

# Typologies of Linear Dynamic Systems and Models

David F. Hendry\*  
Nuffield College, Oxford, U.K.

## Abstract

Open and closed linear dynamic systems are formulated and considered for stationary and integrated data processes. A typology of linear dynamic systems is developed, extending that available for individual dynamic equations. Then an overlapping typology of models of these systems is examined and methods for analyzing econometric models are described. General to simple modelling strategies are briefly considered.

Running head:

## Dynamic Typologies

Key Words: linear dynamic systems; cointegrated processes; model typology; econometric modelling.

AMS Classification Numbers: 62 M10, 62 P 20, 90 A 16

Address for correspondence:

David F. Hendry,  
Nuffield College,  
Oxford, OX1 1NF,  
UK.

---

\*This research was financed in part by grants R000233447 and R231184 from the UK Economic and Social Research Council. Helpful comments and suggestions from Mike Clements, Søren Johansen, Grayham Mizon, V.K. Srivastava and two referees are gratefully acknowledged.

# 1 Introduction

Economic behaviour is inherently dynamic for many reasons including inertia, expectations, inter-temporal optimization, stock-flow relations, time to produce, and smoothing to mitigate the impact of unanticipated events. Economists seek to explain observed outcomes using dynamic systems which characterize the joint distributions of the variables under analysis. A dynamic system is any time-dependent representation of the variables to be explained in terms of available information. Spanos (1986) refers to the system as the Haavelmo distribution, after Haavelmo (1944). Often economic theory induces a formulation of the economic mechanism in terms of sets of simultaneous equations which impose restrictions on the dynamic system, and we refer to any such representation as a model. A system is therefore a model of itself and it is well known that all models which impose no over-identifying restrictions on a common set of variables correspond to the same system.

This paper formulates and analyzes a double typology of linear dynamic systems and models thereof, focusing on the implications for empirical modelling of economic time series. The former builds on the typology in Hendry and Richard (1983) and Hendry, Pagan and Sargan (1984), extending their analysis for individual dynamic equations to linear systems. It highlights the basic properties of certain system types and cautions against assuming any specific form without investigating its implications. The latter aims to clarify relations between such ostensibly different model forms as vector autoregressions, vector error correction systems, simultaneous equations, reduced forms, conditional models, causal chains, block-recursive representations and triangular systems. It is shown that in a linear formulation, when a common information set is given throughout, all the available model types impose testable restrictions on the dynamic system. The analysis is conducted for closed, open, complete and incomplete systems for both stationary and integrated time series (i.e.  $I(0)$ ,  $I(1)$  and  $I(2)$  data, where  $I(d)$  denotes a series which is integrated of order  $d$ , and so is stationary only after differencing exactly  $d$  times). We consider what hypothetical experiments are feasible in each model form, and hence how each model type can be analyzed. Thus, the paper is also a follow-up to Hendry, Neale and Srba (1988), Hendry and Ericsson (1991), Hendry and Mizon (1993), Doornik and Hendry (1994) and Hendry and Doornik (1994). The introduction defines the notation and introduces the general linear dynamic system.

It is desired to model a vector of  $n$  observable real random variables  $\mathbf{x}_t = (x_{1t} \dots x_{nt})'$  defined on a probability space  $(\Omega, \mathcal{F}, P)$  and characterized by the continuous joint data density function  $D_X(\cdot)$  for  $T$  observations:

$$D_X(\mathbf{x}_1 \dots \mathbf{x}_T \mid \mathbf{X}_0, \mathbf{Q}_T^1, \boldsymbol{\theta}) \quad \text{where } \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k \quad (1)$$

when  $\mathbf{X}_0$  denotes the (pre-sample) initial conditions,  $|$  separates the outcomes from the conditioning set,  $\mathbf{Q}_T^1 = (\mathbf{q}_1 \dots \mathbf{q}_T)$  is a set of  $m$  deterministic conditioning variables (such as constant, seasonals, trend and dummy variables) and there are  $k$  parameters in  $\boldsymbol{\theta}$  with a parameter space  $\Theta$  which is a subset of  $k$ -dimensional real space  $\mathbb{R}^k$ .

Denote the history of  $\{\mathbf{x}_t\}$  by  $\mathbf{X}_{t-1} = (\mathbf{X}_0 : \mathbf{X}_{t-1}^1)$  where  $\mathbf{X}_{t-1}^1 = (\mathbf{x}_1 \dots \mathbf{x}_{t-1})$  and  $\mathbf{X}_{t-1}$  is a shorthand for the sigma-field generated by the past of the process under study. Write the sequential conditional density of the outcome  $\mathbf{x}_t$  at time  $t$  as  $D_X(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{Q}_T^1, \boldsymbol{\theta})$ , then from (1) for the complete sample  $\mathbf{X}_T^1 = (\mathbf{x}_1 \dots \mathbf{x}_T)$ :

$$D_X(\mathbf{X}_T^1 \mid \mathbf{X}_0, \mathbf{Q}_T^1, \boldsymbol{\theta}) = \prod_{t=1}^T D_X(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{Q}_T^1, \boldsymbol{\theta}). \quad (2)$$

Because of the dynamic nature of an economy, no assumption of independence is made about the successive drawings of  $\{\mathbf{x}_t\}$ . Nevertheless, the factorization in (2) generates a mean innovation process defined by (see Hendry, 1995):

$$\boldsymbol{\epsilon}_t = \mathbf{x}_t - E[\mathbf{x}_t \mid \mathbf{X}_{t-1}].$$

The forms of  $D_X(\cdot)$  and  $D_X(\cdot)$  could be very different: for example, when autoregressive heteroscedastic (ARCH) errors are present (see Engle, 1982),  $D_X(\cdot)$  could be a normal density although  $D_X(\cdot)$  is not a multivariate normal.

In (2),  $\boldsymbol{\theta}$  is allowed to contain transient parameters, namely parameters of dummy variables (perhaps interactive with lagged elements of  $\mathbf{x}_t$ ). The parameters of interest to the investigator comprise a subset  $\boldsymbol{\psi} \in \Psi$  of

$\lambda = \mathbf{g}(\boldsymbol{\theta}) \in \Lambda$  where  $\mathbf{g}(\cdot)$  is a 1-1 function of  $\boldsymbol{\theta}$ . Since:

$$D_X(\mathbf{X}_T^1 | \mathbf{X}_0, \mathbf{Q}_T^1, \boldsymbol{\theta}) = D_X(\mathbf{X}_T^1 | \mathbf{X}_0, \mathbf{Q}_T^1, \lambda) \quad (3)$$

then  $D_X(\cdot)$  is invariant under 1-1 parameter transformations and the resulting class of densities is isomorphic. Such a result is regularly used in derivations, as when a variance or a standard deviation is thought to be the more useful parameter to consider. A major objective of econometric modelling is to discover and estimate sustainable relationships between economic magnitudes. The parameter  $\psi$  is assumed to be structural in that it is invariant and directly characterizes the relations of the economy under analysis (see Hendry, 1993). Consequently,  $\psi$  must be constant over time, invariant to policy interventions, and be unaltered by extensions of the information set.

It is unclear how it could be known *a priori* that the elements of  $\mathbf{Q}_T^1$  are valid conditioning variables when more than just a constant and linear trend are included (e.g. dummy variables for policy changes) and we assume that no such difficulty arises in the following analysis. However, we only require conditioning on  $\mathbf{q}_t$  to be valid, which is a weaker (albeit not unexceptional) assumption, in that the following reduction is valid and is sufficient for modelling:

$$D_X(\mathbf{x}_t | \mathbf{X}_{t-1}, \mathbf{Q}_T^1, \boldsymbol{\theta}) = D_X(\mathbf{x}_t | \mathbf{X}_{t-1}, \mathbf{q}_t, \boldsymbol{\theta}) \quad (4)$$

Thus, only contemporaneous information on the deterministic terms is used. It might be thought that by their deterministic nature, future values of  $\{\mathbf{q}_t\}$  are known with certainty. This is certainly true of polynomials in time, but not of (say) dummy variables for special events (e.g. the oil crisis, a government's political persuasion etc.). When conducting dynamic analyses of systems (e.g. dynamic simulation), the treatment of such dummies can play a crucial role in the ostensible performance. Equally, as shown below, the way  $\mathbf{q}_t$  enters the system is important for its properties and for parameter estimators (see e.g. Johansen, 1988).

Finally, a maximum lag length or memory of  $s$  periods is imposed for tractability: it may prove possible in practice to obtain more parsimonious representations using moving average components but that aspect will not be considered here. Combining all of these ingredients, we represent the data generating process (DGP) as:

$$D_X(\mathbf{X}_T^1 | \mathbf{X}_0, \mathbf{Q}_T^1, \boldsymbol{\theta}) = \prod_{t=1}^T D_X(\mathbf{x}_t | \mathbf{X}_{t-1}^{t-s}, \mathbf{q}_t, \boldsymbol{\theta}) \quad (5)$$

When  $D_X(\cdot)$  is close to joint normality, the resulting system is a vector autoregressive representation (VAR), but no issue of principle is involved in extending this to have vector moving average errors (see e.g. Lütkepohl, 1991).

First, we note four major categorizations common to all such models, namely whether the model is open or closed (meaning that all variables are modelled), whether it is complete or incomplete (i.e. whether there are as many equations as variables to be modelled), whether the model is linear or non-linear and whether the data are stationary  $I(0)$  or are integrated of order 1 or 2 ( $I(1)$  or  $I(2)$ ). Other forms of non-stationarity (such as regime shifts) are assumed to be adequately accounted for by the introduction of transient parameters and the presence of  $\mathbf{q}_t$ . We are mainly concerned with systems which are linear in the variables, although it is recognized that the generic form of model in macro-econometrics must be non-linear due to linear accounting identities, multiplicative price, quantity and value relations and (usually) log-linear behavioural equations. Further, we will only analyze complete models, taking incomplete models as obtainable by omitting some of the equations while retaining knowledge that the associated variables are endogenous to the system and are determined by a subset of non-modelled or lagged variables (see Hendry and Richard, 1983). Finally, the  $I(0)$  formulation is the general one, with  $I(1)$  or  $I(2)$  being obtained as restrictions on the basic form, as shown below (see Johansen, 1992a).

