Explaining Cointegration Analysis: Part II

David F. Hendry
Department of Economics, Oxford University, UK

and

Katarina Juselius
Department of Economics, University of Copenhagen, Denmark

November 12, 2000

Abstract We describe the concept of cointegration, its implications in modelling and forecasting, and discuss inference procedures appropriate in integrated-cointegrated vector autoregressive processes (VARs). Particular attention is paid to the properties of VARs, to the modelling of deterministic terms, and to the determination of the number of cointegration vectors. The analysis is illustrated by empirical examples.

1 Introduction

Hendry and Juselius (2000) investigated the properties of economic time series that were integrated processes, such as random walks, which contained a unit root in their dynamics. Here we extend the analysis to the multivariate context, and focus on cointegration in systems of equations.

We showed in Hendry and Juselius (2000) that when data were non-stationary purely due to unit roots (integrated once, denoted $\mathsf{I}(1)$), they could be brought back to stationarity by the linear transformation of differencing, as in $x_t - x_{t-1} = \Delta x_t$. For example, if the data generation process (DGP) were the simplest random walk with an independent normal (IN) error having mean zero and constant variance σ_ϵ^2:

$$x_t = x_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right], \tag{1}$$

then subtracting x_{t-1} from both sides delivers $\Delta x_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ which is certainly stationary.[1] Since Δx_t cannot have a unit root, it must be $\mathsf{I}(0)$. Such an analysis generalizes to (say) twice-integrated series – which are $\mathsf{I}(2)$– so must become $\mathsf{I}(0)$ after differencing twice.

It is natural to enquire if other linear transformations than differencing will also induce stationarity. The answer is 'possibly', but unlike differencing, there is no guarantee that the outcome must be $\mathsf{I}(0)$: cointegration analysis is designed to find linear combinations of variables that also remove unit roots. In a bivariate context, if y_t and x_t are both $\mathsf{I}(1)$, there may (but need not) be a unique value of β such that $y_t - \beta x_t$ is $\mathsf{I}(0)$: in other words, there is no unit root in the relation linking y_t and x_t. Consequently, cointegration is a restriction on a dynamic model, and so is testable. Cointegration vectors are of considerable interest when they exist, since they determine $\mathsf{I}(0)$ relations that hold between variables which are individually non-stationary. Such relations are often called 'long-run equilibria', since it can be proved that they act as 'attractors' towards which convergence occurs whenever there are departures therefrom (see e.g., Granger, 1986, and Banerjee, Dolado, Galbraith and Hendry, 1993, ch. 2).

Since $\mathsf{I}(1)$ variables 'wander' (often quite widely) because of their stochastic trends, whereas (weakly) stationary variables have constant means and variances, if there exists a linear combination that delivers an $\mathsf{I}(0)$ relation, it might be thought that it would be obvious from graphs of the variables. Unfortunately, that need not be the case, as figure 1 shows. In panel a, three variables (denoted y_a, y_b, y_c) are plotted that are actually very strongly cointegrated, whereas panel b plots another two (denoted Y_a, Y_b) that are neither cointegrated, nor linked in any causal way. Panels c and d respectively show the changes in these variables. It is not obvious from the graphs that the first set are closely linked whereas the second are not connected at all. Nevertheless, the *PcGive* cointegration test, described in Hendry and Juselius (2000), applied to simple dynamic models relating y_a to y_b and y_c, and Y_a to Y_b, respectively takes the values $\mathsf{t}_{ur} = -6.25^{**}$ and $\mathsf{t}_{ur} = -2.31$, so the first strongly rejects a unit root in the relation, whereas the second does not. Thus, cointegration may or may not exist between variables that do or do not 'look cointegrated', and the only way to find out is through a careful statistical analysis, rather than rely on visual inspection. These two points, namely the importance but non-obvious nature of cointegrated relations, motivates our discussion.

The organization of this paper is as follows. Section 2 begins by illustrating the inherently multivariate nature of cointegration analysis: several variables must be involved, and this determines the form of the statistical tools required. Section 3 then discusses the conditions under which a vector autoregressive process (VAR) would provide a feasible empirical model for integrated economic time series, spelling out both its statistical and economic requirements, illustrated by the empirical example used in

[1] Notice that differencing is not an *operator* for equations: one can difference data (to create Δx_t), but attempting to difference equation (1) would lead to $\Delta x_t = \Delta x_{t-1} + \Delta \epsilon_t$. Such an equation is not well defined, since Δ can be cancelled on both sides, so is redundant.

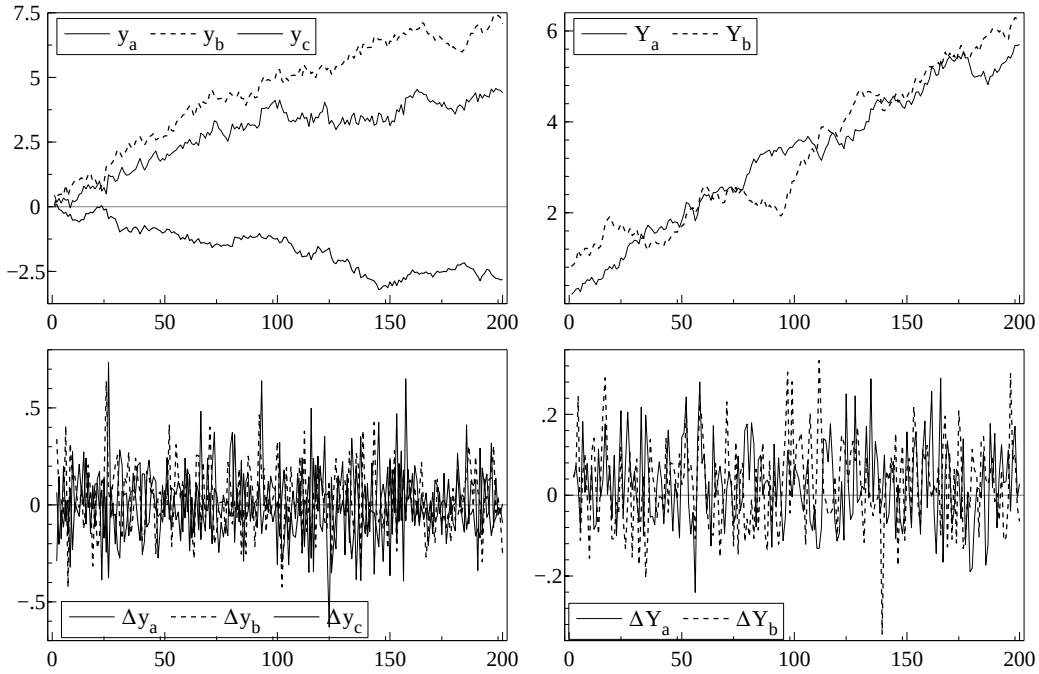

Figure 1 Cointegrated and non-cointegrated time series.

Hendry and Juselius (2000). In section 4, we consider alternative representations of the VAR that yield different insights into its properties under stationarity, and also set the scene for deriving the necessary and sufficient conditions that deliver an integrated-cointegrated process. The purpose of section 5 is to define cointegration via restrictions on the VAR model, and relate the properties of the vector process to stochastic trends and stationary components based on the moving-average representation. Section 6 considers the key role of deterministic terms (like constants and trends) in cointegration analyses.

At that stage, the formalization of the model and analysis of its properties are complete, so we turn to issues of estimation (section 7) and inference (section 8), illustrated empirically in section 9. Section 10 considers the identification of the cointegration parameters, and hypothesis tests thereon, and section 11 discusses issues that arise in the analysis of partial systems (conditional on a subset of the variables) and the closely related concept of exogeneity. Finally, we discuss forecasting in cointegrated systems (section 12) and the associated topic of parameter constancy (section 13), also relevant to any policy applications of cointegrated systems. Section 14 concludes. The paper uses matrix algebra extensively to explain the main ideas, so we adopt the following notation: bold-face capital letters for matrices, bold-face lower case for vectors, normal case for variables and coefficients, and Greek letters for parameters. We generally assume all variables are in logs, which transformation produces more homogeneous series for inherently-positive variables (see e.g., Hendry, 1995a, ch. 2), but we will not distinguish explicitly between logs and the original units.[2]

[2]If a variable had a unit root in its original units of measurement, it would become essentially deterministic over time if it had a constant error variance. Thus, absolute levels must have heteroscedastic errors to make sense; but if so, that is not a sensible place to start modeling. Moreover, if the log had a unit root, then the original must be explosive. Many economic variables seem to have that property, appearing to show quadratic trends in absolute levels.

2 The multivariate nature of cointegration analysis

Cointegration analysis is inherently multivariate, as a single time series cannot be cointegrated. Consequently, consider a set of integrated variables, such as gasoline prices at different locations as in Hendry and Juselius (2000), where each individual gasoline price (denoted $p_{i,t}$) is I(1), but follows a common long-run path, affected by the world price of oil ($p_{o,t}$). Cointegration between the gasoline prices could arise, for example, if the price differentials between any two locations were stationary. However, cointegration as such does not say anything about the direction of causality. For example, one of the locations could be a price leader and the others price followers; or, alternatively, none of the locations might be more important than the others. In the first case, the price of the leading location would be driving the prices of the other locations (be 'exogenous' to the other prices) and cointegration could be analyzed from the equations for the other 'adjusting' prices, given the price of the leader. In the second case, all prices would be 'equilibrium adjusting' and, hence, all equations would contain information about the cointegration relationships. In the bivariate analysis in Hendry and Juselius (2000), cointegration was found in a single-equation model of $p_{1,t}$ given $p_{2,t}$, thereby assuming that $p_{2,t}$ was a price leader. If this assumption was incorrect, then the estimates of the cointegration relation would be inefficient, and could be seriously biased. To find out which variables adjust, and which do not adjust, to the long-run cointegration relations, an analysis of the full system of equations is required, as illustrated in Section 11.

Here, we will focus on a vector autoregression (VAR) as a description of the system to be investigated. In a VAR, each variable is 'explained' by its own lagged values, and the lagged values of all other variables in the system. To see which questions can be asked within a cointegrated VAR, we postulate a trivariate VAR model for the two gasoline prices $p_{1,t}$ and $p_{2,t}$, together with the price of crude oil, $p_{o,t}$. We restrict the analysis to one lagged change for simplicity, and allow for 2 cointegration relations. Then the system can be written as:

$$\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{o,t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{o,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} (p_1 - p_2)_{t-1} \\ (p_2 - p_o)_{t-1} \end{pmatrix} + \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} d_{1,t} + \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\epsilon}_t$ is assumed $\mathsf{IN}_3[\mathbf{0}, \boldsymbol{\Omega}_\epsilon]$, and $\boldsymbol{\Omega}_\epsilon$ is the (positive-definite, symmetric) covariance matrix of the error process.

Within the hypothetical system (2), we could explain the three price changes from period $t-1$ (previous week) to t (this week) as a result of:

(i) an adjustment to previous price changes, with impacts ϕ_{ij} for the j^{th} lagged change in the i^{th} equation;

(ii) an adjustment to previous disequilibria between prices in different locations, $(p_1 - p_2)$, and between the price in location 2 and the price of crude oil, $(p_2 - p_o)$, with impacts α_{i1} and α_{i2} respectively in the i^{th} equation;

(iii) an extraordinary intervention in the whole market, such as the outbreak of the Kuwait war, described by the intervention dummy $d_{1,t}$;

(iv) a constant term $\boldsymbol{\pi}$; and

(v) random shocks, $\boldsymbol{\epsilon}_t$.

