered by David Cox [9]; Steffen Lauritzen and Thomas Richardson (2002) [38] seek causal interpretations in chain graph models; and Richard Doll [10] discusses principles and practices for establishing causality from empirical evidence. Jim Heckman (2000,2005) [23] [24] offers an economic-theory concept of causality, Kevin Hoover [35] considers its role in macroeconomics; in [26], I sought to generalize notions of causality and exogeneity to non-stationary time series; Kun Zhang, Jiji Zhang and Bernhard Schölkopf (2015) [51] proposed distinguishing cause from effect using exogeneity; and Vassilios Bazinas and Bent Nielsen (2014) [2] seek to determine causal transmission from specific shifts affecting different aspects of an economic system.

Expressing it somewhat anachronistically (as many of the above references come later), into this intellectual minefield steps Clive Granger [18], defining causality in relation to changes in the joint distributions of observables: if deleting the complete history of a variable from the universe of information does not alter the joint distribution of the remaining variables, then the omitted variable does not (Granger) cause the remainder. Granger non-causality (denoted GNC) depends on 'time's arrow' being unidirectional, so only the past can cause the present.† Thus, Granger's concept of causality takes place in time, and is asymmetric as it is the past of a variable that does or does not induce changes in other variables. An implication is that there is no Granger causality from the eliminated variable to the remaining variables if the conditional distributions of the remaining variables given the history of the eliminated variable is the same as their marginal distributions after elimination. Since the universe of information is never available, and in any case, it could never be known that it was all possible information, such a definition is clearly non-operational. Like all other characterizations of causality, one can never be certain of evidence for, or against, Granger non-causality.

In practice, the non-operational nature of his definition has been finessed by testing whether dropping a subset of lagged variables from a larger set affects the goodness of fit of models of the remaining variables. However, it is inappropriate to test **causality** by simply deleting lags of a variable from, or adding them to, a model and seeing if the effects are significant: that certainly confuses correlations with causes. Such a procedure tests in-sample relevance within the subset of variables under analysis, and may presage forecasting improvements if the world does not change, but cannot possibly be viewed as revealing causality unless the initial set is somehow provable to be the universe of all variables.

*As an aside, their correspondence is still a wonderful, and humourous, read. For example, Hume wrote to Smith about Smith's earlier famous work, [48] on its publication: "I proceed to tell you the melancholy News, that your book has been very unfortunate: For the public seem disposd to applaud it extremely.... You may conclude what Opinion true Philosophers will entertain of it..."

†Expectations about the future are necessarily based on past information, so any causal impact on the present is a product of the past.

Unfortunately, such a confusion abounds in the empirical literature, with studies often claiming evidence for Granger causality from bivariate models. This paper notes several other important drawbacks that arise from such a route to making his concept of causality operational, but it also emphasises the remarkably pervasive role of empirical GNC in econometric modelling of time series, based on Hendry and Grayham Mizon (1999) [29], many in areas not envisaged by Granger's initial concept. In Granger (1980,1988) [19], [21], he sought to clarify aspects of his concept of causality, and together with [18], these three articles have attracted more than 20,000 citations from other scholars.

The structure of the paper is as follows. Section 2 formalizes Granger non-causality in terms of the impacts of eliminating one time-series variable from the joint density of all variables. Section 3 discusses how seeking to make GNC operational creates the different concept of empirical Granger non-causality (EGNC), then considers the resulting drawbacks of that approach. Section 4 shows that notwithstanding any drawbacks EGNC may have as an articulation of 'causality', it remains relevant to ten major areas of econometric modelling of time series. Section 5 concludes.

2. Formalizing Granger non-causality

Let $\mathbf{X}_0$ denote the initial conditions up to time 0 of the N-dimensional real-valued vector time series $\{\mathbf{x}_t\}$, so records its complete history, and let $\mathbf{X}_{t-1} = (\mathbf{X}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{t-1}) = (\mathbf{X}_0, \mathbf{X}^1_{t-1})$. The joint density of $\mathbf{x}_t$ for $t = 1, \ldots, T$ is denoted by $\mathsf{D}_{\mathsf{X}^1_\mathsf{T}}(\mathbf{X}^1_T|\mathbf{X}_0)$. Because nothing precludes the distributions of variables changing over time, $\mathsf{D}_{\mathsf{X}^1_\mathsf{T}}(\cdot)$ needs subscripted by both the variables in question ($\mathbf{X}$) and the time period under analysis $(1, \ldots, T)$. At this stage, $\mathsf{D}_{\mathsf{X}^1_\mathsf{T}}(\cdot)$ is not characterized by any parameters.

Sequentially factorize $\mathsf{D}_{\mathsf{X}^1_\mathsf{T}}(\cdot)$ as (see Joseph Doob (1953), [11]):

$$\mathsf{D}_{\mathsf{X}^1_T}\left(\mathbf{X}^1_T \mid \mathbf{X}_0\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}\right) \tag{1}$$

This step creates a martingale difference sequence when all the $\mathsf{D}_{\mathsf{x}_t}(\cdot)$ are known, since letting $\mathbf{v}_t = \mathbf{x}_t - \mathsf{E}_{\mathsf{x}_t}[\mathbf{x}_t|\mathbf{X}_{t-1}]$, then $\mathsf{E}_{\mathsf{x}_t}[\mathbf{v}_t|\mathbf{X}_{t-1}] = \mathbf{0}$ and hence $\mathsf{E}_{\mathsf{x}_t}[\mathbf{v}_t|\mathbf{V}_{t-1}] = \mathbf{0}$ as well.