Many possible system forms could be adopted to characterize the behaviour of the  $\{\mathbf{x}_t\}$  and learn about  $\psi$ . To formulate these, we next extend the single equation typology in Hendry and Richard (1983) to systems, briefly commenting on the styles of system that result. Providing a common information set is maintained, a clear 'ordering' in complexity, or nesting, between types results: given the contents of the vector  $\mathbf{x}_t$ , the VAR embeds all conditional systems (see Hendry and Mizon, 1993).

Finally, we develop the analysis specifically for different model representations of the same system, including error-correction models, open and closed simultaneous models, conditional models, causal chains, block-recursive models and triangular representations. The relations between these 'rival' model forms are examined and related to the properties of the assumed DGP. The two typologies overlap to the extent that some systems are their own model.

## 2 Closed linear systems

### 2.1 Formulation

Given the restriction of the analysis to linear models, the basic closed form of  $D_x(\mathbf{x}_t|\cdot)$  in (5) is a VAR. The second typology below will consider special cases of this formulation. Assuming that normality is a good approximation to the shape of  $D_x(\cdot)$ , consistent with linearity of conditional models:

$$\mathbf{x}_t = \sum_{j=1}^s \mathbf{A}_j \mathbf{x}_{t-j} + \mathbf{F} \mathbf{q}_t + \boldsymbol{\nu}_t \text{ where } \boldsymbol{\nu}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}]. \quad (6)$$

In (6),  $\boldsymbol{\nu}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}]$  denotes an  $n$ -dimensional independent normal density with mean zero and covariance matrix  $\boldsymbol{\Omega}$  assumed to be (symmetric) positive definite. Clearly,  $\boldsymbol{\nu}_t$  is also a mean innovation process. Despite the presence of deterministic (non-modelled) variables in (6), we refer to it as a closed model. To keep the notation simple, we have not explicitly allowed for interactions between the  $\mathbf{q}_t$  and lagged  $\{\mathbf{x}_{t-j}\}$ , but have assumed that the parameters  $(\mathbf{A}_1 \dots \mathbf{A}_s, \mathbf{F}, \boldsymbol{\Omega})$  are constant in the postulated representation. In practice, interactions such as  $q_{it}x_{j,t-l}$  may be used, especially when  $q_{it}$  is a regime shift dummy, so that  $x_{j,t-l}$  is zero for a proportion of the sample, but little seems to be gained in a theoretical treatment by their retention. Then relative to (1),  $k = n(ns + m + \frac{1}{2}(n+1))$  so that even for (e.g.) 4 equations with 2 lags and 2 dummies,  $k = 50$ ; for 8 equations with 4 lags and 5 dummy variables,  $k = 332$  and hence the dimensionality of the parameter space rises rapidly. This curse of dimensionality is a major driving force behind the desire to develop more parsimonious representations of the initial specification as discussed in section [5].

### 2.2 Invariance under linear transformations

We now consider the set of non-singular linear transformations of (6) which are isomorphic in that they retain the same basic innovation process  $\{\boldsymbol{\nu}_t\}$  and hence the same likelihood. This also generates the class of all equivalent models of (6). We take  $s = 2$  and  $m = 0$  for simplicity, so the system can be written as:

$$\mathbf{x}_t = \mathbf{A}_1 \mathbf{x}_{t-1} + \mathbf{A}_2 \mathbf{x}_{t-2} + \boldsymbol{\nu}_t \quad (7)$$

Let  $\mathbf{f}'_t = (\mathbf{x}'_t : \mathbf{x}'_{t-1} : \mathbf{x}'_{t-2})$  and rewrite (7) as:

$$\mathbf{G} \mathbf{f}_t = \boldsymbol{\nu}_t \text{ where } \mathbf{G} = (\mathbf{I} : -\mathbf{A}_1 : -\mathbf{A}_2). \quad (8)$$

Longer lags merely extend  $\mathbf{f}_t$ . Two types of linear transform are feasible. Let  $\mathbf{M}$  and  $\mathbf{P}$  be known  $n \times n$  and  $\ell \times \ell$  non-singular matrices where  $\ell = (s+1)n$  and consider:

$$\mathbf{M} \mathbf{G} \mathbf{P}^{-1} \mathbf{P} \mathbf{f}_t = \mathbf{M} \boldsymbol{\nu}_t \text{ or } \mathbf{G}^* \mathbf{f}_t^* = \boldsymbol{\nu}_t^* \quad (9)$$

where  $\mathbf{f}_t^* = \mathbf{P} \mathbf{f}_t$  and  $\boldsymbol{\nu}_t^* \sim \text{IN}(\mathbf{0}, \mathbf{M} \boldsymbol{\Omega} \mathbf{M}')$  with:

$$\mathbf{G}^* = \mathbf{M} \mathbf{G} \mathbf{P}^{-1} = (\mathbf{M} : -\mathbf{A}_1^* : -\mathbf{A}_2^*).$$

Clements and Hendry (1993) refer to these as  $M$ - and  $P$ -transforms. No restrictions are imposed by these reversible transforms, so the systems (8) and (9) are isomorphic and have the same likelihood.  $M$ -transforms combine contemporaneous elements of  $\mathbf{x}_t$  and can induce simultaneity. When elements of  $\mathbf{M}$  are unknown, statistical equivalence need not hold, and such cases are discussed in section [5]: in particular, there must be sufficient restrictions on the resulting coefficient matrices to ensure unique identification (for an example, see Hendry and Mizon (1993); Koopmans (1950) provides the classic discussion of identification conditions in econometric systems).  $P$ -transforms induce inter-temporal data combinations, such as creating differences. All types of model are generated by alternative choices for  $\mathbf{M}$  and  $\mathbf{P}$ . If scale-preservation is also desired,  $|\mathbf{M}| = 1$  and  $\mathbf{P}$  should be upper block-triangular with an  $n \times n$  unit matrix in the top-left position (for details, see Ericsson, 1993).

A transformation to levels and differences is achievable by  $P$ -transforms and so leaves the system invariant. One useful reformulation of the system (6) is (see Hendry *et al.*, 1984, Engle and Granger, 1987, Johansen, 1988,

and Banerjee, Dolado, Galbraith and Hendry, 1993):

$$\Delta \mathbf{x}_t = \sum_{j=1}^{s-1} \mathbf{A}_j^* \Delta \mathbf{x}_{t-j} + \boldsymbol{\pi} \mathbf{x}_{t-1} + \mathbf{F} \mathbf{q}_t + \boldsymbol{\nu}_t. \quad (10)$$

No restrictions are imposed by this transform and the resulting system is called a vector error correction system (VECS). Its parameterization is more orthogonal than that in (6) since the variables are usually less inter-correlated after transformation from levels to differences and differentials: the resulting parameters are often more interpretable. When  $s = 2$ , (10) becomes:

$$\Delta \mathbf{x}_t = \mathbf{A}_1^* \Delta \mathbf{x}_{t-1} + \boldsymbol{\pi} \mathbf{x}_{t-1} + \mathbf{F} \mathbf{q}_t + \boldsymbol{\nu}_t. \quad (11)$$

Although it is less usual to consider ‘structural’ representations of closed systems,  $M$ -transforms of (11) yield:

$$\mathbf{M} \Delta \mathbf{x}_t = \mathbf{M}_1^* \Delta \mathbf{x}_{t-1} + \boldsymbol{\pi}^* \mathbf{x}_{t-1} + \mathbf{F}^* \mathbf{q}_t + \boldsymbol{\nu}_t^*. \quad (12)$$

We refer to (12) as a structural VAR, recognizing that the issue of its actual structurality remains to be resolved. When  $\mathbf{M}$  is known:

$$\Delta \mathbf{x}_t^* = \mathbf{A}_1^0 \Delta \mathbf{x}_{t-1}^* + \boldsymbol{\pi}^0 \mathbf{x}_{t-1}^* + \mathbf{F}^0 \mathbf{q}_t^* + \boldsymbol{\nu}_t^* \quad (13)$$

where  $\mathbf{x}_t^* = \mathbf{M} \mathbf{x}_t$ ,  $\mathbf{q}_t^* = \mathbf{M} \mathbf{q}_t$ ,  $\mathbf{A}_1^0 = \mathbf{M} \mathbf{A}_1^* \mathbf{M}^{-1}$ ,  $\boldsymbol{\pi}^0 = \mathbf{M} \boldsymbol{\pi} \mathbf{M}^{-1}$ ,  $\mathbf{F}^0 = \mathbf{M} \mathbf{F} \mathbf{M}^{-1}$  and  $\boldsymbol{\nu}_t^* = \boldsymbol{\nu}_t \sim \text{IN}[\mathbf{0}, \mathbf{M} \boldsymbol{\Omega} \mathbf{M}']$ . In effect, the  $\mathbf{x}_t^*$  are the linear data combinations of economic interest. Although the resulting system is invariant, its parameterization is equivariant: for example, the maximum likelihood (or any linear) estimator  $\hat{\boldsymbol{\pi}}^0$  of  $\boldsymbol{\pi}^0$  is equal to  $\mathbf{M} \hat{\boldsymbol{\pi}} \mathbf{M}^{-1}$  and hence varies in line with the parameter transforms. Both  $M$ - and  $P$ -transforms are used below.