When all three prices are I(1), whereas $(p_{1,t} - p_{2,t})$ and $(p_{2,t} - p_{o,t})$ are I(0), then the latter describe cointegrated relations, i.e., relations that are stationary even when the variables themselves are non-stationary. Cointegration between the prices means that the three prices follow the same long-run trends, which then cancel in the price differentials. This may seem reasonable *a priori*, but could nevertheless be incorrect empirically: using multivariate cointegration analysis, we can formally test whether such is indeed the case. In general, we write cointegration relations in the form:

$$\beta_{11}p_{1,t} + \beta_{12}p_{2,t} + \beta_{13}p_{o,t}; \text{ and } \beta_{21}p_{1,t} + \beta_{22}p_{2,t} + \beta_{32}p_{o,t} \tag{3}$$

etc., where, in (2), we have normalized $\beta_{11} = \beta_{22} = 1$, and set $\beta_{13} = \beta_{21} = 0$. Such restrictions cannot be imposed arbitrarily in empirical research, so we will discuss how to test restrictions on cointegration relations in Section 10.

The existence of cointegration by itself does imply which prices 'equilibrium adjust' and which do not; nor does it entail whether any adjustment is fast or slow. Information about such features can be provided by the α_{ij} coefficients. For example, $\alpha_{31} = \alpha_{32} = 0$, would tell us that there were no feed-back effects onto the price of crude oil from 'deviant' price behavior in the gasoline market. In this case, the price of crude oil would influence gasoline prices, but would not be influenced by them. Next, consider, for example, when $\alpha_{11} = -0.6$ whereas $\alpha_{21} = -0.1$. Then, gasoline prices at location 1 adjust more quickly to restore an imbalance between its own price and the price at location 2 than the other way around. Finally, consider $\alpha_{22} = -0.4$: then the location-2 price would adjust quite quickly to changes in the level of crude oil price. In that case, we would be inclined to say that the price of crude oil influenced the price of location 2 which influenced the price at location 1. This would certainly be the case if the covariance matrix $\mathbf{\Omega}_\varepsilon$ was diagonal, so there were no contemporaneous links: if $\mathbf{\Omega}_\varepsilon$ was not diagonal, revealing cross-correlated residuals, one would have to be careful about 'causal' interpretations. The interpretation of the parameter estimates is generally more straightforward when $\mathbf{\Omega}_\varepsilon$ is diagonal, but this is seldom the case; price shocks are often correlated, sometimes indicating an modeled causal link. For example, if a shock to crude-oil prices affects gasoline prices within the same week (say), the correlations between the residuals from the crude-oil equation and the gasoline equations are the result of a current oil-price effect. Another explanation for residual cross-correlation is omitted variables that simultaneously influence several variables in the system.

As already discussed in Hendry and Juselius (2000), the constant terms, π_i, can both describe an intercept in the cointegration relations and linear trends in the variables, and the empirical analysis can be used to estimate both effects. Finally, the random shocks, $\epsilon_{i,t}$, are assumed to be serially independent, and homoscedastic. All these, and other issues, will be discussed and illustrated with an empirical example below.

It should now be evident that a cointegrated VAR provides a rich model: the β_{ij} coefficients characterize long-run relationships between levels of variables; the α_{ij} coefficients describe changes that help restore an equilibrium market position; the ϕ_{ij} coefficients describe short-term changes resulting from previous changes in the market – which need not have permanent effects on the levels; we will comment on the intercepts in Section 6; and the intervention effects, ψ_i, describe extraordinary events in the market, like the Kuwait war. We might, for example, ask whether such an event affected the gasoline prices differently at different locations (i.e., whether $\psi_1 \neq \psi_2$), and hence, permanently changed the mean of the price differential $(p_1 - p_2)$: this will be investigated empirically in Section 9.1. First, we must analyze the VAR model more generally.

3 The statistical adequacy of a VAR model

To understand when a VAR is an adequate description of reality, it is important to know the limitations as well as the possibilities of that model. The purpose of this section is, therefore, to demonstrate that a VAR model can be a convenient way of summarizing the information given by the autocovariances of the data under certain assumptions about the DGP: see Hendry (1995a) for details. However, the required assumptions may not hold in any given instance, so the first step in any empirical analysis of a VAR is to test if these assumptions are indeed appropriate.

In section 3.1, we first assume that there are $p \geq 2$ variables $x_{i,t}$ under analysis, and that the p-dimensional process $\mathbf{x}_t$ is stationary, so does not contain any unit roots. We then derive the VAR model under this simplifying assumption. In section 3.2, we discuss an interpretation of the VAR in terms of rational economic behavior, and finally, in section 3.4, we extend the discussion to consider stability and unit-root properties. Notice that unit roots are a restriction of the initial VAR model, so can be tested, but it transpires that the tests are not standard t, F or χ^2.

3.1 Stochastic properties

A stationary VAR arises naturally as a model of a data set $(\mathbf{x}_1, \ldots, \mathbf{x}_T)'$ viewed as a sequence of T realizations from the p-dimensional process $\{\mathbf{x}_t\}$, given the two general simplifying assumptions of multivariate normality and time-invariant covariances. Derivations of a VAR from a general DGP are described in (e.g.) Hendry (1995a, ch. 9). The resulting VAR could have many lagged variables, but for simplicity of notation, we restrict attention here to the case with 2 lags denoted VAR(2), which suffices to illustrate all the main properties – and problems (the results generalize easily, but lead to more cumbersome notation). We write the simplest VAR(2) as:

$$\mathbf{x}_t = \boldsymbol{\pi} + \boldsymbol{\Pi}_1 \mathbf{x}_{t-1} + \boldsymbol{\Pi}_2 \mathbf{x}_{t-2} + \boldsymbol{\epsilon}_t \tag{4}$$

where

$$\boldsymbol{\epsilon}_t \sim \mathsf{IN}_p \left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon \right], \tag{5}$$

$t = 1, \ldots, T$ and the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \boldsymbol{\Omega}_\epsilon)$ are constant and unrestricted, except for $\boldsymbol{\Omega}_\epsilon$ being positive-definite and symmetric.

Given (4), the conditional mean of $\mathbf{x}_t$ is:

$$\mathsf{E}\left[\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2} \right] = \boldsymbol{\pi} + \boldsymbol{\Pi}_1 \mathbf{x}_{t-1} + \boldsymbol{\Pi}_2 \mathbf{x}_{t-2} = \mathbf{m}_t,$$

say, and the deviation of $\mathbf{x}_t$ from $\mathbf{m}_t$ defines $\boldsymbol{\epsilon}_t$:

$$\mathbf{x}_t - \mathbf{m}_t = \boldsymbol{\epsilon}_t.$$

Hence, if the assumptions of multivariate normality, time-constant covariances, and truncation at lag 2 are correct, then (4):

- is linear in the parameters;
- has constant parameters;
- has normally distributed errors $\boldsymbol{\epsilon}_t$, with:
- (approximate) independence between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\epsilon}_{t-h}$ for lags $h = 1, 2, \ldots$.

These conditions provide the model builder with testable hypotheses on the assumptions needed to justify the VAR. In economic applications, the multivariate normality assumption is seldom satisfied.

This is potentially a serious problem, since derivations of the VAR from a general DGP rely heavily on multivariate normality, and statistical inference is only valid to the extent that the assumptions of the underlying model are correct. An important question is, therefore, how we should modify the standard VAR model in practice. We would like to preserve its attractiveness as a reasonably tractable description of the basic characteristics of the data, while at the same time, achieving valid inference. Simulation studies have demonstrated that statistical inference is sensitive to the validity of some of the assumptions, such as, parameter non-constancy, serially-correlated residuals and residual skewness, while moderately robust to others, such as excess kurtosis (fat-tailed distributions) and residual heteroscedasticity. Thus, it seems advisable to ensure the first three are valid. Both direct and indirect testing of the assumptions can enhance the success of the empirical application. It is often useful to calculate descriptive statistics combined with a graphical inspection of the residuals as a first check of the adequacy of the VAR model, then undertake formal mis-specification tests of each key assumption (see Doornik and Hendry, 1999, ch. 10: all later references to specific tests are explained there). Once we understand why a model fails to satisfy the assumptions, we can often modify it to end with a 'well-behaved' model. Precisely how depends on the application, as will be illustrated in section 3.3 for the gasoline price series discussed in Hendry and Juselius (2000).

3.2 Economic interpretation and estimation

As discussed in Hendry (1995a), the conditional mean $\mathbf{m}_t$ can be given an economic interpretation as the agents' plans at time $t-1$ given the past information of the process, $\mathbf{x}_{t-1}$, $\mathbf{x}_{t-2}$, etc., denoted $\mathbf{X}_{t-1}^0$. The IN distributional assumption in (5) implies that agents are rational, in the sense that the deviation between the actual outcome $\mathbf{x}_t$ and the plan $\mathsf{E}_{t-1}[\mathbf{x}_t|\mathbf{X}_{t-1}^0]$ is a white-noise innovation, not explicable by the past of the process. Thus, the VAR model is consistent with economic agents who seek to avoid systematic forecast errors when they plan for time t based on the information available at time $t-1$.

By way of contrast, a VAR with autocorrelated residuals would describe agents that do not use all information in the data as efficiently as possible. This is because they could do better by including the systematic variation left in the residuals, thereby improving the accuracy of their expectations about the future. Checking the assumptions of the model, (i.e., checking the white-noise requirement of the residuals, and so on), is not only crucial for correct statistical inference, but also for the economic interpretation of the model as a description of the behavior of rational agents.

To derive a full-information maximum likelihood (FIML) estimator requires an explicit probability formulation of the model. Doing so has the advantage of forcing us to take the statistical assumptions seriously. Assume that we have derived an estimator under the assumption of multivariate normality. We then estimate the model, and find that the residuals are not normally distributed, or that the residual variance is heteroscedastic instead of homoscedastic, or that residuals exhibit significant autocorrelation, etc. The parameter estimates (based on an incorrectly-derived estimator) may not have any meaning, and since we do not know their 'true' properties, inference is likely to be hazardous. Therefore, to claim that conclusions are based on FIML inference is to claim that the empirical model is capable of accounting for all the systematic information in the data in a satisfactory way.

Although the derivation of a FIML estimator subject to parameter restrictions can be complicated, this is not so when the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \boldsymbol{\Omega}_\epsilon)$ of the VAR model (4) are unrestricted. In that case, the ordinary least squares (OLS) estimator is equivalent to FIML. After the model has been estimated by OLS, the IN distributional assumption can be checked against the data using the residuals $\widehat{\boldsymbol{\epsilon}}_t$. As already mentioned, the white-noise assumption is often rejected for a first tentatively-estimated model, and one has to modify the specification of the VAR model accordingly. This can be done, for example, by:

1987:26–98:29	$\overline{\mathbf{x}}$	$s_{\overline{x}}$	*Skew*	*Ex.K.*	*Jarq.Bera*	*min*	*max*
p_1	-0.489	0.149	0.39	0.66	15.5	-0.91	0.02
p_2	-0.499	0.139	0.17	0.42	6.1	-0.92	-0.06
Δp_1	-0.00010	0.032	0.04	1.38	210.3	-0.13	0.21
Δp_2	-0.00026	0.025	-0.03	-0.33	101.9	-0.11	0.14
$\widehat{\epsilon}_{p_1}$	0.0	0.025	0.27	4.64	214.9	-0.09	0.17
$\widehat{\epsilon}_{p_2}$	0.0	0.020	0.01	2.47	90.7	-0.08	0.11

Table 1 Descriptive statistics.

- investigating parameter constancy (e.g., 'is there a structural shift in the model parameters'?);
- increasing the information set by adding new variables;
- increasing the lag length;
- changing the sample period;
- adding intervention dummies to account for significant political or institutional events;
- conditioning on weakly-exogenous variables;
- checking the adequacy of the measurements of the chosen variables.