Next, partition $\mathbf{x}'_t = (\mathbf{x}'_{1,t}, \mathbf{x}'_{2,t})$ into $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, and factorize each $\mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t|\mathbf{X}_{t-1})$ in (1) as:

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

To match this conditional-marginal factorization, partition $\mathbf{X}_{t-1}$ into $\mathbf{X}_{1,t-1}$ and $\mathbf{X}_{2,t-1}$. Then:

$$\mathsf{D}_{\mathsf{X}^1_T}\left(\mathbf{X}^1_T \mid \mathbf{X}_0\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{1,t}|\mathsf{x}_{2,t}}\left(\mathbf{x}_{1,t} \mid \mathbf{x}_{2,t}, \mathbf{X}_{1,t-1}, \mathbf{X}_{2,t-1}\right) \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{1,t-1}, \mathbf{X}_{2,t-1}\right) \tag{3}$$

Definition 1. *If $\mathsf{D}_{\mathsf{x}_{2,t}}(\cdot)$ does not depend on $\mathbf{X}_{1,t-1}$ so that:*

$$\mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{1,t-1}, \mathbf{X}_{2,t-1}\right) = \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{2,t-1}\right) \forall t, \tag{4}$$

then $\mathbf{x}_1$ does not Granger cause $\mathbf{x}_2$.

Thus, Granger non-causality (GNC) entails from (3) that marginalizing the distribution of $\{\mathbf{x}_{2,t}\}$ with respect to the history of $\{\mathbf{x}_{1,t}\}$, namely $\mathbf{X}_{1,t-1}$, has no effect:

$$\prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{1,t-1}, \mathbf{X}_{2,t-1}\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{2,t-1}\right) = \mathsf{D}_{(\mathsf{X}_2)_T^1}\left((\mathbf{X}_2)_T^1 \mid (\mathbf{X}_2)_0\right) \tag{5}$$

Consequently a test for GNC can be based on the irrelevance of $\mathbf{X}_{1,t-1}$ as a determinant of $\mathbf{x}_{2,t}$ $\forall t = 1, \ldots, T$. At this stage, leaving open any question of contemporaneous causality as a separate issue, then *if* $\mathbf{X}_{t-1}$ was provably the universal information set over $t = 1, \ldots, T$, it would seem that $\mathbf{X}_{1,t-1}$ is irrelevant to the properties of $\mathbf{x}_{2,t}$, so cannot be a cause.

Granting that $\mathsf{D}_{\mathsf{X}_T^1}\left(\mathbf{X}_T^1 | \mathbf{X}_0\right)$ is the universal distribution for $t = 1, \ldots, T$, then it must be the data generation process (DGP) of $\mathbf{x}_t$. In economics at least, outcomes are determined by human behaviour interacting with their environment, and economic agents appear to have parameters, or relatively stable characteristics, that determine their decisions. Denoting all the parameters of the DGP by $\pi \in \mathbf{\Pi} \subseteq \mathbb{R}^d$, the DGP can be written as $\mathsf{D}_{\mathsf{X}_\mathsf{T}^1}(\mathbf{X}_T^1 | \mathbf{X}_0, \pi)$. Now to sequentially factorize $\mathsf{D}_{\mathsf{X}_\mathsf{T}^1}(\cdot)$ let $\mathbf{g}(\pi) = (\psi_1 \ldots \psi_T)$ for a $1-1$ function $\mathbf{g}(\cdot)$, allowing that $\{\psi_t\} \in \mathbf{\Psi} \subseteq \mathbb{R}^d$ need not be constant over time if regime shifts, structural breaks or changes in behaviour occur. Then:

$$\mathsf{D}_{\mathsf{X}_T^1}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \pi\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \psi_t\right) \tag{6}$$

leading to the conditional-marginal sequential factorization:

$$\mathsf{D}_{\mathsf{X}_T^1}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \pi\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{1,t}|\mathsf{x}_{2,t}}\left(\mathbf{x}_{1,t} \mid \mathbf{x}_{2,t}, \mathbf{X}_{t-1}, \psi_{1,t}\right) \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{t-1}, \psi_{2,t}\right) \tag{7}$$

where $\psi_t' = (\psi_{1,t}', \psi_{2,t}')$ sustains this factorization. Now, Granger non-causality (GNC) entails from (7) that:

$$\prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{1,t-1}, \mathbf{X}_{2,t-1}, \psi_{2,t}\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{x}_{2,t}}\left(\mathbf{x}_{2,t} \mid \mathbf{X}_{2,t-1}, \psi_{2,t}\right) \tag{8}$$

Ignoring parameters in Definition 1 is an important lacuna. The added complications from tracking what happens to the parameters are not just that the $\{\psi_{2,t}\}$ need not be constant over time, although validly testing the irrelevance of $\mathbf{X}_{1,t-1}$ in (8) requires such knowledge, but also that $\{\psi_{1,t}\}$ and $\{\psi_{2,t}\}$ may be linked. In particular, to ascertain GNC of $\mathbf{X}_{1,t-1}$ for $\mathbf{x}_{2,t}$ requires that changes in $\{\psi_{1,t}\}$ do not alter $\{\psi_{2,t}\}$. This is not mere sophistry but a key issue: if changes in $\psi_{1,t}$ shifted $\psi_{2,t}$, but $\psi_{2,t}$ was taken as given in (8), it would appear that there was no link between $\mathbf{x}_1$ and $\mathbf{x}_2$ even though it was changes in the parameters of the former that were causing changes in the parameters of the latter. Conversely, if (3) is used, so parameter changes are ignored, GNC would probably be rejected as $\mathbf{X}_{1,t-1}$ proxied those unmodelled shifts (albeit that such a test would be based on a mis-specified model).

3. Empirical Granger non-causality

As GNC is non-operational, many studies switch to testing empirical Granger non-causality (EGNC), defined as trying to test *causality* by adding lags of an excluded variable to a model and seeing if they are significant. That certainly tests relevance within the subset of variables under analysis, and possibly leads to forecasting improvements if the world does not change, but cannot possibly be viewed as revealing causality, or its absence, irrespective of the test rejecting or not.