### 2.3 Cointegration

The data  $\{\mathbf{x}_t\}$  are said to be integrated of order  $d$ , denoted  $I(d)$ , when the maximal order of integration of any element is  $d$ . This formulation is needed to accommodate the invariance of linear systems to linear transforms. When the data  $\{\mathbf{x}_t\}$  are integrated of order 1,  $I(1)$ , then  $\Delta \mathbf{x}_t$  is  $I(0)$  and the system specification in (10) is balanced only if  $\boldsymbol{\pi} \mathbf{x}_{t-1}$  is  $I(0)$ : otherwise there arises a contradiction between the left and right hand sides. Clearly  $\boldsymbol{\pi}$  cannot be full rank in such a state of nature since that would contradict the assumption that  $\mathbf{x}_t$  was  $I(1)$ , so let  $\text{rank}(\boldsymbol{\pi}) = r < n$ . Then  $\boldsymbol{\pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$  where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are  $n \times r$  matrices of rank  $r$ , and  $\boldsymbol{\beta}' \mathbf{x}_t$  must comprise  $r$  cointegrating ( $I(0)$ ) relations, inducing a restricted  $I(0)$  representation; when  $s = 1$ , this has the form (see Engle and Granger, 1987 and Johansen, 1988):

$$\Delta \mathbf{x}_t = \mathbf{A}_1^* \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{x}_{t-1}) + \mathbf{F} \mathbf{q}_t + \boldsymbol{\nu}_t. \quad (14)$$

Following Davidson, Hendry, Srba and Yeo (1978), this is called a vector error-correction representation.

To ensure that  $\mathbf{x}_t$  is not  $I(2)$ , a further requirement is that:

$$\text{rank}(\boldsymbol{\alpha}_\perp \boldsymbol{\Phi} \boldsymbol{\beta}'_\perp) = n - r \quad \text{when} \quad \boldsymbol{\Phi} = - \sum_{j=1}^s j \mathbf{A}_j$$

so  $\boldsymbol{\Phi}$  is the general mean-lag matrix,  $\boldsymbol{\alpha}_\perp$  and  $\boldsymbol{\beta}_\perp$  are  $n \times (n - r)$  matrices such that  $\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha} = \mathbf{0}$ ,  $\boldsymbol{\beta}'_\perp \boldsymbol{\beta} = \mathbf{0}$  with  $(\boldsymbol{\alpha} : \boldsymbol{\alpha}_\perp)$  and  $(\boldsymbol{\beta} : \boldsymbol{\beta}_\perp)$  full rank ( $n$ ) matrices. Should the analysis commence in  $I(2)$  space, then  $\boldsymbol{\alpha}_\perp \boldsymbol{\Phi} \boldsymbol{\beta}'_\perp = \boldsymbol{\gamma} \boldsymbol{\delta}'$  is also reduced rank, so some linear combinations first cointegrate from  $I(2)$  to  $I(1)$  and then  $r$  others (perhaps with  $I(1)$  differences of  $I(2)$  variables) cointegrate to  $I(0)$  (see Johansen, 1992a). Thus, both  $I(2)$  and  $I(1)$  impose reduced rank restrictions on the initial formulation in (10), and the former imposes further restrictions on (14).

The behaviour of the system depends greatly on the properties of  $\mathbf{F} \mathbf{q}_t$ . We consider the case where  $\mathbf{q}'_t = (1, t)$  so that  $\mathbf{F} \mathbf{q}_t = \mathbf{f}_1 + \mathbf{f}_2 t$ . When  $\mathbf{f}_2$  is unrestricted, there will be a quadratic trend in the levels of the variables in (14), so to ensure only a linear trend we must force  $t$  to enter the cointegration space: i.e.  $\mathbf{f}_2 = \boldsymbol{\alpha} \boldsymbol{\varphi}$  say, where  $\boldsymbol{\varphi}$  is  $r \times 1$ . Thus, (14) becomes:

$$\Delta \mathbf{x}_t = \mathbf{A}_1^* \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\varphi} t) + \mathbf{f}_1 + \boldsymbol{\nu}_t. \quad (15)$$

There is at most a linear trend in levels in (15). A similar treatment could be given to the intercept, and would force driftless processes on any variables which did not cointegrate: this would be acceptable for interest rates, but not for say GNP (see section [5.2]).

### 3 Open linear systems

#### 3.1 Conditional systems

Let  $\mathbf{x}'_t = (\mathbf{y}'_t : \mathbf{z}'_t)$  where  $\mathbf{y}_t$  and  $\mathbf{z}_t$  are  $n_1 \times 1$  and  $n_2 \times 1$  vectors of endogenous and non-modelled variables respectively with  $n_1 + n_2 = n$ . Partition  $\boldsymbol{\lambda}' = (\boldsymbol{\lambda}'_1 : \boldsymbol{\lambda}'_2)$  conformably with  $(\mathbf{y}'_t : \mathbf{z}'_t)$  and factorize the joint sequential density (conditional on  $\mathbf{q}_t$ ) as:

$$D_{\mathbf{x}}(\mathbf{x}_t | \mathbf{X}_{t-1}^{t-s}, \mathbf{q}_t, \boldsymbol{\lambda}) = D_{\mathbf{y}|\mathbf{z}}(\mathbf{y}_t | \mathbf{z}_t, \mathbf{X}_{t-1}^{t-s}, \mathbf{q}_t, \boldsymbol{\lambda}_1) D_{\mathbf{z}}(\mathbf{z}_t | \mathbf{X}_{t-1}^{t-s}, \mathbf{q}_t, \boldsymbol{\lambda}_2). \quad (16)$$

The validity of conditioning depends on the weak exogeneity of  $(\mathbf{z}_t : \mathbf{q}_t)$  for the parameters of interest in (19) as discussed in Engle, Hendry and Richard (1983). In terms of (16),  $\mathbf{z}_t$  is weakly exogenous for  $\boldsymbol{\psi}$  if (i)  $\boldsymbol{\psi} = \mathbf{f}(\boldsymbol{\lambda}_1)$  alone and (ii)  $\Lambda = \Lambda_1 \times \Lambda_2$  where  $\boldsymbol{\lambda}_1 \in \Lambda_1$  and  $\boldsymbol{\lambda}_2 \in \Lambda_2$ , so that the joint parameter space is the Cartesian product of the conditional and marginal parameter spaces (see Ericsson, 1992 for an exposition).

Keeping to linear systems conditional on  $\mathbf{z}_t$  and  $\mathbf{q}_t$ :

$$\mathbf{y}_t = \boldsymbol{\Xi} \mathbf{q}_t + \boldsymbol{\Gamma}_0 \mathbf{z}_t + \boldsymbol{\Gamma}_1 \mathbf{z}_{t-1} + \cdots + \boldsymbol{\Gamma}_s \mathbf{z}_{t-s} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_s \mathbf{y}_{t-s} + \boldsymbol{\nu}_t \quad (17)$$

Given data up to and including time  $t-1$  on  $(\mathbf{Y}_{t-1}^0 : \mathbf{Z}_{t-1}^0)$  together with  $(\mathbf{z}_t : \mathbf{q}_t)$  and the values of the parameters  $(\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_s, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_s)$ , the next value of  $\mathbf{y}_t$  is generated by equation (17). When  $s = 1$ :

$$\mathbf{y}_t = \boldsymbol{\Xi} \mathbf{q}_t + \boldsymbol{\Gamma}_0 \mathbf{z}_t + \boldsymbol{\Gamma}_1 \mathbf{z}_{t-1} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\nu}_t. \quad (18)$$

One lag is in fact fully general and is the only case we need to consider as follows. Let  $\bar{\mathbf{y}}'_t = (\mathbf{y}'_t : \mathbf{y}'_{t-1} : \cdots : \mathbf{y}'_{t-s+1})$  and  $\bar{\mathbf{z}}'_t = (\mathbf{z}'_t : \mathbf{z}'_{t-1} : \cdots : \mathbf{z}'_{t-s})$ , then stack the coefficients in the second line of (17) in conformable  $(n_1 s \times n_2 s$  and  $n_1 s \times n_1 s)$  matrices:

$$\boldsymbol{\Gamma}^+ = \begin{pmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_s \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Phi}^+ = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{s-1} & \boldsymbol{\Phi}_s \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}$$

By direct multiplication (ignoring the deterministic terms for simplicity), it can be checked that (17) maps into (for  $\bar{\boldsymbol{\nu}}_t = (\boldsymbol{\nu}'_t : \mathbf{0}' : \cdots : \mathbf{0}')'$ ):

$$\bar{\mathbf{y}}_t = \boldsymbol{\Gamma}^+ \bar{\mathbf{z}}_t + \boldsymbol{\Phi}^+ \bar{\mathbf{y}}_{t-1} + \bar{\boldsymbol{\nu}}_t,$$

where the top block delivers (17), augmented by identities of the form  $\mathbf{y}_{t-j} \equiv \mathbf{y}_{t-j}$ . In some respects the resulting notation can be awkward, so we will concentrate on 1 lag in both typology expositions. Equally, 2-equation systems will be used to illustrate the later model analyses.

#### 3.2 Simultaneous equations

Interest in economics often centers on a simultaneous representation such as:

$$\mathbf{B}_0 \mathbf{y}_t + \mathbf{B}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{B}_s \mathbf{y}_{t-s} = \boldsymbol{\Upsilon} \mathbf{q}_t + \mathbf{C}_0 \mathbf{z}_t + \mathbf{C}_1 \mathbf{z}_{t-1} + \cdots + \mathbf{C}_s \mathbf{z}_{t-s} + \boldsymbol{\epsilon}_t \quad (19)$$

where  $\mathbf{B}_0$  is the  $n_1 \times n_1$  matrix of coefficients of all the endogenous variables,  $\mathbf{B}_1$  is the coefficient matrix of the 1-lagged endogenous variables ...,  $\mathbf{C}_0$  is the  $n_2 \times n_2$  matrix of the current dated non-modelled variables etc, and  $\boldsymbol{\epsilon}_t \sim \text{IN}(\mathbf{0}, \boldsymbol{\Sigma})$ . As it stands, (19) is a partial or incomplete model unless supplemented by a marginal model for the  $\{\mathbf{z}_t\}$  process. For  $s = 1$  lag:

$$\mathbf{B}_0 \mathbf{y}_t + \mathbf{B}_1 \mathbf{y}_{t-1} = \boldsymbol{\Upsilon} \mathbf{q}_t + \mathbf{C}_0 \mathbf{z}_t + \mathbf{C}_1 \mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t \quad (20)$$

The condition  $|\mathbf{B}_0| \neq 0$  is needed to ensure that the system of equations is complete (i.e. as many independent equations as endogenous variables) and that the equations are mutually consistent with none redundant. Sufficient restrictions on the coefficient matrices are required to ensure a unique representation. The model in (19) is a special case of a structural form, which claims to model the decision equations of the relevant agents: as before, the structurality of the representation is a claim that must be established.