Any or all of these steps may be needed, but we stress the importance of checking that the initial VAR is 'congruent' with the data evidence before proceeding with empirical analysis.

3.3 A tentatively-estimated VAR

As a first step in the analysis, the unrestricted VAR(2) model, with a constant term and without dummy variables, was estimated by OLS for the two gasoline prices at different locations. Table 1 reports some descriptive statistics for the logs of the variables in levels, differences, and for the residuals. As discussed in Hendry (1995a), since the gasoline prices are apparently non-stationary, the empirical density is not normal, but instead bimodal. The price changes on the other hand, seem to be stationary around a constant mean. From Table 1, the mean is not significantly different from zero for either price change. Normality is tested with the Jarque–Bera test, distributed as $\chi^2(2)$ under the null, so is strongly rejected for both series of price changes and the VAR residuals. Since rejection could be due to either excess kurtosis (the normal has kurtosis of 3), or skewness, we report these statistics separately in Table 1. It appears that excess kurtosis is violated for both equations, but that for $p_{1,t}$ also seems to be skewed to the right.

The graphs (in logs) of the two gasoline prices are shown in Figure 2 in levels and differences (with 99% confidence bands), as well as the residuals (similarly with 99% confidence bands). There seem to be some outlier observations (larger than $\pm 3\widehat{\sigma}$), both in the differenced prices and in the residuals. The largest is at 1990:31, the week of the outbreak of the Kuwait war. This is clearly not a 'normal' observation, but must be adequately accounted for in the specification of the VAR model. The remaining large price changes ($> \pm 3\widehat{\sigma}$) can (but need not necessarily) be intervention outliers. It is, however, always advisable to check whether any large changes in the data correspond to some extraordinary events: exactly because they are big, they will influence the estimates with a large weight and, hence, potentially bias the estimates if they are indeed outliers. The role of deterministic components, such as intervention dummies, will be discussed in more detail in Section 6.

In Figure 3, the autocorrelograms and the empirical densities (with the normal density) are reported for the two VAR residuals. There should be no significant autocorrelation, if the truncation after the

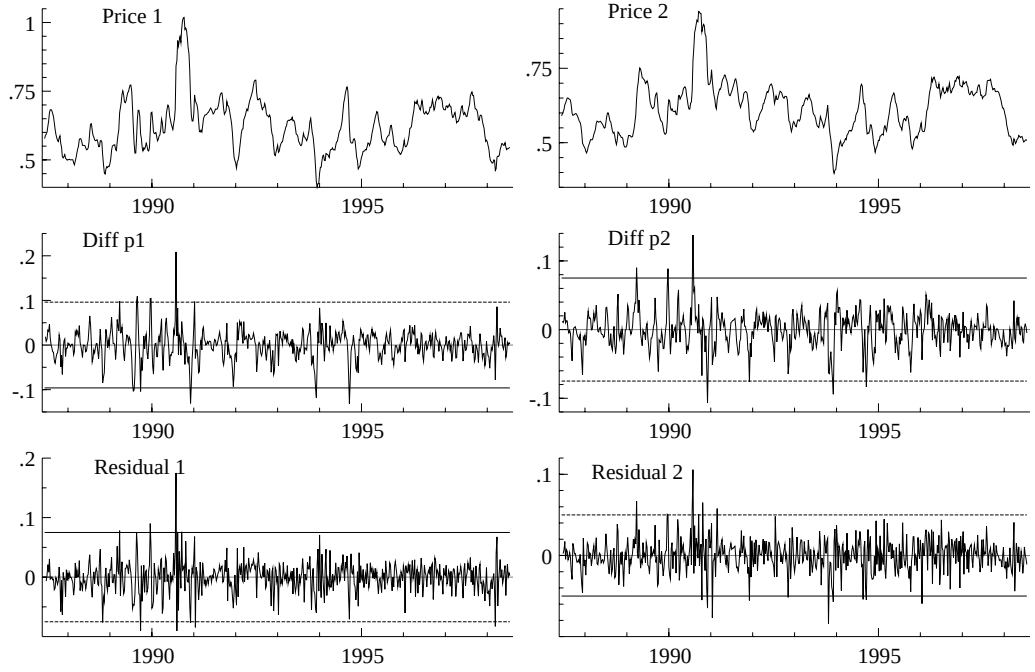

Figure 2 Graphs of gasoline prices 1 and 2 in levels and differences, with residuals from a VAR(2) and 99% confidence bands.

second lag is appropriate. Since all the autocorrelation coefficients are very small, this seems to be the case. Furthermore, the empirical density should not deviate too much from the normal density; and the residuals should be homoscedastic, so have similar variances over time. The empirical densities seem to have longer tails (excess kurtosis) than the normal density, and the Kuwait war outlier sticks out, confirming our previous finding of non-normality.

3.4 Stability and unit-root properties

Up to this point, we have discussed (and estimated) the VAR model as if it were stationary, i.e., without considering unit roots.[3] The dynamic stability of the process in (4) can be investigated by calculating the roots of:

$$\left(\mathbf{I}_p - \mathbf{\Pi}_1 L - \mathbf{\Pi}_2 L^2\right) \mathbf{x}_t = \mathbf{\Pi}(L)\mathbf{x}_t,$$

where $L^i \mathbf{x}_t = \mathbf{x}_{t-i}$. Define the characteristic polynomial:

$$\mathbf{\Pi}(z) = \left(\mathbf{I}_p - \mathbf{\Pi}_1 z - \mathbf{\Pi}_2 z^2\right).$$

The roots of $|\mathbf{\Pi}(z)| = 0$ contain all necessary information about the stability of the process and, therefore, whether it is stationary or non-stationary. In econometrics, it is more usual to discuss stability in terms of the companion matrix of the system, obtained by stacking the variables such that a first-order system results. Ignoring deterministic terms, we have:

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 \\ \mathbf{I}_p & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \end{pmatrix}, \tag{6}$$

[3] One can always estimate the unrestricted VAR with OLS, but if there are unit roots in the data, some inferences are no longer standard, as discussed in Hendry and Juselius (2000).

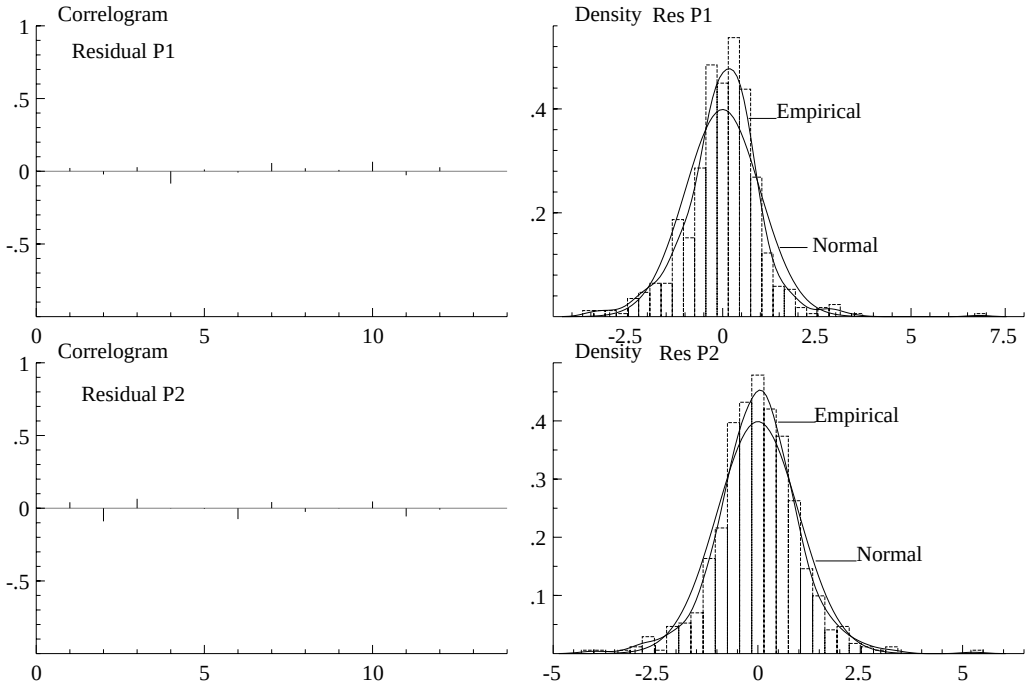

Figure 3 The autocorrelogram and the empirical and normal density of the two VAR residuals.

where the first block is the original system, and the second merely an identity for $\mathbf{x}_{t-1}$. Now, stability depends on the eigenvalues of the coefficient matrix in (6), and these are precisely the roots of $\left|\mathbf{\Pi}\left(z^{-1}\right)\right| = 0$ (see e.g., Banerjee *et al.*, 1993). For a p-dimensional VAR with 2 lags, there are $2p$ eigenvalues. The following results apply:

(a) if all the eigenvalues of the companion matrix are inside the unit circle, then $\{\mathbf{x}_t\}$ is stationary;
(b) if all the eigenvalues are inside or on the unit circle, then $\{\mathbf{x}_t\}$ is non-stationary;
(c) if any of the eigenvalues are outside the unit circle, then $\{\mathbf{x}_t\}$ is explosive.

For the bivariate gasoline-price VAR(2) model, we have $2 \times 2 = 4$ roots, the moduli of which are:

$$0.93, 0.72, 0.72, 0.56.$$

Figure 4 illustrates these in relation to the unit circle.

We note that the system is stable (no explosive roots), that there is one near-unit root, suggesting the presence of a stochastic trend, as well as a pair of complex roots (whether a pair of roots is real or complex can depend on the third or smaller digits of the estimated coefficients, so is not usually a fundamental property of such a figure). Since there is only one root close to unity for the two variables, the series seem non-stationary and possibly cointegrated.

When there are unit roots in the model, it is convenient to reformulate the VAR into an equilibrium-correction model (EqCM). The next section discusses different ways of formulating such models.