A number of drawbacks of EGNC were considered in [29], including:

1. rejecting EGNC of a subset of variables $\mathbf{x}_1$ for another subset $\mathbf{x}_2$ in an estimated model does not entail either the presence or absence of GNC between them in the DGP of which they are part;

2. not rejecting GNC in a DGP need not entail EGNC in a model thereof;

3. the existence of GNC may only hold for some time period, although detecting EGNC requires extended sample periods, almost certainly without knowledge of what periods are relevant.

To understand 1., consider a bivariate model in which a variable $x_{2,t}$ depends on $x_{1,t-1}$ as well as its own lags, where 'depends' is judged by the statistical significance of the coefficient of $x_{1,t-1}$. However, $x_{2,t}$ is in fact determined by a variable $x_{3,t-1}$ which is sufficiently highly correlated with $x_{1,t-1}$ that the latter acts as a good proxy. Conversely, a genuine link between $x_{2,t}$ and $x_{1,t-1}$ can be masked by confounding from other omitted relevant determinants, or even by mis-measurement of the correct values of $x_{1,t-1}$. Then 2. follows from x_1 and x_2 being unrelated in their DGP, but marginalising that DGP with respect to all other variables leads to their being linked in their bivariate sequential distribution. Finally 3. is certainly relevant in economics, and may be in other disciplines, where changes in policy regimes can alter what causes what in the economy. A 'classic' example, albeit where the regime switching dates are known ex post, is the sequence from a 19th and early 20th Century 'gold standard' for exchange rates, so monetary movements determined interest rates in participating countries, through 'fixed but adjustable' exchange rates after Bretton Woods, to 'floating' exchange-rate values where interest rates can affect inflows and outflows of currency: Jean-François Richard [45] provides an insightful analysis.

These three drawbacks can seriously confound empirical tests of GNC as informing about causality in the sense that changes in $\mathbf{X}_{1,t-1}$ directly alter the values of $\mathbf{x}_{2,t}$ in (3), or that changes in $\psi_{1,t}$, which will alter $\mathbf{x}_1$, also shift $\psi_{2,t}$ so thereby change $\mathbf{x}_2$. In addition, as noted above, contemporaneous links (simultaneity relative to the frequency of the data being analyzed) are not taken into account. Nevertheless, EGNC plays a fundamental role in empirical modelling of time series and so is pervasive in econometrics, irrespective of its ability or otherwise to determine 'genuine DGP causes', as we now discuss.

4. Empirical Granger non-causality in 10 areas of econometric modelling

[29] discuss ten areas of econometric modelling of time series where EGNC plays an important role, in that its presence or absence changes the implications that can be drawn:
§4.1 marginalizing with respect to non-modelled lagged variables;
§4.2 validity of contemporaneous conditioning;
§4.3 encompassing;
§4.4 cointegration;
§4.5 distributions of estimators and tests;
§4.6 simulation-based inference and the bootstrap;
§4.7 forecasting;
§4.8 dynamic simulation;
§4.9 policy analysis; and
§4.10 impulse-response analyses.
The first four are aspects of the theory of reduction (see e.g., Hendry and Jurgen Doornik, 2014, [27] for a recent explanation), the next two concern statistical inference, and the last four the application of estimated models. We briefly reconsider these in turn as there are a number of more recent results of relevance.

4.1. Marginalizing with respect to non-modelled lagged variables

Given the above definition of GNC in terms of marginalizing, this is naturally its area of most importance to econometric modelling of time series. Both EGNC and GNC are concepts specifically relevant to the validity of marginalizing as in (8), which is a key step in the theory of reduction explaining the origin of empirical models as approximations to DGPs for the variables under study, called local DGPs (LDGPs).

From (7), it is clear that even when $\{\mathbf{x}_t\}$ is not the universal information set, the presence or absence of EGNC between $\mathbf{X}_{1,t-1}$ and $\mathbf{x}_{2,t}$ will affect the specification of any empirical model for $\{\mathbf{x}_t\}$. Finding that such EGNC holds appears to allow a more parsimonious analysis of the determinants of $\{\mathbf{x}_{2,t}\}$ without having to include $\mathbf{X}_{1,t-1}$ or model the perhaps complex process for $\{\mathbf{x}_{1,t}\}$. However, an additional condition affects whether or not there will be a loss of information from that marginalization, namely the existence or not of any links between $\{\psi_{1,t}\}$ and $\{\psi_{2,t}\}$. When $(\psi_{1,t};\psi_{2,t})$ are variation free, so $(\psi_{1,t};\psi_{2,t}) \in \mathbf{\Psi}_{1,t} \times \mathbf{\Psi}_{2,t}\ \forall t$ always satisfy a cut as in Ole [1], then knowledge about $\{\psi_{1,t}\}$ is uninformative about $\{\psi_{2,t}\}$, and no useful information [1] is lost by analyzing the marginalized density on the right-hand side of (8) when $\{\mathbf{x}_{2,t}\}$ alone is the focus of modelling. As noted in section 2, when there are links between the parameters, especially if changes in $\{\psi_{1,t}\}$ alter $\{\psi_{2,t}\}$, a full understanding of $\{\mathbf{x}_{2,t}\}$ entails joint modelling of $\{\mathbf{x}_t\}$ which might be a difficult task.