The reduced, or solved, form expresses  $\mathbf{y}_t$  as a function of the parameters, non-modelled variables and lagged variables:

$$\mathbf{y}_t = \mathbf{f}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-s}, \mathbf{z}_t, \mathbf{z}_{t-1}, \dots, \mathbf{z}_{t-s}, \mathbf{q}_t, \boldsymbol{\nu}_t). \quad (21)$$

When the system is linear and  $\mathbf{B}_0$  is non-singular, the solution to (19) is:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}_0^{-1} \boldsymbol{\Upsilon} \mathbf{q}_t + \mathbf{B}_0^{-1} (\mathbf{C}_0 \mathbf{z}_t + \dots + \mathbf{C}_s \mathbf{z}_{t-s}) - \mathbf{B}_0^{-1} (\mathbf{B}_1 \mathbf{y}_{t-1} + \dots + \mathbf{B}_s \mathbf{y}_{t-s}) + \mathbf{B}_0^{-1} \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\Xi} \mathbf{q}_t + \boldsymbol{\Gamma}_0 \mathbf{z}_t + \boldsymbol{\Gamma}_1 \mathbf{z}_{t-1} + \dots + \boldsymbol{\Gamma}_s \mathbf{z}_{t-s} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \dots + \boldsymbol{\Phi}_s \mathbf{y}_{t-s} + \boldsymbol{\nu}_t \end{aligned} \quad (22)$$

which returns us to a restricted version of the original system representation and reveals the implicit, testable, restrictions imposed by the simultaneous formulation (see Monfort and Rabemananjara, 1990 for an analysis of open stationary systems).

### 3.3 Cointegration

In practice,  $\mathbf{y}_t$  and  $\mathbf{z}_t$  are often  $I(1)$  and the balance of (20) must be checked since combinations of orders of integration for  $y$ s and  $z$ s need not be mutually consistent with the claim that  $\boldsymbol{\epsilon}_t$  is  $I(0)$ . For example, when  $\mathbf{y}_t$  is  $I(1)$ , the unit roots must cancel between  $\{\mathbf{B}_0 \mathbf{y}_t\}$  and  $\{\mathbf{C}_0 \mathbf{z}_t + \mathbf{C}_1 \mathbf{z}_{t-1} - \mathbf{B}_1 \mathbf{y}_{t-1}\}$ . One possibility is when  $\mathbf{B}_1 = -\mathbf{B}_0$  and  $\mathbf{C}_1 = -\mathbf{C}_0$  so (20) holds in differences:

$$\mathbf{B}_0 \Delta \mathbf{y}_t = \mathbf{C}_0 \Delta \mathbf{z}_t + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\epsilon}_t. \quad (23)$$

All terms are  $I(0)$  when the levels data are  $I(1)$ , and balance occurs. Notice that the system has not been differenced: rather, the model happens to satisfy coefficient restrictions which sustain a differenced representation. In particular, its error process remains  $IN[0, \Sigma]$  and no negative moving average has been induced as would occur when the data are differenced. This is an important practical issue and can be clarified by expressing the system in lag-operator notation as:

$$(\mathbf{B}_0 + \mathbf{B}_1 L) \mathbf{y}_t = (\mathbf{C}_0 + \mathbf{C}_1 L) \mathbf{z}_t + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\epsilon}_t \quad (24)$$

where  $L^k x_t = x_{t-k}$ . Imposing the parameter restrictions that  $\mathbf{B}_1 = -\mathbf{B}_0$  and  $\mathbf{C}_1 = -\mathbf{C}_0$  yields:

$$(\mathbf{B}_0 - \mathbf{B}_0 L) \mathbf{y}_t = (\mathbf{C}_0 - \mathbf{C}_0 L) \mathbf{z}_t + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\epsilon}_t$$

or:

$$\mathbf{B}_0 (1 - L) \mathbf{y}_t = \mathbf{C}_0 (1 - L) \mathbf{z}_t + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\epsilon}_t \quad (25)$$

leading to (23). As a consequence, the reduced form (18) becomes:

$$\mathbf{y}_t = \boldsymbol{\Gamma}_0 \mathbf{z}_t - \boldsymbol{\Gamma}_0 \mathbf{z}_{t-1} + \mathbf{y}_{t-1} + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\nu}_t. \quad (26)$$

This is obviously a restrictive parameterization, and when the highest order of polynomial in time in  $\mathbf{q}_t$  is  $h$ , then (26) includes a polynomial of order  $h + 1$ .

More generally, when  $\mathbf{y}_t$  and  $\mathbf{z}_t$  are  $I(1)$  and cointegrate (see Engle and Granger, 1987), then some linear combinations are  $I(0)$ : we denote these by  $\mathbf{u}_t = (\mathbf{y}_t - \boldsymbol{\Psi} \mathbf{z}_t)$ . Many possibilities arise depending on the dimensionality of  $\mathbf{y}_t$  and  $\mathbf{z}_t$  ( $n_1$  and  $n_2$ ) and the number of cointegrating relations. We assume complete cointegration for simplicity here, so the system can be written as (see Davidson and Hall, 1991):

$$\mathbf{B}_0 \Delta \mathbf{y}_t = \mathbf{C}_0 \Delta \mathbf{z}_t - \mathbf{D}_0 (\mathbf{y}_{t-1} - \boldsymbol{\Psi} \mathbf{z}_{t-1}) + \boldsymbol{\Upsilon} \mathbf{q}_t + \boldsymbol{\epsilon}_t, \quad (27)$$

where every term is  $I(0)$ ,  $\mathbf{D}_0 - \mathbf{B}_0 = \mathbf{B}_1$  and  $\mathbf{D}_0 \boldsymbol{\Psi} = \mathbf{C}_0 + \mathbf{C}_1$ . When  $\mathbf{D}_0$  induces a full-rank stable process, the system adjusts to the equilibrium or steady-state path  $(\mathbf{y}^* - \boldsymbol{\Psi} \mathbf{z}^*) = \boldsymbol{\phi}(\mathbf{q})$  driven by the vector error-correction mechanism (VECM)  $(\mathbf{y}_{t-1} - \boldsymbol{\Psi} \mathbf{z}_{t-1})$ . If only a subset of  $m_1 < n_1$  equations cointegrates, then these alone have

‘levels equilibria’ and the remaining variables have steady-state growth rates (of the form  $E[\Delta y_{it}] = g_i$ ). The behaviour of the system depends greatly on  $\mathbf{\Upsilon} \mathbf{q}_t$  as in the closed case.

We now consider the typology of open linear systems. At most one lag is allowed (generalizations may change the resultant type) and deterministic factors are omitted (depending on how they enter as discussed above, no major surprises result from their inclusion). From the preceding analysis it should be clear that conditioning induces a reduction in the system under analysis from an  $n$ -dimensional to an  $n_1$ -dimensional density and weak exogeneity ensures that such a reduction is without loss of information about the parameters of interest. Thus, the closed system in  $\mathbf{x}_t$  is the most general under consideration and the open system in  $\mathbf{y}_t$  conditional on  $\mathbf{z}_t$  is a special case; naturally, the closed system in  $\mathbf{y}_t$  is even more specialized and does not entail that an open system (which includes  $\mathbf{z}_t$ ) is somehow more general than a closed one. Further, the imposition of a ‘structure’ via simultaneity also entails restrictions on the system and again is less general. These points are important in understanding the two typologies that follow.

## 4 A typology of open linear dynamic systems

In this section, the conditional statistical generating mechanism is given by (28) without deterministic components:

$$\mathbf{y}_t = \mathbf{\Gamma}_0 \mathbf{z}_t + \mathbf{\Gamma}_1 \mathbf{z}_{t-1} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\nu}_t. \quad (28)$$

where  $\boldsymbol{\nu}_t \sim \text{IN}(\mathbf{0}, \mathbf{\Omega})$ . We refer to (28) as a vector autoregressive-distributed lag system denoted by  $VAD(1, 1; n_1, n_2)$  where the first two arguments note the maximum lags and the second two the dimensions of the  $\mathbf{y}_t$  and  $\mathbf{z}_t$  vectors; information on the detailed lags of each variable is, of course, thereby lost. Each system is postulated as a stochastic economic model of a DGP which is unknown to the system’s proprietor. When the system coincides with the DGP, then usually it will work well and estimation and forecasting will deliver the anticipated results. The case of interest in practice is when the system does not coincide with the DGP, but is an invalid reduction of it.

Many economic models fail to specify the complete distribution of all the random variables under analysis, due to implicit exogeneity assumptions. A necessary condition for valid analysis of (28) is that  $\mathbf{z}_t$  is weakly exogenous for the parameters of interest  $\boldsymbol{\psi}$ . Such a requirement already precludes many possible systems since no cointegration vector relevant to (28) may enter the marginal process for  $\mathbf{z}_t$ . The systems are ‘reduced forms’: pre-multiply by  $\mathbf{B}_0 \neq \mathbf{I}$  to obtain the simultaneous form. This ease of transform between system types and corresponding model types is one reason why the typologies overlap. Nevertheless, as will be seen, models of systems raise new issues.