4 Different representations of the VAR

The purpose of this section is to demonstrate that the unrestricted VAR can be given different parametrizations without imposing any binding restrictions on the parameters of the model (i.e., without

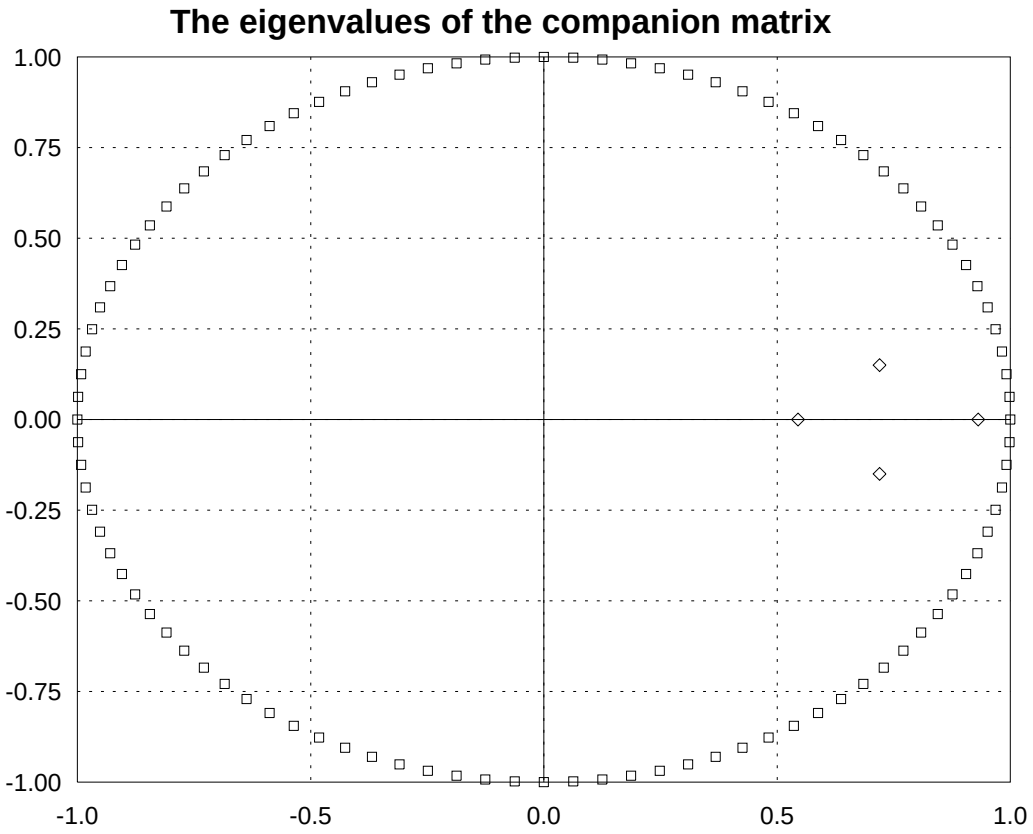

Figure 4 The roots of the characteristic polynomial for the VAR(2) model.

changing the value of the likelihood function). At this stage, we do not need to specify the order of integration of $\mathbf{x}_t$: as long as the parameters $(\boldsymbol{\pi}, \boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \boldsymbol{\Omega}_\epsilon)$ are unrestricted, OLS can be used to estimate them, as discussed in the previous section. Thus, any of the four parameterizations below, namely (4), (7), (8), or (9) can be used to obtain the first unrestricted estimates of the VAR. Although the parameters differ in the four representations, each of them explains exactly as much of the variation in $\mathbf{x}_t$.

The first reformulation of (4) is into the following equilibrium-correction form:

$$\Delta \mathbf{x}_t = \boldsymbol{\Phi}_1 \Delta \mathbf{x}_{t-1} - \boldsymbol{\Pi} \mathbf{x}_{t-1} + \boldsymbol{\pi} + \boldsymbol{\epsilon}_t, \tag{7}$$

where $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_p[\mathbf{0}, \boldsymbol{\Omega}_\epsilon]$, with the lagged levels matrix $\boldsymbol{\Pi} = \mathbf{I}_p - \boldsymbol{\Pi}_1 - \boldsymbol{\Pi}_2$ and $\boldsymbol{\Phi}_1 = -\boldsymbol{\Pi}_2$.[4] In (7), the lagged levels matrix, $\boldsymbol{\Pi}$, has been placed at time $t-1$, but could be chosen at any feasible lag without changing the likelihood. For example, placing the $\boldsymbol{\Pi}$ matrix at lag 2 yields the next parameterization:

$$\Delta \mathbf{x}_t = \boldsymbol{\Phi}_1^* \Delta \mathbf{x}_{t-1} - \boldsymbol{\Pi} \mathbf{x}_{t-2} + \boldsymbol{\pi} + \boldsymbol{\epsilon}_t \tag{8}$$

where $\boldsymbol{\Phi}_1^* = (\boldsymbol{\Pi}_1 - \mathbf{I}_p)$, with an unchanged $\boldsymbol{\Pi}$ matrix.[5] In a sense, (7) is more appropriate if one wants to discriminate between the short-run adjustment effects to the long-run relations (in levels), and the effects of changes in the lagged differences (the transitory effects). The estimated coefficients and their p-values can vary considerably between the two formulations (7) and (8), despite their being identical in terms of explanatory power, and likelihood (note that they have identical errors $\{\boldsymbol{\epsilon}_t\}$). Often, many

[4] In the general VAR(k) model, $\boldsymbol{\Phi}$ matrices cumulate the longer lag coefficients of the levels representation.

[5] In the general VAR(k), $\boldsymbol{\Phi}^*$ matrices cumulate the earlier lag coefficients.

more significant coefficients are obtained with (8) than with (7), illustrating the increased difficulty of interpreting coefficients in dynamic models relative to static regression models: many significant coefficients need not imply high explanatory power, but could result from the parameterization of the model.

The final convenient reformulation of the VAR model is into second-order differences (acceleration rates), changes, and levels:

$$\Delta^2 \mathbf{x}_t = \mathbf{\Phi} \Delta \mathbf{x}_{t-1} - \mathbf{\Pi} \mathbf{x}_{t-1} + \boldsymbol{\pi} + \boldsymbol{\epsilon}_t \tag{9}$$

where $\mathbf{\Phi} = \mathbf{\Phi}_1 - \mathbf{I}_p = -\mathbf{I}_p - \mathbf{\Pi}_2$, and $\mathbf{\Pi}$ remains as before. This formulation is most convenient when $\mathbf{x}_t$ contains $\mathsf{I}(2)$ variables, but from an economic point of view it also provides a natural decomposition of economic data which cover periods of rapid change, when acceleration rates (in addition to growth rates) become relevant, and more stable periods, when acceleration rates are zero, but growth rates still matter. It also clarifies that the 'ultimate' variable to be explained is $\Delta^2 \mathbf{x}_t$, which is often treated as a 'surprise', but as (9) demonstrates, can be explained by the determinants of the model. Indeed, it can be seen that treating $\Delta^2 \mathbf{x}_t$ purely as a 'surprise' imposes $\mathbf{\Phi} = \mathbf{0}$ and $\mathbf{\Pi} = \mathbf{0}$, and makes the differences behave as random walks.

Although the above reformulations are equivalent in terms of explanatory power, and can be estimated by OLS without considering the order of integration, inferences on some the parameters will not be standard unless $\mathbf{x}_t \sim \mathsf{I}(0)$. For example, when $\mathbf{x}_t$ is non-stationary, the joint significance of the estimated coefficients cannot be based on standard F-tests (see Hendry and Juselius, 2000, for a discussion in the context of single-equation models). We will now turn to the issue of non-stationarity in the VAR model.

5 Cointegration in the VAR

We first note that the general condition for $\mathbf{x}_t \sim \mathsf{I}(0)$ is that $\mathbf{\Pi}$ has full rank, so is non-singular. In this case, $|\mathbf{\Pi}(1)| = |\mathbf{\Pi}| \neq 0$, which corresponds to condition (a) in Section 3.4 that all the eigenvalues should lie within the unit circle. Stationarity can be seen as follows: stationary variables cannot grow systematically over time (that would violate the constant-mean requirement), so if $\mathbf{x}_t \sim \mathsf{I}(0)$ in (7), then $\mathsf{E}[\Delta \mathbf{x}_t] = \mathbf{0}$. Taking expectations yields:

$$-\mathbf{\Pi} \mathsf{E}[\mathbf{x}_{t-1}] + \boldsymbol{\pi} = \mathbf{0} \tag{10}$$

so when $\mathbf{\Pi}$ has full rank, $\mathsf{E}[\mathbf{x}_t] = \mathbf{\Pi}^{-1} \boldsymbol{\pi}$. Thus, the levels of stationary variables have a unique equilibrium mean – this is precisely why stationarity is so unreasonable for economic variables which are usually evolving! When $\mathbf{\Pi}$ is not full rank (i.e., when $\mathbf{x}_t$ exhibits $\mathsf{I}(1)$ behavior), (10) leaves some of the levels indeterminate. At the other extreme, where $\mathbf{\Pi} = \mathbf{0}$, the VAR becomes one in the differences $\Delta \mathbf{x}_t$, and these are stationary if $\mathbf{\Phi}_1 - \mathbf{I}_p$ has full rank, in which case $\mathbf{x}_t \sim \mathsf{I}(1)$. Notice that $\mathbf{\Phi}_1 = \mathbf{I}_p$ when $\mathbf{\Pi} = \mathbf{0}$ makes $\Delta \mathbf{x}_t$ a vector of random walks, so $\mathbf{x}_t \sim \mathsf{I}(2)$.

Section 5.1 presents the conditions for cointegration in the $\mathsf{I}(1)$ model as restrictions on the $\mathbf{\Pi}$ matrix and Section 5.2 discusses the properties of the vector process when the data are $\mathsf{I}(1)$ and cointegrated, based on the moving-average representation.

5.1 Determining cointegration in the VAR model

As discussed in Section 3.4, when some of the roots of the system (4) are on the unit circle (case (b)), the vector process $\mathbf{x}_t$ is non-stationary. However, some linear combinations, denoted $\boldsymbol{\beta}' \mathbf{x}_t$, might be

stationary even though the variables themselves are non-stationary. Then the variables are cointegrated from $\mathsf{I}(1)$, down one step to $\mathsf{I}(0)$, which Engle and Granger (1987) expressed as being $CI(1,1)$. There are two general conditions for $\mathbf{x}_t \sim \mathsf{I}(1)$, which we now discuss.

The first condition, needed to ensure that the data are not $\mathsf{I}(0)$, is that $\mathbf{\Pi}$ has reduced rank $r < p$, so can be written as:

$$\mathbf{\Pi} = -\boldsymbol{\alpha}\boldsymbol{\beta}' \tag{11}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $p \times r$ matrices, both of rank r. Substituting (11) into (9) delivers the cointegrated VAR model:

$$\Delta^2 \mathbf{x}_t = \mathbf{\Phi}\Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1}\right) + \boldsymbol{\pi} + \boldsymbol{\epsilon}_t. \tag{12}$$

An important feature of 'reduced rank' matrices like $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is that they have orthogonal complements, which we denote by $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$: i.e., $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are $p \times (p-r)$ matrices orthogonal to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (so $\boldsymbol{\alpha}'_\perp\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta}'_\perp\boldsymbol{\beta} = \mathbf{0}$), where the $p \times p$ matrices $(\boldsymbol{\alpha}\ \boldsymbol{\alpha}_\perp)$ and $(\boldsymbol{\beta}\ \boldsymbol{\beta}_\perp)$ both have full rank p. These orthogonal matrices play a crucial role in understanding the relationship between cointegration and 'common trends' as we explain below (a simple algorithm for constructing $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ from $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is given in Hendry and Doornik, 1996). Note that multiplying (12) by $\boldsymbol{\alpha}'_\perp$ will eliminate the cointegrating relations since $\boldsymbol{\alpha}'_\perp\boldsymbol{\alpha} = \mathbf{0}$.

The second condition, which is needed to ensure that the data are not $\mathsf{I}(2)$, is somewhat more technical, and requires that a transformation of $\mathbf{\Phi}$ in (12) must be of full rank.[6] Here, we will disregard the I(2) problem and only discuss the case when the footnoted condition is satisfied.

If $r = p$, then $\mathbf{x}_t$ is stationary, so standard inference (based on t, F, and χ^2) applies. If $r = 0$, then $\Delta\mathbf{x}_t$ is stationary, but it is not possible to obtain stationary relations between the levels of the variables by linear combinations. Such variables do not have any cointegration relations, and hence, cannot move together in the long run. In this case, each of (7)–(9) becomes a VAR model in differences but, since $\Delta\mathbf{x}_t \sim \mathsf{I}(0)$, standard inference still applies. If $p > r > 0$, then $\mathbf{x}_t \sim \mathsf{I}(1)$ and there exist r directions in which the process can be made stationary by linear combinations, $\boldsymbol{\beta}'\mathbf{x}_t$. These are the cointegrating relations, exemplified in (2) above by the two price differentials.