4.2. Validity of contemporaneous conditioning

In the conditional-marginal sequential factorization (7), the previous subsection considered the marginal component $\mathsf{D}_{\mathsf{x}_{2,t}}(\mathbf{x}_{2,t}|\mathbf{X}_{t-1},\psi_{2,t})$, specifically whether $\mathbf{X}_{t-1}$ could be

reduced to $\mathbf{X}_{2,t-1}$ without loss. The first right-hand side term was the conditional distribution $\mathsf{D}_{\mathsf{x}_{1,t}|\mathsf{x}_{2,t}}(\mathbf{x}_{1,t}|\mathbf{x}_{2,t}, \mathbf{X}_{t-1}, \psi_{1,t})$, and if that is to be analyzed, a key requirement is the validity of the contemporaneous conditioning of $\mathbf{x}_{1,t}$ on $\mathbf{x}_{2,t}$, which requires weak exogeneity as defined by Robert Engle *et al.* (1983) [15]. Two conditions must be satisfied to achieve that:
(a) the parameters of interest in the analysis, θ say, only depend on $\{\psi_{1,t}\}$; and
(b) $(\psi_{1,t}; \psi_{2,t}) \in \mathbf{\Psi}_{1,t} \times \mathbf{\Psi}_{2,t}$ $\forall t$ as in §4.1, so are variation free.
Then the model of $\{\mathbf{x}_{1,t}\}$ can be analyzed given $\{\mathbf{x}_{2,t}\}$. These requirements do not explicitly include EGNC, so hold irrespective of the reduction in (8). However, [15] defined strong exogeneity as weak exogeneity plus (8) holding, so $(\mathbf{X}_2)^1_T$ can be treated as given when analyzing $\{\mathbf{x}_{1,t}\}$, which will be important for conditional forecasting in §4.7.

[15] also defined a concept of super exogeneity as weak exogeneity combined with invariance of the parameters of the conditional model to shifts in the distributions of the conditioning variables, which will also prove relevant in §4.9. Recent methods of model selection based on indicator saturation as discussed in [27] would allow the validity of conditioning to be tested when marginal distributions shifted.‡

4.3. Encompassing

Encompassing is when a model M_1 can explain the results obtained by a rival model M_2, measured statistically for regressions (say) by the variables in M_2 being irrelevant when added to those of M_1: see Christophe Bontemps and Mizon [3] for an excellent overview. [41] formalize encompassing using a Wald encompassing test, and Bernadette Govaerts et al. (1993) [17] highlight the role of Granger causality in the distributions of some encompassing test statistics when 'completing' models are needed to link the variables in the two models under consideration. Denote the regressors in the two models of the same dependent variable by $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, without precluding there may be variables in common in practice, where some of those variables may also enter lagged. Then a completing model M_c that merely projected the $\mathbf{x}_{2,t}$ on the $\mathbf{x}_{1,t}$, so without including lags although each vector actually Granger caused the other, would lead to an invalid test. As a consequence, [17] prefer an F-test of adding the regressors of M_2 to M_1 (and conversely) to test their marginal significance. As with the class of non-nested tests in Cox (1961,1962), [7], [8], possible outcomes include rejecting M_1, or M_2, both, or neither model against the other.

4.4. Cointegration

In our article in this special issue, Jennifer L. Castle and I discuss Granger's key contributions to analyzing cointegration, for which he was awarded the 2003 *The Sveriges Riksbank Prize in Economic Science in Memory of Alfred Nobel.* Here, we merely note a

‡The concept of weak exogeneity has not been much discussed in the statistics literature, but [31] illustrate its relevance to even the venerable Gauss–Markov theorem. When the mean (μ say) of the conditioning variable is the same parameter as the slope in the conditional model, which link is not excluded by the usual assumptions for the theorem's proof, then a more efficient linear unbiased estimator can be based on the former.

couple of additional implications for GNC of cointegration: for a more extensive treatment, see [42].

So far, the integration status of the variables has not been addressed explicitly, although they have been treated as wide-sense non-stationary. As initially formulated by [20] and [14], cointegration links a system of variables that are integrated of first order, denoted $\mathsf{I}(1)$, so are non-integrated after differencing once. Cointegrating combinations are then $\mathsf{I}(0)$, and by being equilibrium correcting, must enter lagged in at least one equation in the system, so entail empirical Granger causality of the dependent variable of that equation by the other variables in the equilibrium correction. There could of course be feedbacks between the differenced variables in addition.

Although in general there are no implications of Granger causality for weak exogeneity in $\mathsf{I}(0)$ processes, feedbacks of a cointegrating vector onto more than one variable both entail empirical Granger causality and a violation of weak exogeneity for the parameters of the cointegrating combination, sometimes called a failure of long-run weak exogeneity. This example corresponds to condition (b) of §4.2 (variation free) not holding.

4.5. Distributions of estimators and tests

Notwithstanding the analyses in the previous subsections, even for $\mathsf{I}(0)$ variables [29] show the potential impact of the presence of lagged feedbacks of $\mathbf{X}_{1,t-1}$ onto $\mathbf{x}_{2,t}$ when estimating relationships between $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ if the errors on the relation are unknowingly autocorrelated. The resulting correlation induced between the conditioning variables and the errors entails that estimators are inconsistent for the parameters of interest and have inconsistently estimated standard errors, leading to invalid interpretations of test statistic outcomes.

It is now well established that different distributions are required for estimators and tests in a variety of settings for $\mathsf{I}(1)$ variables. However, the forms these distributions take are dependent on the weak exogeneity status of the variables in the analysis (see Peter Phillips and Mico Loretan, 1991, [44]), as well as on cointegration, what deterministic terms enter the DGP, and how they are treated in the cointegration model: see e.g., Søren Johansen (1996) [37] and [25].

4.6. Simulation-based inference and the bootstrap

Similarly, simulation based estimation of dynamic latent-variables models as in Daniel McFadden (1989) [40], and Ariel Pakes and David Pollard (1989) [43] for example, and bootstrap-derived distributions in dynamic models (see *inter alia*, Bradley Efron, 1979, [12] and Efron and Robert Tibshirani, 1993, [13]) are both affected by the existence of lagged feedbacks from $\mathbf{X}_{1,t-1}$ to $\mathbf{x}_{2,t}$. To treat unmodelled variables as 'fixed in repeated samples' requires them to be strongly exogenous, and if not, unless the joint distribution is simulated, invalid outcomes will result.

4.7. Forecasting

Granger's contributions to the theory and practice of economic forecasting are discussed extensively in the paper by Michael P. Clements: here we only consider the impact of Granger causality on forecasts. Thus, attention turns to the opposite of EGNC, so lagged information about one set of variables does alter the outcomes of a second set.