The distributions of estimators and tests often differ asymptotically between  $I(0)$  and  $I(1)$  processes, although in finite samples it can be difficult to discriminate between roots close to, but less than, unity (the unit root case is also a shorthand for near non-stationarity). The analysis is conditional on the initial value  $\mathbf{x}_0 = (\mathbf{y}_0 : \mathbf{z}_0)'$ , so the process is only asymptotically stationary for  $I(0)$ . As can be seen from Table 1, the system typology is an exact analogy of the individual equation one. The restrictions as listed are dependent on maintaining the contents of  $(\mathbf{y}_t : \mathbf{z}_t)$ : as stressed above, the VAR nests the VAD when  $\mathbf{y}_t = \mathbf{x}_t$ . We briefly comment on each system type in turn.



Table 1: Vector autoregressive-distributed lag system typology<sup>a</sup>

Type	System	Restrictions
VAD	$\mathbf{y}_t = \Gamma_0 \mathbf{z}_t + \Gamma_1 \mathbf{z}_{t-1} + \Phi_1 \mathbf{y}_{t-1} + \epsilon_t$	None
VECM	$\Delta \mathbf{y}_t = \Gamma_0 \Delta \mathbf{z}_t - \mathbf{D}_0 (\mathbf{y}_{t-1} - \Psi \mathbf{z}_{t-1}) + \epsilon_t$	None
Static system	$\mathbf{y}_t = \Gamma_0 \mathbf{z}_t + \mathbf{e}_t$	$\Phi_1 = \Gamma_1 = \mathbf{0}$
Levels VAR	$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \nu_t$	$\Gamma_0 = \Gamma_1 = \mathbf{0}$
Differenced system	$\Delta \mathbf{y}_t = \Gamma_0 \Delta \mathbf{z}_t + \mathbf{v}_t$	$\Phi_1 = \mathbf{I},$ $\Gamma_1 = -\Gamma_0$
Leading indicator	$\mathbf{y}_t = \Gamma_1 \mathbf{z}_{t-1} + \xi_t$	$\Phi_1 = \Gamma_0 = \mathbf{0}$
Partial adjustment	$\mathbf{y}_t = \Gamma_0 \mathbf{z}_t + \Phi_1 \mathbf{y}_{t-1} + \varepsilon_t$	$\Gamma_1 = \mathbf{0}$
Common factor	$\mathbf{y}_t = \Gamma_0 \mathbf{z}_t + \mathbf{u}_t,$ $\mathbf{u}_t = \Phi_1 \mathbf{u}_{t-1} + \zeta_t$	$\Gamma_1 = -\Phi_1 \Gamma_0$
Finite distributed lag	$\mathbf{y}_t = \Gamma_0 \mathbf{z}_t + \Gamma_1 \mathbf{z}_{t-1} + \mathbf{w}_t$	$\Phi_1 = \mathbf{0}$
Dead start	$\mathbf{y}_t = \Gamma_1 \mathbf{z}_{t-1} + \Phi_1 \mathbf{y}_{t-1} + \varsigma_t$	$\Gamma_0 = \mathbf{0}$

<sup>a</sup> For simplicity, the typology is illustrated by a  $VAD(1, 1; n_1, n_2)$  without deterministic elements.

#### 4.1 Vector error-correction system

The VECS is isomorphic to the VAD with  $\mathbf{D}_0 \Psi = \Gamma_1 + \Gamma_0$  and  $\mathbf{D}_0 = \mathbf{I} - \Phi_1$ . We have discussed its forms and related problems in section [3.3] above. When  $\mathbf{z}_t$  is weakly exogenous for  $\psi$ , then the cointegration vectors  $(\mathbf{I}_{n_1} : \Psi) (\mathbf{y}'_t : \mathbf{z}'_t)' = \beta'_1 \mathbf{x}_t$  must coincide with a non-singular linear transformation of those in the VAR for  $\mathbf{x}_t$  other than vectors entering only the marginal process.

#### 4.2 Static system

Such a system might be used to investigate cointegration (see Engle and Granger, 1987, and Phillips, 1991), with  $\{\mathbf{e}_t\}$  generally being highly autocorrelated. This type recurs in the second typology as part of a triangular system representation.

#### 4.3 Vector autoregression

When the information set is restricted to  $\{\mathbf{y}_t\}$  alone, a VAR is unlikely to provide a dominant system form. Allowing the information to be  $\{\mathbf{x}_t\}$  often leads to a system that accurately characterizes the available data, albeit with a profligate parameterization. There are several solutions to that dilemma, including constructing parsimonious VARs (see Hendry and Mizon, 1993) and imposing exact or loose prior information as in Bayesian VARs (see Doan, Litterman and Sims, 1984). The main problem arising from not conditioning on  $\mathbf{z}_t$  is usually apparent parameter change: this seems likely to arise when  $\mathbf{z}_t$  includes policy determined variables which are not well represented by linear autoregressions. The appropriate solution need not be to condition on the variables which are difficult to model and hope that weak exogeneity is not violated; instead, modelling the regime shifts (e.g. by adding dummy variables) allows tests for the absence of the additional variables in the marginal system from the conditional system and hence evaluates an aspect of super exogeneity (see Engle and Hendry, 1993).

#### 4.4 Differenced system

Estimating a system in levels entails estimating any unit roots that may be present and may require the use of Dickey and Fuller (1979) type critical values for hypothesis tests (see Sims, Stock and Watson, 1990). Consequently, some investigators consider systems in the first differences of the original variables, thereby imposing all the unit roots and sustaining conventional inference when the levels are  $I(1)$ . However, as noted above, if a subset of the levels cointegrate, then the differenced system excludes these, which may induce a negative moving-average error that reflects the loss of long-run information. Clements and Hendry (1994) find that such a loss can be important for forecasting even in 2-dimensional systems.

#### 4.5 Leading indicator

Leading indicator systems have seen a revival of interest recently (see *inter alia* Neftci, 1979, Stock and Watson, 1989, 1992, Diebold and Rudebusch, 1991, and Artis, Bladen-Hovell, Osborn, Smith and Zhang, 1993). The evidence that there exist systematic leading and lagging links that can be exploited for forecasting without econometric modelling of the relevant system is scanty (see e.g. Emerson, 1994). This is unsurprising in that the *VAD* is just a generalized ‘leading indicator’ system which incorporates both economic theory and empirical information.

#### 4.6 Partial adjustment

The exclusion of lagged non-modelled variables can be seen in the system context to be an unwarranted restriction, dependent on the timing of measurements and the length of observation intervals. The rationale for the partial adjustment principle is discussed in Hendry (1995) and relative to a VECS does not seem sustainable. Should the lag length restrictions on non-modelled variables be valid, commencing from the *VAD*  $(1, 1; \cdot, \cdot)$  allows them to be tested.

#### 4.7 Common factor

As in individual equations, autoregressive errors impose common factor restrictions on the dynamics and do not represent a generalization of the system. The common factor (that  $\Gamma_1 = -\Phi_1\Gamma_0$ ) can be seen using lag-operator notation, since:

$$(\mathbf{I} - \Phi_1 L) \mathbf{y}_t = (\Gamma_0 + \Gamma_1 L) \mathbf{z}_t + \epsilon_t = (\mathbf{I} - \Phi_1 L) \Gamma_0 \mathbf{z}_t + \epsilon_t \quad (29)$$

or:

$$\begin{aligned} \mathbf{y}_t &= \Gamma_0 \mathbf{z}_t + \mathbf{u}_t \\ \mathbf{u}_t &= \Phi_1 \mathbf{u}_{t-1} + \epsilon_t \end{aligned} \quad (30)$$

which holds only if the lag polynomial matrices  $(\mathbf{I} - \Phi_1 L)$  and  $(\Gamma_0 + \Gamma_1 L)$  have the matrix common factor  $(\mathbf{I} - \Phi_1 L)$ . Further,  $\Phi_1 = \mathbf{I}$  is the restriction for no cointegration, and hence induces a differenced representation (see Hendry and Mizon, 1978, and Mizon, 1993).

#### 4.8 Finite distributed lag

Again, the system context suggests that the restriction of excluding all lagged endogenous information is untenable:  $\mathbf{y}_{t-1}$  is almost certain to matter in time-series economics. Arguments for long lags on non-modelled variables may be based on a misconception about dynamic simulation. This is sometimes used as a model evaluation device, and it is clear that improved performance can appear to result from replacing endogenous dynamics by non-modelled variables: however, such an outcome is spurious unless the non-modelled variables are both strongly exogenous and the assumed information on them would be available in an operational context. For a general critique, see Chong and Hendry (1986).

#### 4.9 Dead start

The presence or absence of the contemporaneous value of  $\mathbf{z}_t$  turns on three considerations. First, is  $\mathbf{z}_t$  weakly exogenous for  $\psi$  in the resulting system; secondly, if so, is the coefficient matrix  $\Gamma_0 \neq \mathbf{0}$ ; finally, if so is  $\Gamma_0$  constant over time and across regimes? A positive answer to all three questions entails that the dead-start model is a reduced form and hence it may omit important information if theory testing is an objective of the analysis. Interestingly, a negative answer to the first two also precludes the possibility that dead start is a structural representation in its own right, noting that the presence of  $\mathbf{z}_t$  in the conditional model is irrelevant to its exogeneity status.

A great deal more could be written about each of the typology forms, their empirical performance, specification tests, and detailed interpretations. Most of the criticisms of each type as a single equation model carry over to the system context when that type does not coincide with the DGP (see Hendry, 1995). Some of these issues will arise in the typology concerned with alternative model representations of linear systems.