5.2 The VAR model in moving-average form

When the characteristic polynomial $\mathbf{\Pi}(z) = \mathbf{I}_p - \mathbf{\Pi}_1 z - \mathbf{\Pi}_2 z^2$ contains a unit root, the determinant $|\mathbf{\Pi}(z)| = 0$ for $z = 1$, so $\mathbf{\Pi}(z)$ cannot be inverted to express $\mathbf{x}_t$ as a moving average of current and past $\boldsymbol{\epsilon}_t$. Instead, we must decompose the characteristic polynomial into a unit-root part and a stationary invertible part, written as the product:

$$\mathbf{\Pi}(z) = (1-z)\mathbf{\Pi}^*(z),$$

where $\mathbf{\Pi}^*(z)$ has no unit roots, and is invertible. The VAR model can now be written as:

$$(1-L)\mathbf{x}_t = \Delta\mathbf{x}_t = \left[\mathbf{\Pi}^*(L)\right]^{-1}\boldsymbol{\epsilon}_t. \tag{14}$$

Thus, $\Delta\mathbf{x}_t$ is a moving average of current and past $\boldsymbol{\epsilon}_t$. To see the nature of that relation, expand $\left[\mathbf{\Pi}^*(L)\right]^{-1}$ as a power series in L:

$$\left[\mathbf{\Pi}^*(L)\right]^{-1} = \mathbf{C}_0 + \mathbf{C}_1 L + \mathbf{C}_2 L^2 + \ldots = \mathbf{C}(L) \text{ (say)}.$$

[6] Specifically:

$$\boldsymbol{\alpha}'_\perp\mathbf{\Phi}\boldsymbol{\beta}_\perp = \boldsymbol{\zeta}\boldsymbol{\eta}' \tag{13}$$

where $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are $(p-r) \times s$ matrices for $s = p - r$, which is the number of 'common stochastic trends' of first order when $\mathbf{x}_t \sim \mathsf{I}(1)$. Thus, (13) must have full rank ($s = p - r$) for $\mathbf{x}_t \sim \mathsf{I}(1)$. If $s < p - r$, then the model contains $p - r - s$ second-order stochastic trends, and $\mathbf{x}_t \sim \mathsf{I}(2)$.

In turn, express $\mathbf{C}(L)$ as:

$$\mathbf{C}(L) = \mathbf{C} + \mathbf{C}^*(L)(1-L),$$

so $\mathbf{C}(1) = \mathbf{C}$ as $\mathbf{C}^*(1)(1-1) = \mathbf{0}$. We can now rewrite (14) as:

$$\Delta \mathbf{x}_t = \left[\mathbf{C} + \mathbf{C}^*(L)(1-L)\right] \boldsymbol{\epsilon}_t.$$

By integration (dividing by the difference operator, $(1-L)$):

$$\mathbf{x}_t = \mathbf{C}\left(\frac{\boldsymbol{\epsilon}_t}{1-L}\right) + \mathbf{C}^*(L)\boldsymbol{\epsilon}_t + \mathbf{x}_0,$$

for some initial condition denoted $\mathbf{x}_0$, we can express $\mathbf{x}_t$ as:

$$\mathbf{x}_t = \mathbf{C}\sum_{i=1}^{t} \boldsymbol{\epsilon}_i + \mathbf{C}^*(L)\boldsymbol{\epsilon}_t + \mathbf{x}_0. \tag{15}$$

In (15), $\mathbf{x}_t$ is decomposed into a stochastic trend, $\mathbf{C}\sum_{i=1}^{t} \boldsymbol{\epsilon}_i$, and a stationary stochastic component, $\mathbf{e}_t = \mathbf{C}^*(L)\boldsymbol{\epsilon}_t$.[7] There are $(p-r)$ linear combinations between the cumulated residuals, $\boldsymbol{\alpha}'_\perp \sum_{i=1}^{t} \widehat{\boldsymbol{\epsilon}}_i$ which define the common stochastic trends that affect the variables $\mathbf{x}_t$ with weights $\mathbf{B}$, where $\mathbf{C} = \mathbf{B}\boldsymbol{\alpha}'_\perp$. In this sense, there exists a beautiful duality between cointegration and common trends. The following example illustrates.

Assume that there exists one common trend between the two gasoline price series, and hence one cointegration relation as reported in Hendry and Juselius (2000). Then, $r = 1$ and $p - r = 1$, and we can write the moving-average (common-trends) representation as:

$$\begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix} = \begin{pmatrix} b_{11} \\ b_{21} \end{pmatrix} \left(\sum_{i=1}^{t} u_i\right) + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} t + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}, \tag{17}$$

where $\mathbf{B}' = (b_{11}, b_{21})$ are the weights of the estimated common trend given by $\widehat{u}_i = \boldsymbol{\alpha}'_\perp \widehat{\boldsymbol{\epsilon}}_i$, and $\delta_1,\ \delta_2$ are the coefficients of linear deterministic trends in $p_{1,t},\ p_{2,t}$, respectively. Hendry and Juselius (2000) showed that $(p_{1,t} - p_{2,t}) \sim \mathsf{I}(0)$, i.e., that the two prices were cointegrated with $\boldsymbol{\beta}' = (1, -1)$. Furthermore, the assumption behind the single-equation model in Hendry and Juselius (2000) was that $\boldsymbol{\alpha}' = (*, 0)$. These estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ correspond to $\mathbf{B}' = (1, 1)$ and $\boldsymbol{\alpha}'_\perp = (0, 1)$ in (17). We express this outcome as: cumulated shocks to $p_{2,t}$ give an estimate of the common stochastic trend in this small system. Both prices are similarly affected by the stochastic trend, so that in:

$$(p_{1,t} - p_{2,t}) = (b_{11} - b_{21}) \sum_{i=1}^{t} u_i + (\delta_1 - \delta_2)t + (e_{1,t} - e_{2,t}) \tag{18}$$

we have $(b_{11} - b_{21}) = 0$, and the linear relation $(p_{1,t} - p_{2,t})$ has no stochastic trend left, thereby defining a cointegration relation.

Note that (17) also allows for a deterministic linear trend t in $\mathbf{x}_t$. If, in addition, $\delta_1 = \delta_2$, then both the stochastic and the linear trend will cancel in the linear relation $(p_{1,t} - p_{2,t})$ in (18). If $\delta_1 \neq \delta_2$, then we need to allow for a linear trend in the cointegration relation, such that $(p_{1,t} - p_{2,t} - d_1 t)$ contains neither stochastic nor deterministic trends. In this case, we say that the price differential is trend stationary, and that there is a trend in the cointegration space. We have reached the stage where we need a more complete discussion of the key role played by deterministic terms in cointegrated models.

[7] It can be shown that:

$$\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\Phi} \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp, \tag{16}$$

so the $\mathbf{C}$ matrix is directly related to (13), and can be calculated from estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Phi}$: see e.g., Johansen (1992a). Letting $\mathbf{B} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\Phi} \boldsymbol{\beta}_\perp)^{-1}$, then $\mathbf{C} = \mathbf{B}\boldsymbol{\alpha}'_\perp$, so the common stochastic trends have a reduced-rank representation similar to the stationary cointegration relations.

6 Deterministic components in a cointegrated VAR

A characteristic feature of the equilibrium-correction formulation (12) is the inclusion of both differences and levels in the same model, allowing us to investigate both short-run and long-run effects in the data. As discussed in Hendry and Juselius (2000), however, the interpretation of the coefficients in terms of dynamic effects is difficult. This is also true for the trend and the constant term,as well as other deterministic terms like dummy variables. The following treatment starts from the discussion of the dual role of the constant term and the trend in the dynamic regression model in Hendry and Juselius (2000), and extends the results to the cointegrated VAR model.

When two (or more) variables share the same stochastic and deterministic trends, it is possible to find a linear combination that cancels both the trends. The resulting cointegration relation is not trending, even if the variables by themselves are. In the cointegrated VAR model, this case can be accounted for by including a trend in the cointegration space. In other cases, a linear combination of variables removes the stochastic trend(s), but not the deterministic trend, so we again need to allow for a linear trend in the cointegration space. Similar arguments can be used for an intervention dummy: the intervention might have influenced several variables similarly, such that the intervention effect cancels in a linear combination of them, and no dummy is needed. Alternatively, if an intervention only affects a subset of the variables (or several, but asymmetrically), the effect will not disappear in the cointegration relation, so we need to include an intervention dummy.

These are only a few examples showing that the role of the deterministic and stochastic components in the cointegrated VAR is quite complicated. However, it is important to understand their role in the model, partly because one can obtain misleading (biased) parameter estimates if the deterministic components are incorrectly formulated, partly because the asymptotic distributions of the cointegration tests are not invariant to the specifications of these components. Furthermore, the properties of the resulting formulation may prove undesirable for (say) forecasting, by inadvertently retaining unwanted components – such as quadratic trends, as illustrated by Case 1 below. In general, parameter inference, policy simulations, and forecasting are much more sensitive to the specification of the deterministic than the stochastic components of the VAR model. Doornik, Hendry and Nielsen (1998) provide a comprehensive discussion.

6.1 Intercepts in cointegration relations and growth rates

Another important aspect is to decompose the intercept, $\boldsymbol{\pi}$, into components that induce growth in the system, and those that capture the means of the cointegration relations.[8] Reconsider the VAR representation (7):

$$\Delta \mathbf{x}_t = \boldsymbol{\Phi}_1 \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}' \mathbf{x}_{t-1}\right) + \boldsymbol{\pi} + \boldsymbol{\epsilon}_t \tag{19}$$

where $\Delta \mathbf{x}_t \sim \mathsf{I}(0)$ and $\boldsymbol{\epsilon}_t \sim \mathsf{I}(0)$. To 'balance' (19), $\boldsymbol{\beta}' \mathbf{x}_{t-1}$ must be $\mathsf{I}(0)$ also. Since the r cointegration relations $\boldsymbol{\beta}' \mathbf{x}_{t-1}$ are stationary, each of them has a constant mean**.** Similarly, $\Delta \mathbf{x}_t$ is stationary with a constant mean, which we denote by $\mathsf{E}[\Delta \mathbf{x}_t] = \boldsymbol{\gamma}$, describing a $(p \times 1)$ vector of growth rates. This was illustrated in (17) by allowing for linear trends in the two prices with slope coefficients δ_1 and δ_2. Furthermore, let $\mathsf{E}\left[\boldsymbol{\beta}' \mathbf{x}_{t-1}\right] = \boldsymbol{\mu}$ describe a $(r \times 1)$ vector of intercepts in the cointegrating relations. We now take expectations in (19):

$$\left(\mathbf{I}_p - \boldsymbol{\Phi}_1\right) \boldsymbol{\gamma} = \boldsymbol{\alpha} \mathsf{E}\left[\boldsymbol{\beta}' \mathbf{x}_{t-1}\right] + \boldsymbol{\pi} = \boldsymbol{\alpha} \boldsymbol{\mu} + \boldsymbol{\pi}.$$

[8] One of the reasons we assume all the variables are in logs is to avoid the growth rates depending on the levels of the variables.

Consequently, $\boldsymbol{\pi} = (\mathbf{I}_p - \boldsymbol{\Phi}_1)\boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu}$. Note that the constant term $\boldsymbol{\pi}$ in the VAR model does indeed consist of two components: one related to the linear growth rates in the data, and the other related to the mean values of the cointegrating relations (i.e., the intercepts in the long-run relations). This decomposition is similar to the simpler single-equation case discussed in Hendry and Juselius (2000). When the cointegration relations are trend free as in (19):

$$\boldsymbol{\beta}'\mathsf{E}\left[\Delta\mathbf{x}_t\right] = \mathsf{E}\left[\Delta\boldsymbol{\beta}'\mathbf{x}_t\right] = \boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0},$$

so we can express (19) in mean-deviation form as:

$$(\Delta\mathbf{x}_t - \boldsymbol{\gamma}) = \boldsymbol{\Phi}_1(\Delta\mathbf{x}_{t-1} - \boldsymbol{\gamma}) + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t. \tag{20}$$

There are two forms of equilibrium correction in (20): that of the growth $\Delta\mathbf{x}_t$ in the system to its mean $\boldsymbol{\gamma}$; and of the cointegrating vectors $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ to their mean $\boldsymbol{\mu}$. The two mean values, $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$, play an important role in the cointegrated VAR model, and it is important to ascertain whether they are significantly different from zero or not at the outset of the empirical analysis. In the next subsection, we will present five baseline cases describing how the trend and the intercept can enter the VAR specification.