Despite its poor reputation, economic forecasting remains a key task for most policy agencies internationally, and more generally the need for forecasts includes those concerned with demographics, health, and climate among other disciplines. Returning to the formulation in (7), whether or not (8) holds determines the validity of forecasts of $\mathbf{x}_{1,T+h}$ for $h > 1$ conditional on given future values of $\{\mathbf{x}_{2,T+h}\}$. Expressing the joint distribution on the left-hand side of (7) as a correctly-specified first-order vector autoregression, and assuming for the moment that the parameterization is constant, then:

$$\mathbf{x}_t = \gamma + \mathbf{\Gamma}\mathbf{x}_{t-1} + \mathbf{u}_t \quad \text{where} \quad \mathbf{u}_t \sim \mathsf{IN}\left[\mathbf{0}, \mathbf{\Omega}\right] \tag{9}$$

where $\mathsf{IN}\left[\mathbf{0}, \mathbf{\Omega}\right]$ denotes an independent normal random variable with the given mean and variance, being identically distributed when both moments are constant, and all the eigenvalues λ_i of $\mathbf{\Gamma}$ lie inside the unit circle, with the estimated model:

$$\widehat{\mathbf{x}}_t = \widehat{\gamma} + \widehat{\mathbf{\Gamma}}\mathbf{x}_{t-1} \tag{10}$$

This formulation can be generalized to longer lags, $\mathsf{I}(1)$ variables, mis-specification and more complicated deterministic terms, but suffices to illustrate the analysis. Then a sequence of 1-step ahead forecasts from time T (the forecast origin) is given by:

$$\widehat{\mathbf{x}}_{T+h|T+h-1} = \widehat{\gamma} + \widehat{\mathbf{\Gamma}}\mathbf{x}_{T+h-1} \quad h = 1, \ldots, H \tag{11}$$

At each h, the past value $\mathbf{x}_{T+h-1}$ is already known, so obtaining $\left\{\widehat{\mathbf{x}}_{T+h|T+h-1}\right\}$ does not require knowledge of GNC or EGNC between the components $\mathbf{x}_1$ and $\mathbf{x}_2$.

There are two ways to produce multi-step forecasts, namely unconditionally or conditionally on $\{\mathbf{x}_{2,T+h}\}$. The former merely iterates (11):

$$\begin{aligned}\widehat{\mathbf{x}}_{T+H|T} &= \widehat{\gamma} + \widehat{\mathbf{\Gamma}}\widehat{\mathbf{x}}_{T+H-1|T} = \widehat{\gamma} + \widehat{\mathbf{\Gamma}}\left(\widehat{\gamma} + \widehat{\mathbf{\Gamma}}\widehat{\mathbf{x}}_{T+H-2|T}\right) = \cdots \\ &= \sum_{h=1}^{H}\widehat{\mathbf{\Gamma}}^{h-1}\widehat{\gamma} + \widehat{\mathbf{\Gamma}}^{H}\mathbf{x}_T \end{aligned} \tag{12}$$

so does not depend on whether or not (8) holds.

Sometimes multi-step forecasts are made conditional on a sequence of future values denoted here by $\{\overline{\mathbf{x}}_{2,T+h}, h = 1, \ldots, H\}$. To formulate this approach, factorize $\mathbf{x}_t$ in (9) into $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$, and similarly for (11):

$$\begin{pmatrix} \widehat{\mathbf{x}}_{1,T+h|T+h-1} \\ \widehat{\mathbf{x}}_{2,T+h|T+h-1} \end{pmatrix} = \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} + \begin{pmatrix} \widehat{\mathbf{\Gamma}}_{11} & \widehat{\mathbf{\Gamma}}_{12} \\ \widehat{\mathbf{\Gamma}}_{21} & \widehat{\mathbf{\Gamma}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1,T+h-1} \\ \mathbf{x}_{2,T+h-1} \end{pmatrix} \tag{13}$$

so denoting forecasts given $\{\overline{\mathbf{x}}_{2,T+h}\}$ by $\widetilde{\mathbf{x}}_{1,T+H|T}$:

$$\begin{aligned} \widetilde{\mathbf{x}}_{1,T+H|T} &= \widehat{\gamma}_1 + \widehat{\mathbf{\Gamma}}_{12}\overline{\mathbf{x}}_{2,T+H-1} + \widehat{\mathbf{\Gamma}}_{11}\widetilde{\mathbf{x}}_{1,T+H-1|T} \\ &= \sum_{h=1}^{H} \widehat{\mathbf{\Gamma}}_{11}^{h-1}\left(\widehat{\gamma}_1 + \widehat{\mathbf{\Gamma}}_{12}\overline{\mathbf{x}}_{2,T+H-h}\right) + \widehat{\mathbf{\Gamma}}_{11}^{H}\mathbf{x}_{1,T} \end{aligned} \tag{14}$$

as $\{\overline{\mathbf{x}}_{2,T+H-h}\}$ is taken as known. Then (14) will be more accurate than (12) provided $\overline{\mathbf{x}}_{2,T+h} = \mathbf{x}_{2,T+h}$ which requires that (8) holds, as the additional forecast errors from $\{\widehat{\mathbf{x}}_{2,T+H-h}\}$ are not cumulated. To simplify the algebra, assume parameter estimation uncertainty can be ignored, so the forecast errors from (12) are then given by $\widehat{\mathbf{u}}_{T+H|T} = \mathbf{x}_{T+H} - \widehat{\mathbf{x}}_{T+H|T}$:

$$\widehat{\mathbf{u}}_{T+H|T} = \sum_{h=0}^{H-1} \mathbf{\Gamma}^{h}\mathbf{u}_{T+H-h} \tag{15}$$

whereas when $\overline{\mathbf{x}}_{2,T+h} = \mathbf{x}_{2,T+h}$ those from (14) are $\widetilde{\mathbf{u}}_{1,T+H|T} = \mathbf{x}_{1,T+H} - \widetilde{\mathbf{x}}_{1,T+H|T}$:

$$\widetilde{\mathbf{u}}_{1,T+H|T} = \sum_{h=0}^{H-1} \mathbf{\Gamma}_{11}^{h}\mathbf{u}_{1,T+H-h} \tag{16}$$

thereby eliminating the components from both the variances and covariances of $\mathbf{u}_{2,T+H-h}$.