### 5 Models of linear systems

Reconsider the formulation of the  $I(0)$  VAR in (6) when  $s = 1$  written as:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{F}\mathbf{q}_t + \boldsymbol{\nu}_t \text{ where } \boldsymbol{\nu}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}]. \quad (31)$$

There are a number of potentially different models of this basic VAR, including error-correction models, open and closed simultaneous systems, conditional models, causal chains, block-recursive models and triangular representations. These alternatives are generated by the operations of conditioning and  $M$ - and  $P$ -transforms. Whereas members of the system typology were not equivalent to each other, many model types are isomorphic in a world of constant parameters. In each case, we will note the number of parameters and hence whether restrictions are or are not being imposed.

A 2-equation maquette will illustrate each type. Although the equations in each model form could be any type from a single equation typology matching that in section [4], we focus on unrestricted  $VAD(1,1;\cdot,\cdot)$  forms. The basic system is when  $n = 2$  and there is only one deterministic component, namely an intercept, so that (31) can be expressed as:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\nu}_t$$

or:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (32)$$

where:

$$\begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \sim \text{IN} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \right]. \quad (33)$$

There are 9 parameters in (32) and (33) namely  $(a_{11}, a_{12}, a_{21}, a_{22}; \mu_1, \mu_2; \omega_{11}, \omega_{12}, \omega_{22})$ , and these are assumed to be variation free. The parameter values differ across the model types but all parameter spaces are subspaces of  $\mathbb{R}^9$ , with  $\omega_{11} > 0$ ,  $\omega_{22} > 0$  and  $\omega_{12}^2 \leq \omega_{11}\omega_{22}$ . The values taken by  $(a_{11}, a_{12}, a_{21}, a_{22})$  determine whether the system is  $I(1)$  or  $I(0)$ . We now consider each type in turn.

#### 5.1 Vector autoregressive representations

Under the present assumptions, both (31) and the specialization in (32) are first-order VARs. There is no conditioning,  $\mathbf{M} = \mathbf{I}$  and  $\mathbf{P} = \mathbf{I}$ . The system is stationary when the latent roots of  $\mathbf{A}$  are within the unit circle (subject to appropriate initial condition distributions), given by solving  $|\lambda\mathbf{I} - \mathbf{A}| = 0$  or:

$$\begin{vmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{vmatrix} = (\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} = \lambda^2 - \text{tr}(\mathbf{A})\lambda + |\mathbf{A}| = 0.$$

This yields the quadratic equation  $\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0$  with roots:

$$\frac{1}{2} \left[ (a_{11} + a_{22}) \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}a_{21})} \right].$$

When either  $a_{12}$  or  $a_{21}$  is zero, the roots are simply  $a_{11}$  and  $a_{22}$ , so for stationarity we require  $|a_{11}| < 1$  and  $|a_{22}| < 1$ . The condition  $a_{21} = 0$  entails that  $x_1$  does not Granger-cause  $x_2$  (vice versa for  $a_{12} = 0$ : see Granger, 1969). More generally, when the latent roots of  $\mathbf{A}$  lie within the unit circle, the matrix  $(\mathbf{I} - \mathbf{A})$  is non-singular, so on average:

$$\mathbf{E}[\mathbf{x}_t] = (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\mu}. \quad (34)$$

Thus, a stationary VAR converges on a unique equilibrium point, or steady-state growth path if a linear deterministic trend is allowed. To consider non-stationarity, we turn to the next member of the typology.

## 5.2 Vector error corrections

Re-express (32) in a vector error correction form as ( $\mathbf{M} = \mathbf{I}$  but  $\mathbf{P}$  differs from a unit matrix by  $-\mathbf{I}$  as its top right block):

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (35)$$

which is a special case of (10), revealing that  $\boldsymbol{\pi} = (\mathbf{A} - \mathbf{I})$ . Thus,  $\text{rank}(\boldsymbol{\pi}) = 2$  is necessary for stationarity and matches (34). When  $a_{11} \neq 1$  and  $a_{21} \neq 0$  (assumptions made for convenience of normalization), (35) can be written as:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} (a_{11} - 1)(x_{1t-1} - \delta_1 x_{2t-1}) \\ a_{21}(x_{1t-1} - \delta_2 x_{2t-1}) \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (36)$$

where  $\delta_1 = a_{12}/(1 - a_{11})$  and  $\delta_2 = (1 - a_{22})/a_{21}$ . This has a more familiar error correction form, and with 9 unrestricted parameters remains isomorphic to (32) (other than the non-zero requirements for normalizing). The static equilibrium defined by taking unconditional expectations yields two equations linking  $x_1$  and  $x_2$  inducing a point as the solution as in (34).

When  $\delta_1 = \delta_2$ , weak exogeneity of either regressor for the parameters of interest of the other equation is violated, and the system becomes  $I(1)$  since  $\text{rank}(\boldsymbol{\pi}) = 1$  (see Boswijk, 1992, and Johansen, 1992b). Explicitly:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} (a_{11} - 1) \\ a_{21} \end{pmatrix} (x_{1t-1} - \delta x_{2t-1}) + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (37)$$

and letting  $g_i = \mathbf{E}[\Delta x_{it}]$ , the long-run solution is the relation:

$$\mathbf{E}[x_{1t} - \delta x_{2t}] = \frac{(g_1 - \mu_1)}{(a_{11} - 1)} = \frac{(g_2 - \mu_2)}{a_{21}} = \kappa. \quad (38)$$

Since (38) also entails that  $\mathbf{E}[\Delta x_{1t}] = \delta \mathbf{E}[\Delta x_{2t}]$ , then  $g_1 = \delta g_2$  and so  $x_{2t}$  grows at the rate:

$$g_2 = \frac{a_{21}\mu_1 - (a_{11} - 1)\mu_2}{[\delta a_{21} - (a_{11} - 1)]} \quad (39)$$

with  $x_{1t}$  growing at  $\delta$  times this rate. Thus, the  $I(1)$  solution set is a line, rather than the point that resulted for  $I(0)$ . However, (36) also becomes  $I(1)$  under other restrictions which do not violate weak exogeneity (such as when  $a_{22} = 1$  and  $a_{21} = 0$ ) so the presence or absence of weak exogeneity conditions is not tied to the degree of integration.

In (38), when  $\mu_1 = (a_{11} - 1)\varphi$  and  $\mu_2 = a_{21}\varphi$  then  $g_2 = 0$  from (39) and:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} (a_{11} - 1) \\ a_{21} \end{pmatrix} (x_{1t-1} - \delta x_{2t-1} + \varphi) + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix}.$$

Thus, the system exhibits no growth when the intercept vector lies in the cointegration space: as noted above, such a restriction might hold for variables like interest rates or the acceleration of the price level ( $\Delta^2 p_t$ ), but seems unlikely for real flow magnitudes like income.

In general, letting  $\boldsymbol{\pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$  where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are  $n \times r$  matrices of rank  $r < n$ , the general VECM representation was noted above as:

$$\Delta \mathbf{x}_t = \sum_{j=1}^{s-1} \mathbf{A}_j^* \Delta \mathbf{x}_{t-j} + \boldsymbol{\alpha}(\boldsymbol{\beta}' \mathbf{x}_{t-1}) + \mathbf{F} \mathbf{q}_t + \boldsymbol{\nu}_t. \quad (40)$$

This model form is in  $I(0)$  space when correctly formulated, so inference about the parameters can be conducted using conventional procedures. Identification restrictions must be imposed on  $\beta$  to ensure uniqueness of  $\alpha$  and  $\beta$ . The parameterization in terms of short-run dynamics and long-run disequilibria is closer to orthogonality than in (31), but the innovation process is unaltered. Restricting  $F\mathbf{q}_t$  to lie in the cointegration space by  $F = \alpha\varphi$  when  $\mathbf{q}_t = 1$  (an intercept) yields an ECM of  $\alpha(\beta'\mathbf{x}_{t-1} + \varphi)$  so has zero long-run growth, as seen for the 2-equation illustration.

### 5.3 Simultaneous models

Haavelmo (1943) described how to estimate this representation, which is often adopted to capture the structure of economic agents' decisions rules. A non-diagonal  $n \times n$  matrix  $\mathbf{B}_0$  premultiplies  $\Delta\mathbf{x}_t$  to allow contemporaneous connections between the variables. Since the system remains isomorphic for all non-singular matrices, normalization and identification requirements must be imposed on the model to ensure a unique representation. Here, identification involves all three attributes of uniqueness, interpretability given the prior theory, and correspondence to the agents' decisions rules. The first can be imposed by arbitrary restrictions (which may even be suggested by theory) but is then unlikely to correspond to autonomous relations. This problem led Sims (1980) to suggest that such restrictions were 'incredible', but Hendry and Mizon (1993) argue that his critique lacks force since even the VAR is a model and there are almost bound to be valid restrictions between elements of the  $\{\mathbf{A}_i\}$ : cointegration restrictions are one possible example. More usefully, the first-order VAR is defined by:

$$E[\mathbf{x}_t | \mathbf{x}_{t-1}] = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\mu} \quad (41)$$

and we construe modelling as attempting to represent this conditional expectation more parsimoniously (see Clements and Mizon, 1991). One possibility, among many others, is that  $\mathbf{A} = -\mathbf{B}^{-1}\mathbf{C}$  say, where  $\{\mathbf{B}, \mathbf{C}\}$  are sparse matrices. In the 2-equation  $I(0)$  case, consider:

$$\begin{pmatrix} 1 & b_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (42)$$

This model is one just-identified representation of (6) but may be more interpretable. Under stationarity it is observationally equivalent to the VAR, but it may be valid to impose (say)  $c_{21} = 0$ . For an  $I(1)$  system, an interesting possibility is:

$$\begin{pmatrix} 1 & b_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} c_{11}^* \\ c_{22}^* \end{pmatrix} (x_{1t-1} - \delta x_{2t-1}) + \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix}. \quad (43)$$

However, this 'triangular' system is unidentified without further restrictions (it has 9 parameters for a cointegrated representation with only 8), and e.g.  $c_{11}^* = 0$  corresponds to a normalization in the original parameter space.