6.2 Five cases for trends and intercepts

The basic ideas are illustrated using the p-dimensional cointegrated VAR with a constant and a linear trend, but to simplify notations we assume that only one lag is needed, so $\boldsymbol{\Phi}_1 = \mathbf{0}$. As before, $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_p\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]$:

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\pi} + \boldsymbol{\delta}t + \boldsymbol{\epsilon}_t. \tag{21}$$

Without loss of generality, the two $(p \times 1)$ vectors $\boldsymbol{\pi}$ and $\boldsymbol{\delta}$ can each be decomposed into two new vectors, of which one is related to the mean value of the cointegrating relations, $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ (case (3) in section 2 of Hendry and Juselius, 2000), and the other to growth rates in $\Delta\mathbf{x}_t$:

$$\begin{array}{rcl} \boldsymbol{\pi} & = & \boldsymbol{\alpha}\boldsymbol{\mu} + \boldsymbol{\gamma}_0 \\ \boldsymbol{\delta} & = & \boldsymbol{\alpha}\boldsymbol{\rho} + \boldsymbol{\tau} \end{array} \tag{22}$$

Substituting (22) into (21) yields:

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\mu} + \boldsymbol{\gamma} + \boldsymbol{\alpha}\boldsymbol{\rho}t + \boldsymbol{\tau}t + \boldsymbol{\epsilon}_t, \tag{23}$$

so, collecting terms in (23):

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\left(\boldsymbol{\beta}' : \boldsymbol{\mu} : \boldsymbol{\rho}\right)\begin{pmatrix} \mathbf{x}_{t-1} \\ 1 \\ t \end{pmatrix} + (\boldsymbol{\gamma} + \boldsymbol{\tau}t) + \boldsymbol{\epsilon}_t. \tag{24}$$

Thus, we can rewrite (21) as:

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\mu}' \\ \boldsymbol{\rho}' \end{pmatrix}'\mathbf{x}^*_{t-1} + (\boldsymbol{\gamma} + \boldsymbol{\tau}t) + \boldsymbol{\epsilon}_t, \tag{25}$$

where $\mathbf{x}^*_{t-1} = (\mathbf{x}'_{t-1}, 1, t)'$. We can always choose $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ such that the equilibrium error $(\boldsymbol{\beta}^*)'\mathbf{x}^*_t = \mathbf{v}_t$ has mean zero (where $\boldsymbol{\beta}^* = \left(\boldsymbol{\beta}', \boldsymbol{\mu}, \boldsymbol{\rho}\right)'$), so the trend component in (25) can be interpreted from the equation:

$$\mathsf{E}\left[\Delta\mathbf{x}_t\right] = \boldsymbol{\gamma} + \boldsymbol{\tau}t. \tag{26}$$

Thus, $\boldsymbol{\gamma} \neq \mathbf{0}$ corresponds to constant growth in the variables $\mathbf{x}_t$ (case 1 in section 2 of Hendry and Juselius, 2000), whereas $\boldsymbol{\tau} \neq \mathbf{0}$ corresponds to linear trends in growth, and so quadratic trends in the variables. Hence, the constant term and the deterministic linear trend play a dual role in the cointegrated model: in the $\boldsymbol{\alpha}$ directions they describe a linear trend and an intercept in the steady-state relations; in the remaining directions, they describe quadratic and linear trends in the data. To correctly interpret the model, one has to understand the distinction between the part of the deterministic component that 'belongs' to the cointegration relations, and the part that 'belongs' to the differences.

In empirical work, usually one has some idea whether there are linear deterministic trends in some (or all) of the variables. It might, however, be more difficult to know if they cancel in the cointegrating relations or not. Luckily, we do not need to know beforehand, because the econometric analysis can be used to find out. As discussed below, all these cases can be expressed as linear restrictions on the deterministic components of the VAR model and, hence, can be tested. We now discuss five of the most frequently used models arising from restricting the deterministic components in (21): see Johansen (1994).

Case 1. No restrictions on $\boldsymbol{\pi}$ and $\boldsymbol{\delta}$, so the trend and intercept are *unrestricted* in the VAR model. With unrestricted parameters, $\boldsymbol{\pi},\ \boldsymbol{\delta}$, the model is consistent with linear trends in the differenced series $\Delta\mathbf{x}_t$ as shown in (26) and, thus, quadratic trends in $\mathbf{x}_t$. Although quadratic trends may sometimes improve the fit within the sample, forecasting outside the sample is likely to produce implausible results. Be careful with this option: it is preferable to find out what induced the apparent quadratic growth and, if possible, increase the information set of the model. Moreover, as shown in Doornik *et al.* (1998), estimation and inference can be unreliable.

Case 2. $\boldsymbol{\tau} = \mathbf{0}$, but $\boldsymbol{\gamma},\ \boldsymbol{\mu},\ \boldsymbol{\rho}$ remain unrestricted, so the trend is *restricted* to lie in the cointegration space, but the constant is unrestricted in the model. Thus, $\boldsymbol{\tau}$ being zero in (25) still allows linear, but precludes quadratic, trends in the data. As illustrated in the previous section, $\mathsf{E}[\Delta\mathbf{x}_t] = \boldsymbol{\gamma} \neq \mathbf{0}$ implies linear deterministic trends in the level $\mathbf{x}_t$. When, in addition, $\boldsymbol{\rho} \neq \mathbf{0}$, these linear trends in the variables do not cancel in the cointegrating relations, so the model contains 'trend-stationary' relations which can either describe a single trend-stationary variable, $(x_{1,t} - b_1 t) \sim \mathsf{I}(0)$, or an equilibrium relation $(\boldsymbol{\beta}_1'\mathbf{x}_t - b_2 t) \sim \mathsf{I}(0)$. Therefore, the hypothesis that a variable is trend-stationary can be tested in this model.

Case 3. $\boldsymbol{\delta} = \mathbf{0}$, so there are no linear trends in (21). Since the constant term $\boldsymbol{\pi}$ is *unrestricted*, there are still linear trends in the data, but no deterministic trends in any cointegration relations. Also, $\mathsf{E}[\Delta\mathbf{x}_t] = \boldsymbol{\gamma} \neq \mathbf{0}$, is consistent with linear deterministic trends in the variables but, since $\boldsymbol{\rho} = \mathbf{0}$, these trends cancel in the cointegrating relations. It appears from (24) that $\boldsymbol{\pi} \neq \mathbf{0}$ accounts for both linear trends in the DGP and a non-zero intercept in the cointegration relations.

Case 4. $\boldsymbol{\delta} = \mathbf{0},\ \boldsymbol{\gamma} = \mathbf{0}$, but $\boldsymbol{\mu} \neq \mathbf{0}$,so the constant term is *restricted* to lie in the cointegration space in (25). In this case, there are no linear deterministic trends in the data, consistent with $\mathsf{E}[\Delta\mathbf{x}_t] = \mathbf{0}$. The only deterministic components in the model are the intercepts in any cointegrating relations, implying that some equilibrium means are different from zero.

Case 5. $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\pi} = \mathbf{0}$, so the model excludes all deterministic components in the data, with both $\mathsf{E}[\Delta\mathbf{x}_t] = \mathbf{0}$ and $\mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_t] = \mathbf{0}$, implying no growth and zero intercepts in every cointegrating relation. Since an intercept is generally needed to account for the initial level of measurements, $\mathbf{x}_0$, only in the exceptional case when the measurements start from zero, or when the measurements cancel in the cointegrating relations, can the restriction $\boldsymbol{\pi} = \mathbf{0}$ be justified.

Turning to our empirical example, Table 1 showed that $\mathsf{E}[\Delta p_{i,t}] = 0$ could not be rejected. Hence, there is no evidence of linear deterministic trends in the gasoline prices, at least not over the sample

period. The graphs in Figure 1 support this conclusion. We conclude that the cointegrated VAR model should be formulated according to case 4 here, with the constant term restricted to the cointegration space, and no deterministic trend terms.

7 The likelihood-based procedure

So far, we have discussed the formulation of the VAR model in terms of well-specified stochastic and deterministic properties. All this can be done before addressing the unit-root problem. As in Hendry and Juselius (2000), we will now assume that some of the roots of the characteristic polynomial are on the unit circle. This means that $\mathbf{x}_t \sim \mathsf{I}(1)$, and we consider the cointegrated VAR(2) model (7) where $\mathbf{\Pi} = -\boldsymbol{\alpha}\boldsymbol{\beta}'$ with a constant, a linear trend and a vector of dummy variables $\mathbf{d}_t$ included:

$$\Delta\mathbf{x}_t = \mathbf{\Phi}_1\Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\pi} + \boldsymbol{\delta}t + \mathbf{\Psi}\mathbf{d}_t + \boldsymbol{\epsilon}_t. \tag{27}$$

Since $\Delta\mathbf{x}_t \sim \mathsf{I}(0)$ and $\boldsymbol{\epsilon}_t \sim \mathsf{I}(0)$, all stochastic components in (27) are stationary by definition except for $\boldsymbol{\beta}'\mathbf{x}_{t-1}$. For (27) to be internally consistent, given that $\mathbf{x}_t \sim \mathsf{I}(1)$, $\boldsymbol{\beta}$ cannot be a full-rank $p \times p$ matrix (because then something stationary would be equal to something non-stationary). The only possible solution is that $\boldsymbol{\beta}$ is a reduced rank $(p \times r)$ matrix with $r < p$, so r linear combinations cancel stochastic trends as shown in section 5. Below, we will only discuss the broad ideas of the maximum likelihood estimation procedure, and will not go through the derivations of the results. The interested reader is referred to Johansen (1988, 1995) and Banerjee *et al.* (1993), *inter alia*, for details.

The derivation of the maximum likelihood estimator (MLE) is done via the 'concentrated likelihood' of the VAR model. Since the latter is crucial for understanding both the statistical and economic properties of the VAR, we will demonstrate how it is defined. We use the following shorthand notation:

$$\begin{aligned} \mathbf{z}_{0,t} &= \Delta\mathbf{x}_t \\ \mathbf{z}_{1,t} &= \mathbf{x}_{t-1} \\ \mathbf{z}_{2,t} &= (\Delta\mathbf{x}_{t-1}, \mathbf{1}, \mathbf{t}, \mathbf{d}_t)\,. \end{aligned}$$

(additional lagged differences are easily included in $\mathbf{z}_{2,t}$). Rewrite (27) as:

$$\mathbf{z}_{0,t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{z}_{1,t-1} + \mathbf{\Theta}\mathbf{z}_{2,t} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{\Theta} = (\mathbf{\Phi}_1, \boldsymbol{\pi}, \boldsymbol{\delta}, \mathbf{\Psi})$. By concentrating out the short-run dynamic adjustment effects, $\mathbf{\Theta}\mathbf{z}_{2,t}$, we are able to obtain a 'simpler' model. This is done by first defining the following auxiliary OLS regressions:

$$\begin{aligned} \mathbf{z}_{0,t} &= \widehat{\mathbf{D}}_1'\mathbf{z}_{2,t} + \Delta\widetilde{\mathbf{x}}_t \\ \mathbf{z}_{1,t} &= \widehat{\mathbf{D}}_2'\mathbf{z}_{2,t} + \widetilde{\mathbf{x}}_{t-1}, \end{aligned}$$

where $\Delta\widetilde{\mathbf{x}}_t = \mathbf{z}_{0,t} - \mathbf{M}_{02}\mathbf{M}_{22}^{-1}\mathbf{z}_{2,t}$ and $\widetilde{\mathbf{x}}_{t-1} = \mathbf{z}_{1,t} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{z}_{2,t}$ are the OLS residuals, and $\mathbf{M}_{ij} = \sum_t(\mathbf{z}_{i,t}\mathbf{z}_{j,t}')/T$ is a product-moment matrix, so that $\widehat{\mathbf{D}}_1' = \mathbf{M}_{02}\mathbf{M}_{22}^{-1}$ and $\widehat{\mathbf{D}}_2' = \mathbf{M}_{12}\mathbf{M}_{22}^{-1}$.