However, the *forecasts* made by (14) do not depend on (8) holding, although their accuracy clearly does. When there is Granger causality from $\mathbf{x}_1$ to $\mathbf{x}_2$, as the former changes, so must the latter and the assumption that $\overline{\mathbf{x}}_{2,T+h} = \mathbf{x}_{2,T+h}$ cannot hold. Consequently, (16) should be:

$$\widetilde{\mathbf{u}}_{1,T+H|T} = \sum_{h=0}^{H-1} \mathbf{\Gamma}_{11}^{h}\left(\mathbf{\Gamma}_{12}\left[\mathbf{x}_{2,T+h} - \overline{\mathbf{x}}_{2,T+h}\right] + \mathbf{u}_{1,T+H-h}\right) \tag{17}$$

which will be biased and could have a larger mean-square forecast error (MSFE) than $\widehat{\mathbf{u}}_{T+H|T}$ in (15).

An important additional consideration that has so far remained undiscussed is that of location shifts, namely changes in previous unconditional means of $\mathsf{I}(0)$ data transformations—often unanticipated as with the Financial Crisis of 2008. Such shifts are pernicious sources of forecast failure, as shown in [6]. Their absence is therefore highly pertinent to successful forecasting, specifically avoiding forecast failure. Indeed, the intermittent occurrence of location shifts is probably the main reason for the poor forecasting record of econometric systems. Return to the DGP underlying (13), now subscripting the intercepts:

$$\begin{pmatrix} \mathbf{x}_{1,t} \\ \mathbf{x}_{2,t} \end{pmatrix} = \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{pmatrix} + \begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1,t-1} \\ \mathbf{x}_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{1,t} \\ \mathbf{u}_{2,t} \end{pmatrix} \tag{18}$$

The simplest setting is one where the $\gamma_{i,t}$ are constant till times $1 < T_i < T$ and then change to another value. When the T_i are close to the forecast origin T, then forecasts based on the previous intercepts will be systematically incorrect. Because intercepts depend on the units of measurement, sometimes including when price indexes are based, large errors can occur.

Even if such shifts occur well inside the sample period, problems can occur in establishing EGNC. Consider a DGP where $\mathbf{\Gamma}_{12} = \mathbf{0}$ and $\mathbf{\Gamma}_{21} = \mathbf{0}$ so neither variable Granger causes the other, but an investigator assumes the γ_i are constant $\forall t$ when in fact $\gamma_{1,t} = \kappa\gamma_{2,t}$. Then $\mathbf{x}_{2,t-1}$ may well be significant when added to the first block as it acts as a proxy for the connection between the intercepts since:

$$\gamma_{1,t} = \kappa\gamma_{2,t} = \kappa\left(\mathbf{x}_{2,t} - \Gamma_{22}\mathbf{x}_{2,t-1} - \mathbf{u}_{2,t}\right) \tag{19}$$

Connections between parameters, as in $\gamma_{1,t} = \kappa\gamma_{2,t}$, are generally bad news because both weak and super exogeneity are violated, but they can also lead to co-breaking as a combination of the variables does not shift even though the marginal process shifts: see Hendry and Michael Massmann (2007) [28]. Here, from (18), the combination $(\mathbf{x}_{1,t} - \kappa\mathbf{x}_{2,t})$ will cancel the shift as $\gamma_{1,t} - \kappa\gamma_{2,t} = \mathbf{0}$. Thus, modelling of the co-breaking relation, which deliberately combines the parameters of the two sub-blocks in (18), will deliver a model that does not change as $\gamma_{2,t}$ changes, so the resulting conditional model will end being valid and invariant, and the combination $(\mathbf{x}_{1,t} - \kappa\mathbf{x}_{2,t})$ can be reasonably forecast. However, $\mathbf{x}_{2,t}$ will remain difficult to forecast unless a model of $\gamma_{2,t}$ can be developed.

4.8. Dynamic simulation

The discussion in the previous subsection bears directly on this and the next subsections. Dynamic simulation is essentially in-sample multi-step forecasting. [32] demonstrated its invalidity when variables that are taken as given in the simulation, such as $\{\overline{\mathbf{x}}_{2,t}\}$ above, are in fact Granger caused by the variables being simulated.[§]

Returning to (14), but now interpreted as being in-sample over $t = 1, \ldots, T$ from the initial state $\mathbf{x}_{1,0}$, then a conditional dynamic simulation delivers:

$$\widetilde{\mathbf{x}}_{1,t|0} = \sum_{j=1}^{t} \widehat{\mathbf{\Gamma}}_{11}^{j-1}\left(\widehat{\gamma}_1 + \widehat{\mathbf{\Gamma}}_{12}\overline{\mathbf{x}}_{2,t-j}\right) + \widehat{\mathbf{\Gamma}}_{11}^{t}\mathbf{x}_{1,0} \tag{20}$$

Under the earlier assumption of weak stationarity, $\left|\widehat{\lambda}_i\right| < 1 \forall i$, then $\widehat{\mathbf{\Gamma}}_{11}^{t} \to \mathbf{0}$ as t increases. Hence, the 'explanation' of $\widetilde{\mathbf{x}}_{1,t|0}$ becomes attributed to $\{\overline{\mathbf{x}}_{2,t-j}\}$, irrespective of GNC or EGNC. When $\mathbf{\Gamma}_{21} \neq \mathbf{0}$, changes in $\mathbf{x}_{1,t}$ alter $\mathbf{x}_{2,t}$ invalidating the claimed conditional outcome.