The general simultaneous equations form is:

$$\mathbf{B}_0\Delta\mathbf{x}_t = \sum_{j=1}^{s-1} \mathbf{B}_j\Delta\mathbf{x}_{t-j} + \alpha^*(\beta'\mathbf{x}_{t-1}) + \mathbf{F}^*\mathbf{q}_t + \mathbf{v}_t \text{ where } \mathbf{v}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Sigma}]. \quad (44)$$

Notice that  $\beta$  is unaffected by the switch of model form and is an invariant of the transformation, but that all other parameters are transformed, including  $\alpha^*$ . Such a formulation can be more interpretable than (40) yet use fewer parameters, allowing tests of over-identifying restrictions by checking whether (44) parsimoniously encompasses (40) (see Hendry and Richard, 1989, and Hendry and Mizon, 1993).

### 5.4 Conditional models

Conditional models correspond to the special case (taking  $x_{2t}$  as the conditioning variable from the 2-equation VAR):

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1^0 \\ \mu_2^0 \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad (45)$$

where:

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \text{IN} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \omega_{22} \end{pmatrix} \right]. \quad (46)$$

There are still 9 parameters so an isomorphic representation again results prior to imposing further restrictions. When the system is a cointegrated VAR, we have:

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{11}^0 \\ \alpha_{21}^0 \end{pmatrix} (x_{1t-1} - \delta x_{2t-1}) + \begin{pmatrix} \mu_1^0 \\ \mu_2^0 \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad (47)$$

which is just identified (8 parameters). However, conditioning is valid only when  $\alpha_{21}^0 = 0$  which entails a restriction that can be evaluated against the data evidence. When it is acceptable, the resulting model yields an ECM as its first equation, with the  $I(1)$  completing model of the conditioning variable being a random walk with drift.

From (6) conditional on  $\mathbf{z}_t$ :

$$\Delta \mathbf{y}_t = \mathbf{C}_0 \Delta \mathbf{z}_t + \sum_{j=1}^{s-1} \mathbf{C}_j \Delta \mathbf{x}_{t-j} + \alpha_0 (\beta' \mathbf{x}_{t-1}) + \mathbf{F}_0 \mathbf{q}_t + \epsilon_{1t} \quad (48)$$

where  $\epsilon_{1t} \sim \text{IN}[0, \Sigma_{11}]$  and  $\Sigma_{11}$  depends on the elements of  $\Omega_\nu$ . As it stands, (48) is a partial or incomplete model, but could be supplemented by marginal models for the  $\{\mathbf{z}_t\}$  process. The validity of conditioning depends on the weak exogeneity of  $\mathbf{z}_t$  for the parameters of interest in (48) as discussed above. As before,  $\beta$  remains an invariant of the reformulated system.

### 5.5 Conditional simultaneous models

It is not possible to represent this case in a 2-equation typology, since these models combine the two preceding types in the form:

$$\mathbf{B}_0 \Delta \mathbf{y}_t = \mathbf{G}_0 \Delta \mathbf{z}_t + \sum_{j=1}^{s-1} \mathbf{G}_j \Delta \mathbf{x}_{t-j} + \alpha_0^* (\beta' \mathbf{x}_{t-1}) + \mathbf{F}_0^* \mathbf{q}_t + \epsilon_{1t}^* \quad (49)$$

where  $\epsilon_{1t}^* \sim \text{IN}[0, \Sigma_{11}^*]$ . The simultaneity is between elements of  $\Delta \mathbf{y}_t$  and the conditioning is on elements of  $\Delta \mathbf{z}_t$  so a 3-variable system is the smallest that can manifest both characteristics. The generic 'structural simultaneous system' is usually written in the form of (49). Often (48) is referred to as the reduced form of (49). Sometimes, models such as (49) have been postulated *a priori* without checking the validity of the entailed reductions from the joint sequential density, then tested against their own derived reduced form when the latter may not be congruent. Such an approach seems unlikely to deliver useful representations of the economy. It is, of course, proceeding from simple to general and either relies on an initial correct specification, or on a sequential testing process that hopes to uncover a congruent specification *en route*.

### 5.6 Causal chains

Wold and Jur  en (1953) placed considerable emphasis on this representation which they argued was closer to the autonomous decision rules of economic agents than Haavelmo (1943) simultaneous system. They proved that a causal chain could always be developed, but did not show that the resulting model embodied the underlying structure. Indeed, since their proof also applies to arbitrary orderings of equations, some other criterion must be used to select between the resulting isomorphic forms. In the 2-equation,  $I(0)$  illustration, one possible causal chain is:

$$\begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (50)$$

where:

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \sim \text{IN} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{pmatrix} \right]. \quad (51)$$

As before, there remain 9 parameters, although some of the  $\{\rho_{ij}\}$  could be restricted. In the  $I(1)$  system with cointegration:

$$\begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{11}^* \\ \alpha_{21}^* \end{pmatrix} (x_{1t-1} - \delta x_{2t-1}) + \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (52)$$

so no testable restrictions are imposed relative to the VAR. Weak exogeneity of the first variable for the parameters of the second equation requires  $\alpha_{11}^* = 0$ .

In general, let  $\Omega^{-1} = \mathbf{H}\mathbf{H}'$  where  $\mathbf{H}$  is upper triangular and pre-multiply (6) by  $\mathbf{H}'$ :

$$\mathbf{H}'\mathbf{x}_t = \sum_{j=1}^s \mathbf{H}'\mathbf{A}_j \mathbf{x}_{t-j} + \mathbf{H}'\mathbf{F}\mathbf{q}_t + \mathbf{H}'\boldsymbol{\nu}_t \text{ where } \mathbf{H}'\boldsymbol{\nu}_t = \boldsymbol{\eta}_t \sim \text{IN}[\mathbf{0}, \mathbf{I}]. \quad (53)$$

Then (53) provides a causal chain in that the error covariance matrix is diagonal (and normalized for simplicity) and each equation in the representation can treat contemporaneous variables as previously determined. However, none of the resulting systems need satisfy weak exogeneity of each equation's contemporaneous variables for the parameters of interest (see Engle *et al.*, 1983). If so, no equation will in fact embody structure. Implicitly, causal chains impose conditioning in a recursive order over equations.

### 5.7 Block recursive representations

Both (50) and (45) are in fact block recursive, in that one of the (single equation) blocks can be taken as determined prior to the other. In a larger system, subsystems could be simultaneous internally, but block recursive externally. Thus, this is an intermediate step towards a causal chain. In general (for 2 blocks only, to save cumbersome notation):

$$\begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{B}_{22} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{y}_t \\ \Delta \mathbf{z}_t \end{pmatrix} = \sum_{j=1}^{s-1} \mathbf{P}_j \Delta \mathbf{x}_{t-j} + \boldsymbol{\alpha}_0^\diamond (\boldsymbol{\beta}' \mathbf{x}_{t-1}) + \mathbf{F}_0^\diamond \mathbf{q}_t + \boldsymbol{\epsilon}_t^\diamond \quad (54)$$

where:

$$\begin{pmatrix} \boldsymbol{\epsilon}_{1t}^\diamond \\ \boldsymbol{\epsilon}_{2t}^\diamond \end{pmatrix} \sim \text{IN} \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11}^\diamond & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^\diamond \end{pmatrix} \right]. \quad (55)$$

is a possible block recursive system, with  $\Delta \mathbf{z}_t$  being predetermined (see Fisher, 1965).

### 5.8 Triangular representations

Phillips (1991) considers this as a basic formulation for examining long-run relations. Omitting deterministic factors, a triangular representation has the form:

$$\begin{pmatrix} \mathbf{x}_{1t} \\ \Delta \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}' \mathbf{x}_{2t} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_{1t} \\ \boldsymbol{\xi}_{2t} \end{pmatrix} \quad (56)$$

where  $\boldsymbol{\xi}_t$  is assumed to be a stationary process which satisfies suitable homogeneity and memory conditions (such as mixing). However, (56) only focuses on the long run and does not consider modelling the short-run dynamics as well. When  $\{\boldsymbol{\xi}_t\}$  has a finite autoregressive representation, as assumed for the other members of the typology above, such that  $\mathbf{R}(L)\boldsymbol{\xi}_t = \mathbf{u}_t$  where  $\mathbf{R}_0 = \mathbf{I}$  then:

$$\mathbf{R}(L) \begin{pmatrix} \boldsymbol{\beta}' \mathbf{x}_t \\ \Delta \mathbf{x}_{2t} \end{pmatrix} = \mathbf{u}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}_u] \quad (57)$$

This expresses the system in terms of the two sets of  $I(0)$  variables,  $\boldsymbol{\beta}' \mathbf{x}_t$  and  $\Delta \mathbf{x}_{2t}$  and lagged values thereof, but could be reparameterized to a VEC by rearranging lags and differences.

In the 2-equation example when  $\mathbf{R}(L)\boldsymbol{\xi}_t = \mathbf{u}_t$  and there is only one lag:

$$\begin{pmatrix} x_{1t} - \delta x_{2t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} r_{11} & 0 \\ r_{21} & 0 \end{pmatrix} \begin{pmatrix} x_{1t-1} - \delta x_{2t-1} \\ \Delta x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (58)$$

where:

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim \text{IN} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right]. \quad (59)$$

This imposes the one cointegration vector, but without additional restrictions (such as  $r_{21} = 0$ ) the system requires joint analysis and no direction of weak exogeneity holds. (58) is readily seen to be a reparameterization of (37).