The concentrated model can now be written as:

$$\Delta\widetilde{\mathbf{x}}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1} + \mathbf{u}_t, \tag{28}$$

so we have transformed the original VAR containing short-run adjustments and intervention effects into the 'baby model' form, in which the adjustments are exclusively towards the long-run steady-state relations.

The MLE is close to limited-information maximum likelihood (LIML: see Hendry, 1976, for a consolidation) in that the key issue is to handle a reduced-rank problem, which essentially amounts to

solving an eigenvalue problem. In practice, the estimators are derived in two steps. First, to derive an estimator of $\boldsymbol{\alpha}$, assume that $\boldsymbol{\beta}$ is known: then $\boldsymbol{\beta}'\widetilde{\mathbf{x}}_{t-1}$ becomes a known variable in (28), so $\boldsymbol{\alpha}$ can be estimated by OLS. Next, insert that $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ in the expression for the concentrated likelihood function, which becomes a function of $\boldsymbol{\beta}$ alone, and no longer depends on $\boldsymbol{\alpha}$. To find the value of $\widehat{\boldsymbol{\beta}}$ that maximizes this likelihood function is a non-linear problem, but one that can be solved by reduced-rank regression (see Johansen, 1988). The solution delivers p eigenvalues λ_i where $0 \leq \lambda_i \leq 1$:

$$\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \ldots, \lambda_p),$$

which are ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. The estimate of $\boldsymbol{\beta}$ for r cointegrating vectors is given by the $p \times r$ matrix of eigenvectors corresponding to the largest r eigenvalues (the selection of r is discussed in Section 8). Given the MLE $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, calculate $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\widehat{\boldsymbol{\beta}})$. The estimates of the two eigenvectors, and the corresponding $\widehat{\boldsymbol{\alpha}}$ weights for the empirical example are reported in Table 3.

Each λ_i can be interpreted as the squared canonical correlation between linear combinations of the levels, $\boldsymbol{\beta}_i'\widetilde{\mathbf{x}}_{t-1}$, and a linear combination of the differences, $\boldsymbol{\varphi}_i'\Delta\widetilde{\mathbf{x}}_t$. In this sense, the magnitude of λ_i is an indication of how strongly the linear combination $\boldsymbol{\beta}_i'\widetilde{\mathbf{x}}_{t-1}$ is correlated with the stationary part of the process $\boldsymbol{\varphi}_i'\Delta\widetilde{\mathbf{x}}_t$. If $\lambda_i \approx 0$, the linear combination $\boldsymbol{\beta}_i'\widetilde{\mathbf{x}}_{t-1}$ is not at all correlated with the stationary part of the process, and hence is non-stationary.

In the situation where both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unrestricted (beyond normalizations), standard errors cannot be obtained, but the imposition of rank and other identifying restrictions usually allows appropriate standard errors to be obtained for $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$.

8 Testing cointegration rank

Given that the unrestricted VAR model has been found to satisfactorily describe the data (is a congruent representation), one can start the simplification search, which means imposing valid restrictions on the model such as a reduced-rank restriction, restrictions on the long-run parameters $\boldsymbol{\beta}$, and finally restrictions on the short-run adjustment parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Phi}$. The first, and most crucial, step is to discriminate empirically between zero and non-zero eigenvalues when allowing for sample variation, and then impose an appropriate cointegration rank restriction r on the $\boldsymbol{\Pi}$ matrix. Note that:

- if we underestimate r, then empirically-relevant equilibrium-correction mechanisms (EqCMs) will be omitted;
- whereas if we overestimate r, the distributions of some statistics will be non-standard, so that incorrect inferences may result from using conventional critical test values (based on t, F, χ^2);
- forecasts will be less accurate due to incorrectly retaining $\mathsf{I}(1)$ components, which will increase forecast variances.

A test for r cointegrating vectors can be based on the maximum likelihood approach proposed by Johansen (1988). The statistical problem is to derive a test procedure to discriminate between the λ_i, $i = 1, \ldots, r$, which are large enough to correspond to stationary $\boldsymbol{\beta}_i'\widetilde{\mathbf{x}}_{t-1}$, and those λ_i, $i = r+1, \ldots, p$, which are small enough to correspond to non-stationary eigenvectors. The rank r is determined by a likelihood-ratio test procedure between the two hypotheses:

$\mathcal{H}_p$: rank $= p$, i.e., full rank, so $\mathbf{x}_t$ is stationary;

$\mathcal{H}_r$: rank $= r < p$, i.e., r cointegration relations.

The test is:

$$LR(\mathcal{H}_r \mid \mathcal{H}_p) = -T \ln\left[(1-\lambda_{r+1}) \cdots (1-\lambda_p)\right] = -T \sum_{i=r+1}^{p} \ln(1-\lambda_i).$$

If $\lambda_{r+1} = \cdots = \lambda_p = 0$, the test statistic should be small (close to zero), which delivers the critical value under the null. The test is based on non-standard asymptotic distributions that have been simulated for the five cases discussed in section 6. There is an additional problem, in that $\mathcal{H}_r$ may be correctly accepted when $\lambda_r = 0$, or even $\lambda_{r-1} = 0$. Therefore, if $\mathcal{H}_r$ is accepted, we conclude that there are at least $p - r$ unit roots, i.e., $p - r$ 'common trends' in the process (but there can be more) corresponding to at most r stationary relations.

However, if $LR(\mathcal{H}_{r-1}|\mathcal{H}_p)$ is calculated, the test statistic includes $\ln(1 - \lambda_r)$, which will not be close to zero, so an outcome in excess of the critical value should be obtained, correctly rejecting the false null of fewer that r cointegration relations.

As discussed above, the asymptotic distributions depend on whether there is a constant and/or a trend; and whether these are unrestricted or not in the model. However other deterministic components, such as intervention dummies, are also likely to influence the shape of the test distributions. In particular, care should be taken when a deterministic component generates trending behavior in the levels of the data such as an unrestricted shift dummy ($\cdots$,0,0,0,1,1,1,$\cdots$): an explanation of the procedure is provided in Johansen, Mosconi and Nielsen (2000), and Juselius (2000): Doornik *et al.* (1998) also consider the estimation and inference problems resulting from including dummies.

Because the asymptotic distributions for the rank test depend on the deterministic components in the model and on whether these are restricted or unrestricted, the rank and the specification of the deterministic components have to be determined jointly. Nielsen and Rahbek (2000) have demonstrated that a test procedure based on a model formulation that allows a deterministic component, for example a deterministic trend t, to be restricted to the cointegration relations and the differenced component, $\Delta t = 1$, to be unrestricted in the model induces similarity in the test procedure (i.e., the critical values do not depend on the parameter values, so can be tabulated). This is because when there are linear trends in the data, i.e. $\mathsf{E}[\Delta \mathbf{x}_t] \neq \mathbf{0}$, they can enter the model through the constant term, $\boldsymbol{\gamma} \neq \mathbf{0}$ in (26) or through the cointegration relations, $\boldsymbol{\rho} \neq \mathbf{0}$ in (25). Hence, given linear trends in the data, case 2 is the most general case. When the rank has been determined, it is always possible to test the hypothesis $\boldsymbol{\rho} = \mathbf{0}$, as a linear restriction on the cointegrating relations.

If, on the other hand, $\mathsf{E}[\Delta \mathbf{x}_t] = \mathbf{0}$, so there are no linear trends in the data, then the baseline model has the constant term restricted to the cointegration space, which is case 4 above. Therefore, based on the similarity argument, the rank should be based on either case 4 (trends in the data) or case 2 (no trends in the data). Nevertheless, if there is strong prior information that there are trends in the data, but they do not appear in the cointegration relations, then case 3 is the appropriate choice.

9 Empirical model specification

The rank test is defined for a correctly specified model. Prior to the determination of the cointegration rank we should make sure that the empirical model is well-behaved. In Section 9.1, we choose the trend and constant in the baseline model, and account for the extraordinary events in the sample period; whereas in Section 9.2, we discuss the difficult choice of cointegration rank.

9.1 Model specification

For the gasoline example, we found in Table 1 that $\mathsf{E}[\Delta \mathbf{x}_t] = \mathbf{0}$ cannot be rejected. Hence, there is little evidence of linear deterministic trends in the data, at least not over the sample period, so we should determine the rank based on a case-4 model. But before testing the rank of the $\boldsymbol{\Pi}$ matrix, we need to

Table 2 The estimates of the short-run effects .

	$\widehat{\Phi}$		The Kuwait war effects		$\widehat{\mathbf{\Omega}}_{\varepsilon}$		
	$\Delta p_{1,t-1}$	$\Delta p_{2,t-1}$	$\Delta Ds90.31_t$	$\Delta Ds90.31_{t-1}$		$\epsilon_{1,t}$	$\epsilon_{2,t}$
$\Delta p_{1,t}$	$\mathbf{0.52}$ (11.3)	$\mathbf{0.14}$ (2.3)	0.18 (7.8)	−0.09 (4.0)	$\epsilon_{1,t}$	1.0	
$\Delta p_{2,t}$	0.05 (1.3)	$\mathbf{0.53}$ (12.0)	0.11 (5.8)	−0.02 (1.2)	$\epsilon_{2,t}$	0.66	1.0
The estimates of the extraordinary price changes ($\lvert\widehat{\epsilon}_{i,t}\rvert > 3.3$)							
	$Di89.13$	$Di89.39$	$Di89.51$	$Di90.49$	$Di9103$	$Di93.43$	$Dti98.11$
$\Delta p_{1,t}$	0.08 (3.6)	−0.08 (3.9)	0.09 (4.1)	−0.08 (3.5)	−0.09 (3.9)	−0.05 (2.5)	−0.08 (4.7)
$\Delta p_{2,t}$	0.07 (3.7)	−0.04 (2.0)	0.05 (2.9)	−0.07 (3.6)	−0.08 (4.2)	−0.08 (4.6)	−0.03 (2.5)

account for the effect of the Kuwait war on gasoline prices as discussed in Section 3.2, plus possibly some of the extraordinary price changes in this period that violate the normality assumption.

The question is now which kind of intervention dummies should be used to describe these extraordinary events. Juselius (2000) shows that a shift in the level of the variables corresponds to a permanent impulse blip in the differences, and an impulse effect in the levels corresponds to a transitory blip in the differences. The Kuwait war could have caused either a permanent or a transitory shift in the level of gasoline prices. In the latter case, we would see an initial rise in the gasoline prices in 1990:31 followed later on by a return to the previous price level. Based on the graphs in Figure 1, the differences show a large positive spike followed by another more moderately sized negative spike a few weeks later. This suggests that the initial increase in prices was partly permanent, partly transitory. To account for this extraordinary event on the price levels, the model includes a step dummy, $Ds90.31_t$ being zero up to $t = 1990$:31 and unity after that, *restricted* to lie in the cointegration space (if significant, then the permanent price increase in crude oil as a result of the war had a different effect on gasoline prices in the two locations). To account for the effect on the price changes, we also need to include current and lagged values of the impulse dummy $Di90.31_t$, defined below, as *unrestricted* in the model. Table 2 shows that the estimated direct price effect at the outbreak of the Kuwait war was an increase of 18% in location 1 and 11% in location 2, followed by a drop in the price by 9% in location 1 and 2% in location 2. Hence, the permanent effect of the war seemed to be approximately 9% in both locations. Altogether, the immediate price reaction in location 1 seemed to be stronger than in location 2.