This problem is especially serious when dynamic tracking is used to select between alternative model specifications. For example, as noted above, formulations like (14) can be more accurate for $\{\mathbf{x}_{1,t}\}$ than modelling it in the entire system in (12) when $\mathbf{\Gamma}_{21} \neq \mathbf{0}$.

[§]Incidentally, this paper was selected by Clive Granger to be reprinted in his edited volume *Modelling Economic Series*, Oxford: Clarendon Press, 1990.

Indeed, from (12):

$$\widehat{\mathbf{x}}_{t|0} = \sum_{j=1}^{t} \widehat{\mathbf{\Gamma}}^{t-1}\widehat{\gamma} + \widehat{\mathbf{\Gamma}}^{t}\mathbf{x}_0 \rightarrow \left(\mathbf{I} - \widehat{\mathbf{\Gamma}}\right)^{-1}\widehat{\gamma} \tag{21}$$

so converges on the unconditional mean and appears not to track, whereas (20) continues to show variation in $\widetilde{\mathbf{x}}_{1,t|0}$ that will seem to track. Moreover, an incorrect model specification of the form, for $L \geq 2$:

$$\mathbf{x}_{1,t} = \gamma_1 + \sum_{l=1}^{L} \mathbf{\Gamma}_{12,l}\mathbf{x}_{2,t-l} + \varsigma_t \tag{22}$$

when dynamically simulated treating all $\{\mathbf{x}_{2,t-l}\}$ as if they were known, can appear to track $\mathbf{x}_{1,t|0}$ better still. Thus, false claims as to either or both GNC or EGNC can lead to selecting incorrect model forms. Problems like those highlighted by (18) where parameter changes are not modelled, or co-breaking does not occur, exacerbate the difficulties of producing valid dynamic simulations.

4.9. Policy analysis

When a policy instrument, $\{z_t\}$ say, is changed in an attempt to alter the future value of some target variable, here denoted by $\{y_t\}$, then the new value chosen for that instrument usually depends on a prior quantitative policy analysis. To deliver an appropriate outcome and avoid the problem highlighted by (17), any analyses extending beyond 1-period ahead conditional on a trajectory for z require that y does not Granger cause z, whereas z must actually affect y, possibly indirectly, if the policy change is to be effective. §4.8 showed that no models should be selected by dynamic simulation properties, and [30] explain why a policy model should also not be selected by its forecasting performance.

Moreover, changes to a policy instrument almost inevitably involve location shifts, where the level of the instrument is changed from z_t to $z_t + \delta$ where $\delta \neq 0$. Consequently, another essential requirement is that the policy shift does not alter the previous relationship between $\{y_t\}$ and $\{z_t\}$. Using (18) as the example and interpreting $\mathbf{x}_{1,t}$ as y_t and $\mathbf{x}_{2,t}$ as z_t, then the link $\gamma_{1,t} = \kappa\gamma_{2,t}$ entails a failure of parameter invariance, as $z_t + \delta$ will both directly affect y_t with a lag when $\mathbf{\Gamma}_{12} \neq \mathbf{0}$, but also shift the intercept in that relation by $\kappa\delta$. Super exogeneity of z_t for the parameters of the relation between y_t and z_t as in [15] is needed, and building on Hendry and Carlos Santos (2010) [33], a step-indicator saturation test is proposed by Castle, Hendry, and Andrew Martinez (2016) [5] that could reveal failures of invariance before a policy is incorrectly implemented.

4.10. Impulse-response analyses

Neil Ericsson, Hendry and Mizon [16] and [29] provide extensive analyses of the problems of evaluating policy by using impulse-response analyses—as proposed by Christopher Sims (1980) [47]—and of the role of EGNC in their resulting properties. Again we use

(18) as an $\mathsf{I}(0)$ example, written explicitly as the bivariate Normal DGP for $(y_t,\, z_t)$:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \gamma_{1,t} + \Gamma_{11} y_{t-1} + \Gamma_{12} z_{t-1} \\ \gamma_{2,t} + \Gamma_{21} y_{t-1} + \Gamma_{22} z_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \tag{23}$$

where:

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \right] = \mathsf{IN}_2 \left[\mathbf{0}, \mathbf{\Omega} \right].$$

The impact trajectory over time of $\{y_t\}$ from an 'impulse' to z_t is the objective of the policy analysis. However, when $\omega_{12} \neq 0$, an impulse to z_t is not unique, so arbitrary 'Choleski' factorizations of $\mathbf{\Omega}$ are first undertaken. Using the ordering $(y_t,\, z_t)$ in (23), implementing such a conditional-marginal factorization and taking into acount that $\gamma_{1,t} = \kappa\gamma_{2,t}$ leads to:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \beta z_t + (\kappa - \beta)\, \gamma_{2,t} + (\Gamma_{11} - \beta\Gamma_{21})\, y_{t-1} + (\Gamma_{12} - \beta\Gamma_{22})\, z_{t-1} \\ \gamma_{2,t} + \Gamma_{21} y_{t-1} + \Gamma_{22} z_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t}|u_{2,t} \\ u_{2,t} \end{pmatrix} \tag{24}$$

where $\beta = \omega_{12}/\omega_{22}$ is the population regression parameter and:

$$\begin{pmatrix} u_{1,t}|u_{2,t} \\ u_{2,t} \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{11} - \beta\omega_{12} & 0 \\ 0 & \omega_{22} \end{pmatrix} \right].$$

The literature then considers how a shock to $u_{2,t}$ will affect $\{y_{t+k}\}$ for $k \geq 0$.