## 6 Analyzing dynamic models

Once a model form is selected as optimal given the theory and data process, the issue of how to analyze it arises. In particular, closed models entail that contemporaneous partial derivatives are hard to interpret, since it is far from obvious which experiment could feasibly hold ‘all other things constant’ to examine the link in question. When conditioning is valid, analysis of the conditional model is straightforward in that derivatives of the form:

$$\frac{\partial \mathbf{y}_t}{\partial \mathbf{z}_{t-j}}$$

are well defined as shown below. Here we consider possible methods for closed systems which have non-diagonal error covariance matrices.

Write the original formulation in (14) in lag polynomial triangular form as in (57) but retaining deterministic terms, and under stationarity invert  $\mathbf{R}(L) = \sum_{j=1}^s \mathbf{R}_j L^j$  to obtain:

$$\begin{pmatrix} \beta' \mathbf{x}_t \\ \Delta \mathbf{x}_{2t} \end{pmatrix} = \mathbf{R}(L)^{-1} (\mathbf{F} \mathbf{q}_t + \mathbf{u}_t) = \mathbf{S}(L) \mathbf{F} \mathbf{q}_t + \boldsymbol{\xi}_t \quad (60)$$

where  $\boldsymbol{\xi}_t = \mathbf{S}(L) \mathbf{u}_t$  and  $\mathbf{S}(L) = \mathbf{R}(L)^{-1}$ . When  $\boldsymbol{\Omega}_u$  is non-diagonal, the reformulation underlying a causal chain could be used by writing  $\boldsymbol{\Omega}_u^{-1} = \mathbf{H}_u \mathbf{H}'_u$  (say) and treating each element of  $\mathbf{w}_t = \mathbf{H}'_u \mathbf{u}_t$  as analyzing ‘autonomous’ changes in  $w_{it}$ . This introduces conditioning indirectly and camouflages the resulting loss of structurality should the selected ordering not coincide with that underlying the economic mechanism. In particular, although the invariance of  $\mathbf{S}(L)$  derives from that of the parameters of the structural mechanism, they are no longer structural and changes in any element of the structure are liable to alter many elements of  $\mathbf{S}(L)$  (see Hendry and Mizon, 1993, for further details). Valid conditional models are easier to analyze from this perspective.

The analysis of simultaneous models has also posed problems for some commentators (see e.g. Pearl, 1993 for a good summary of recent debates). First, granted that conditioning on contemporaneous variables is admissible, then the simultaneous system is a model of the conditional expectation of the endogenous vector given the weakly exogenous and lagged variables as discussed above. However, one cannot consider the first equation as being (say)  $\mathbf{E}[\Delta x_{1t} | \Delta x_{2t}, \mathbf{x}_{t-1}]$ . For example, (42) corresponds to the system:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} c_{11} - b_{12}c_{21} & -b_{12}c_{22} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1^0 \\ \mu_2^0 \end{pmatrix} + \begin{pmatrix} \nu_{1t}^0 \\ \nu_{2t}^0 \end{pmatrix} \quad (61)$$

To see the hidden proportionality between the two equations, consider the case where  $c_{11} = 0$ . There are then 9 coefficients in (61) whereas the simultaneous representation has 8 parameters: more dramatic ‘savings’ result in larger systems. However, ‘simultaneity’ is not invariant under linear transformations; when  $x_{1t}, x_{2t}$  represent (say) consumption and income with  $b_{12} = \delta$  then the first equation in (43) becomes saving, leading to the model:

$$\begin{pmatrix} \Delta (x_{1t} - \delta x_{2t}) \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} c_{11}^* \\ c_{22}^* \end{pmatrix} \begin{pmatrix} x_{1t-1} - \delta x_{2t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} + \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} \quad (62)$$

which is not simultaneous, but could be structural. Once one understands how to analyze (62), then (42) should not prove problematical by following the reverse steps.

## 7 Modelling dynamic systems

The final issue concerns modelling strategies and although these are the subject of Hendry and Doornik (1994), a brief discussion is useful.



General to simple methods have played a major role in econometric modelling and previous analyses have considered their research efficiency (directed versus directionless strategies; interpreting test outcomes; avoiding the *non sequitur* of adopting the alternative when a test rejects; determining the baseline innovation error; and correcting flaws as against commencing from a congruent model). The methodological basis for general to simple modelling is discussed in Hendry and Doornik (1994) who present ten inter-related arguments for commencing econometric analyses of economic time series from the joint density. These include the obvious points that the economy is a system; marginalization cannot otherwise be tested (including hypotheses about Granger causality – see Granger (1969) and Cook and Hendry (1993)); simultaneity; tests of weak exogeneity; checking identification (i.e. the uniqueness of the representation); cointegration; tests of cross equation dependencies; and conducting  $h$ -step forecasts when  $h > 1$ . There are also less obvious requirements such as ensuring that nesting entails encompassing; and testing super exogeneity and invariance (generalizing the tests in Engle and Hendry, 1993). The above analysis reveals the further argument that dynamic adjustment of an individual equation need not be interpretable outside of the system context.

## 8 Conclusion

The key element of ‘choice’ between the preceding representations must be the ones which capture the invariants of the economic mechanism. The two typologies are valuable only in so far as they facilitate useful specifications of systems and models for empirical research. There do not seem to be high-level considerations for any one form being dominant, although some forms (such as VARs) seem more likely to be derived than autonomous whereas others (such as conditional VECMs) could correspond more closely to the underlying economic structure. The properties of the various system types have been studied extensively in single equation forms, although rather less has been written about their comparative behaviour in larger systems. Similarly, a complete comparative study of the properties of models can throw light on specification problems. The next step would be to implement all of the members of the two typologies in both empirical and Monte Carlo analyses (extending the single equation findings in Hendry, 1995) to investigate the extent to which they do or do not capture structure in practice and to examine their behaviour in samples of the size relevant to time-series econometrics.

## References

- Artis, M. J., Bladen-Hovell, R. C., Osborn, D. R., Smith, J. P., and Zhang, W. (1993). Turning point prediction in the UK: Preliminary results using CSO leading indicators. Presented to the Royal Economic Society Conference, York.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.
- Boswijk, H. P. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.
- Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.
- Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.
- Clements, M. P., and Hendry, D. F. (1994). Forecasting in cointegrated systems. *Journal of Applied Econometrics*. Forthcoming.
- Clements, M. P., and Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review*, **35**, 887–932.
- Cook, S., and Hendry, D. F. (1993). The theory of reduction in econometrics. *Poznań Studies in the Philosophy of the Sciences and the Humanities*, **38**, 71–100.
- Davidson, J. E. H., and Hall, S. (1991). Cointegration in recursive systems. *Economic Journal*, **101**, 239–251.

- Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427–431.
- Diebold, F. X., and Rudebusch, G. D. (1991). Forecasting output with the composite leading index: An ex ante analysis. *Journal of the American Statistical Association*, **86**, 603–610.
- Doan, T., Litterman, R., and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**, 1–100.
- Doornik, J. A., and Hendry, D. F. (1994). *PcFiml 8: An Interactive Program for Modelling Econometric Systems*. London: International Thomson Publishing.
- Emerson, R. A. (1994). Essays on leading indicators. Unpublished doctoral thesis, Oxford University, UK.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.
- Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.
- Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Ericsson, N. R. (1992). Cointegration, exogeneity and policy analysis: An overview. *Journal of Policy Modeling*, **14**, 251–280.
- Ericsson, N. R. (1993). Comment on 'On the limitations of comparing mean squared forecast errors', by M.P. Clements and D.F. Hendry. *Journal of Forecasting*, **12**, 617–637.
- Fisher, F. M. (1965). Dynamic structure and estimation in economy-wide econometric models. In Duesenberry, J. S., Klein, L. R., Fromm, G., and Kuh, E. (eds.), *Brookings Quarterly Econometric Model of the United States*. Amsterdam: North-Holland Publishing Company.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, **11**, 1–12.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, **12**, 1–118. Supplement.
- Hendry, D. F. (1993). The roles of economic theory and econometrics in time-series economics. Invited address, European Econometric Society, Stockholm.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F., and Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.
- Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.
- Hendry, D. F., and Mizon, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549–563. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.
- Hendry, D. F., Neale, A. J., and Srba, F. (1988). Econometric analysis of small linear systems using PC-FIML. *Journal of Econometrics*, **38**, 203–226.
- Hendry, D. F., Pagan, A. R., and Sargan, J. D. (1984). Dynamic specification. In Griliches, Z., and Intriligator,

- M. D. (eds.), *Handbook of Econometrics*, Vol. 2–3, Ch. 18. Amsterdam: North-Holland. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F., and Richard, J.-F. (1983). The econometric analysis of economic time series (with discussion). *International Statistical Review*, **51**, 111–163. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F., and Richard, J.-F. (1989). Recent developments in the theory of encompassing. In Cornet, B., and Tulkens, H. (eds.), *Contributions to Operations Research and Econometrics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1992a). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.
- Johansen, S. (1992b). Testing weak exogeneity and the order of cointegration in UK money demand. *Journal of Policy Modeling*, **14**, 313–334.
- 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.
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.
- Mizon, G. E. (1993). A simple message to autocorrelation correctors: Don't. Discussion paper, European University Institute, Florence.
- Monfort, A., and Rabemananjara, R. (1990). From a VAR model to a structural model, with an application to the wage-price spiral. *Journal of Applied Econometrics*, **5**, 203–227.
- Neftci, S. N. (1979). Lead-lag relations, exogeneity and prediction of economic time series. *Econometrica*, **47**, 101–113.
- Pearl, J. (1993). On the statistical interpretation of structural equations. Technical report r-200, Cognitive Systems Lab., UCLA.
- Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica*, **59**, 283–306.
- 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.
- Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, **58**, 113–144.
- Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.
- Stock, J. H., and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macro-Economic Annual*, 351–409.
- Stock, J. H., and Watson, M. W. (1992). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. Working paper 4014, NBER.
- Wold, H. O. A., and Juréen, L. (1953). *Demand Analysis: A Study in Econometrics* 2nd edn. New York: John Wiley.