There are a number of potential problems as follows. Impulse responses describe the dynamic properties of an estimated model, and not necessarily the dynamic characteristics of the variables being analyzed. Even when a model is well-specified and invariant to extensions of the information set, its residuals can only coincide with the DGP errors if the model coincides with the DGP, and otherwise residuals are a sign of 'ignorance not knowledge', so could reflect measurement errors and specification mistakes *inter alia*. Next, perturbation of z_t could come from shifts to $\gamma_{2,t}$ or impulses in $u_{2,t}$, and which generates the response is not identifiable from (24) since z_t is a linear function of both. However, their effects on y_t need not be the same, and indeed will differ unless $\kappa = \beta$, which is a necessary condition for the weak exogeneity of z_t for the parameters of the conditional equation in (24). The converse Choleski factorization, which is often also reported, must violate weak exogeneity if the first is correct, so an incorrect reaction will then be inferred. Finally, when $\kappa \neq \beta$, a location shift to z_t changing $\gamma_{2,t}$ to $\gamma_{2,t} + \delta$ will violate super exogeneity, and the policy response based on assuming $\partial y_t / \partial z_t = \beta$ and then tracing the dynamic responses will not match what happens.

5. Conclusion

Clive Granger (1969) [18] proposed an approach to testing for causality based on the arrow of time and the effects of eliminating the complete history of some variables from the joint distribution of the universe of all variables. There was no Granger causality from the eliminated variables to the remaining variables if their conditional and marginal distributions were the same. In practice, the non-operational nature of his definition was finessed by testing whether adding or dropping a lagged subset of variables from a larger set affected the goodness of fit of models of the remaining variables. This route to making his concept of causality operational has major drawbacks and could not be called causality in any useful sense without knowing that the variables in the analysis comprised the universe of relevant information. Notwithstanding such drawbacks in its interpretation, and often far from the original intent of his concept, Granger non-causality and its empirical equivalent play pervasive roles in ten important areas of econometric modelling of time series, attested by the large number of citations to his formulation.

Acknowledgements

Financial support from the Robertson Foundation (award 9907422), Institute for New Economic Thinking (grant 20029822) and Statistics Norway (through Research Council of Norway Grant 236935), are all gratefully acknowledged, as are helpful comments on an earlier draft by Jennifer L. Castle, Michael P. Clements, Andrew B. Martinez, Felix Pretis and Timo Teräsvirta. Email: david.hendry@nuffield.ox.ac.uk.

References

[1] Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Chichester: John Wiley.

[2] Bazinas, V. and B. Nielsen (2014). Exogeneity and causal transmission in reduced-form models. Discussion paper, Nuffield College, Oxford.

[3] Bontemps, C. and G. E. Mizon (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics 70*, 721–750.

[4] Cartwright, N. (2003). Two theorems on invariance and causality. *Philosophy of Science 70*, 203–224.

[5] Castle, J. L., D. F. Hendry, and A. B. Martinez (2016). Policy Analysis, Forediction, and Forecast Failure. Discussion paper, Economics Department, Oxford University.

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

[7] Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, Berkeley, pp. 105–123. University of California Press.

[8] Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B 24*, 406–424.

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

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

[11] Doob, J. L. (1953). *Stochastic Processes.* New York: John Wiley Classics Library. 1990 edition.

[12] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics 7*, 1–26.

[13] Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap.* London: Chapman and Hall.

[14] Engle, R. F. and C. W. J. Granger (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–276.

[15] Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

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

[17] Govaerts, B., D. F. Hendry, and J.-F. Richard (1994). Encompassing in stationary linear dynamic models. *Journal of Econometrics 63*, 245–270.

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

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

[20] Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics 48*, 213–228.

[21] Granger, C. W. J. (1988). Some recent development in a concept of causality. *Journal of econometrics 39*, 199–211.

[22] Harré, R. and E. H. Madden (1975). *Causal Powers.* Oxford: Blackwell.

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

[24] Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology 35*, 1–98.

[25] Hendry, D. F. (1995). On the interactions of unit roots and exogeneity. *Econometric Reviews 14*, 383–419.

[26] Hendry, D. F. (2004). Causality and exogeneity in non-stationary economic time series. In A. Welfe (Ed.), *New Directions in Macromodelling*, pp. 21–48. Amsterdam: North Holland.

[27] Hendry, D. F. and J. A. Doornik (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge, Mass.: MIT Press.

[28] Hendry, D. F. and M. Massmann (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics 25*, 33–51.

[29] Hendry, D. F. and G. E. Mizon (1999). The pervasiveness of Granger causality in econometrics. In R. F. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp. 102–134. Oxford: Oxford University Press.

[30] Hendry, D. F. and G. E. Mizon (2000). On selecting policy analysis models by forecast accuracy. In A. B. Atkinson, H. Glennerster, and N. Stern (Eds.), *Putting Economics to Work: Volume in Honour of Michio Morishima*, pp. 71–113. London School of Economics: STICERD.

[31] Hendry, D. F. and G. E. Mizon (2016). Improving the teaching of econometrics. *Cogent Economics and Finance*, DOI: 0.1080/23322039.2016.1170096. http://www.tandfonline.com/eprint/7dRymBhRtIQF2xtbvPMC/full.

[32] Hendry, D. F. and J.-F. Richard (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics 20*, 3–33.

[33] Hendry, D. F. and C. Santos (2010). An automatic test of super exogeneity. In M. W. Watson, T. Bollerslev, and J. Russell (Eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

[34] Hicks, J. R. (1979). *Causality in Economics*. Oxford: Basil Blackwell.

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

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

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

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

[39] Mackie, J. L. (1980). *The Cement of the Universe: A Study in Causation*. Oxford: Oxford University Press. 2nd edition.

[40] McFadden, D. L. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

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

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

[43] Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimation. *Econometrica 57*, 1027–1058.

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

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

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

[47] 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.

[48] Smith, A. (1759). *Theory of Moral Sentiments*. Edinburgh: A. Kincaid & J. Bell.

[49] Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan & T. Cadell.

[50] Smith, A. (1795). The history of astronomy. In D. Stewart (Ed.), *Essays on Philosophical Subjects by Adam Smith*, pp. 33–105. Edinburgh: W. Creech. Liberty Classics edition, by I. S. Ross, 1982.

[51] Zhang, K., J. Zhang, and B. Schölkopf (2015). Distinguishing cause from effect based on exogeneity. *http://arxiv.org/abs/1504.05651*.