Based on the criterion $|\widehat{\epsilon}_{i,t}| > 3.3$, we detected seven additional 'outlier' observations, accounted for by dummy variables defined as follows: the impulse dummy $Dixx.yy_t$ is unity for $t = 19xx$:yy, and zero otherwise; the transitory impulse dummy $Dtixx.yy_t$ is unity for $t = 19xx$:yy, -1 for $t = 19xx$:$yy + 1$, and zero otherwise. Table2 reports the estimates of these very large price changes. Again, we note that the price reactions in location 1 seem stronger than in location 2: except for 1993:43, gasoline prices in location 1 change more dramatically than in location 2.

The estimates of the Φ matrix in Table2 demonstrate quite strong autoregressive price behavior: a price change one week tends to be followed by a similar change but only half the size, next week. It also appears that a price change by $p_{2,t}$ is followed by a lagged change in $p_{1,t}$, but not the other way around. The estimates of the residual covariance matrix $\mathbf{\Omega}_{\varepsilon}$ show a large positive correlation between price shocks to the two gasoline prices, which suggests that there may be current, as well as lagged, price effects. This will be further discussed in Section 11.

After having accounted for these extraordinary events, the distributions of the residuals became much closer to a normal distribution than in the first tentatively estimated model. However, the em-

pirical model still showed some evidence of excess kurtosis and ARCH (autoregressive conditional heteroscedasticity, so the squared residuals are serially correlated: see Engle, 1982), as appears from Table 3, but neither of these usually causes serious problems for the properties of the estimates (see Gonzalo, 1994). A plausible explanation for these problems is that the residual variance of the process changes in the middle of the sample, at around 1992:40. We have analyzed the data separately for the split sample, and the ARCH and excess kurtosis disappear. However, the basic results remained unchanged in the two sub-samples and there is, therefore, no obvious need to report the results separately. We conclude that the empirical model is reasonably well specified, and turn to the determination of the cointegration rank.

9.2 Rank determination

As already mentioned, a correct choice of the cointegration rank is crucial for the analysis, but in practice, is far from easy. In many economic applications, the size of the sample is often quite small, and the tabulated asymptotic distributions can be rather poor approximations as has been demonstrated in many papers. See, for instance, Johansen (2000) for analytical results, and Jørgensen (2000) for Monte Carlo results. Another reason for concern is that when using correct small-sample distributions for the trace test, the size of the test is correct, but the power can be low, sometimes even of the same magnitude as the size. In such cases, a 5% test procedure will reject a unit root incorrectly 5% of the time, but accept a unit root incorrectly 95% of the time!

Thus, unless a unit root is given a structural interpretation (and hence, should be tested rigorously), it is important to make the decision based on as much information as possible, including prior economic information, and sensitivity analyses of doubtful cases to find out if important information is lost by leaving out the $r^{th}+1$ cointegration vector, or if anything is gained by including it.

The following information is often useful when deciding on the choice of cointegration rank:

(1) the trace test for cointegration rank;
(2) the characteristic roots of the model: if the $r^{th}+1$ cointegration vector is non-stationary and is wrongly included in the model, then the largest characteristic root will be close to the unit circle;
(3) the t-values of the α-coefficients for the $r^{th}+1$ cointegration vector; if these are all small, say less than 3.0, then one would not gain greatly by including that vector as a cointegrating relation in the model;
(4) the recursive graphs of the trace statistic for $\widetilde{r}=1,2,\ldots,p$: since the variable $T_j \ln(1-\lambda_i)$, for $j=T_1,\ldots,T$, grows linearly over time when $\lambda_i \neq 0$, the recursively-calculated components of the trace statistic should increase linearly for the first r components, but stay constant for the remainder;
(5) the graphs of the cointegrating relations: if the graphs reveal distinctly non-stationary behavior of a cointegration relation, which is supposedly stationary, one should reconsider the choice of r, or find out if the model specification is in fact incorrect, for example, if the data are I(2) instead of $\mathsf{I}(1)$;
(6) the economic interpretability of the results.

We will consider all the above pieces of information in turn:

The trace test (item 1) in the bivariate gasoline price case, should be able to discriminate between the following alternatives: no unit roots, one unit root, or two unit roots. The first case corresponds to prices being stationary, the second to them being $\mathsf{I}(1)$ with one stationary cointegration relation, and the last case to them being $\mathsf{I}(1)$ but with no cointegrating relation between them. The two trace test statistics in Table 3 are larger than their 95% quantiles, so that both λ_1 and λ_2 have to be considered different

Table 3 Rank and specification tests .

Rank determination			**The two largest roots**			**Misspecification tests**		
λ_i	$Trace$	Q_{95}	$(r=2)$	$(r=1)$		ARCH	Norm.	R^2
0.13	91.1	20.0	0.93	1.00	Δp_1	9.8	17.8	0.54
0.02	14.4	9.1	0.72	0.72	Δp_2	8.3	20.3	0.48

The unrestricted cointegration vectors and their weights

	p_1	p_2	$Ds90.31$	*constant*			$\widehat{\boldsymbol{\beta}}_1'\mathbf{x}_t$	$\widehat{\boldsymbol{\beta}}_2'\mathbf{x}_t$
$\widehat{\boldsymbol{\beta}}_1'$	1.0	-0.97	-0.02	0.02	$\Delta p_1:$	$\hat{\boldsymbol{\alpha}}_{1.}$	-0.15 (7.5)	-0.02 (2.1)
$\widehat{\boldsymbol{\beta}}_2'$	-0.25	1.0	0.03	0.41	$\Delta p_2:$	$\hat{\boldsymbol{\alpha}}_{2.}$	-0.02 (1.2)	-0.03 (3.8)

from zero, suggesting that both price series are stationary. As mentioned above λ_i can be interpreted as a squared canonical correlation coefficient. It appears from the table that both λ_1 and λ_2 are small in absolute value, indicating a fairly low correlation with the stationary part, and that λ_1 is much larger than λ_2, suggesting that the adjustment to the first cointegration relation is much stronger than to the second. The graphs of the cointegration vectors (item 5) in Figure 5 indicate that both are mean-reverting, but the second one much more strongly so. The two largest characteristic roots (item 2) show that $r = 2$ implies a fairly large root, 0.93, in the model. The recursive graphs (item 4) of $T_j \ln(1 - \lambda_{1_j})$ and $T_j \ln(1 - \lambda_{2_j})$ in Figure 6 show that the second component remains almost constant, whereas the first grows linearly over time. All this suggests that the first cointegrating relation is stationary and that the second is near integrated, but with significant mean reversion. A similar result was found in Hendry and Juselius (2000).

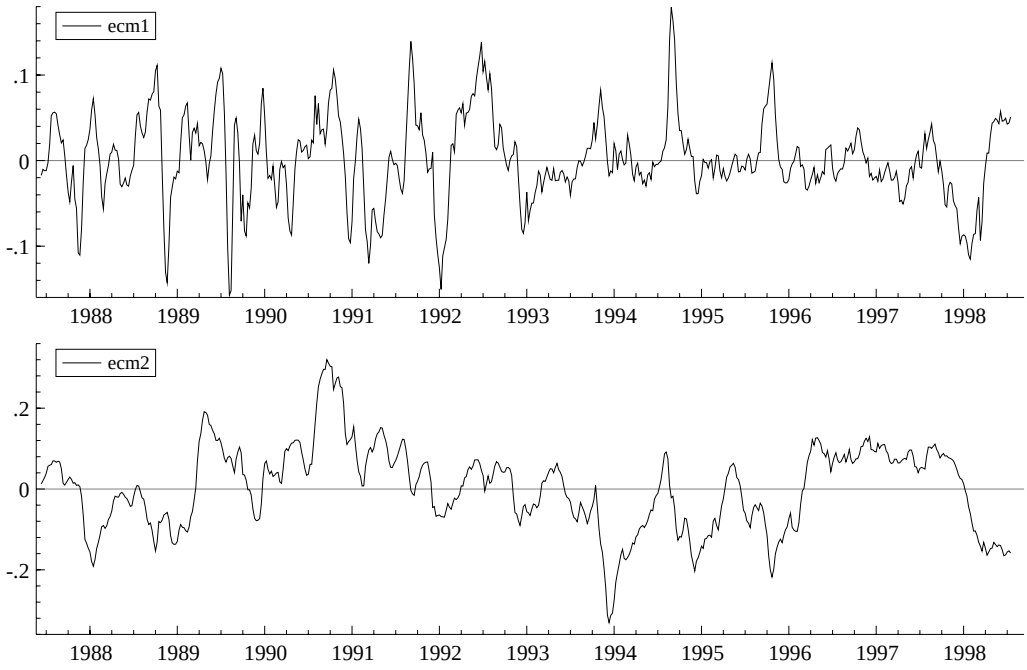

Figure 5 Graphs of the unrestricted cointegration relations $\beta_1' x_t$ and $\beta_2' x_t$.

Why does the trace test find that the small value of $\lambda_2 = 0.02$ is still significantly different from

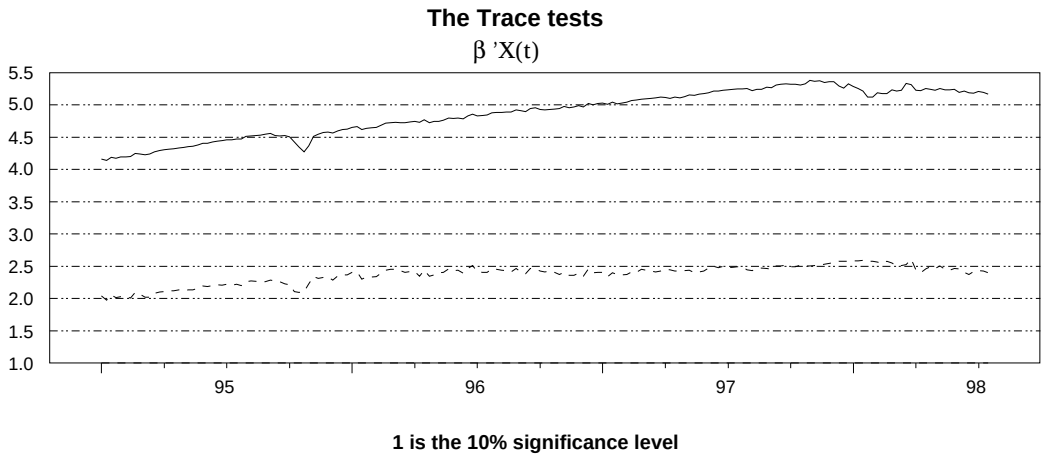

Figure 6 Recursively calculated trace tests.

zero? The simple explanation is that the sample size is very large here, 576 observations. Because the trace test is calculated as $T\ln(1-\boldsymbol{\lambda})$, even a small deviation from zero can be found to be significant when T is large enough. However, inference is much closer to the so called Dicky–Fuller distributions than to standard t-, F-, and χ^2-distributions when there is a near unit root in the model. Hence, to make inference more robust, it is often a good idea to approximate a near unit root by a unit root even when it is found to be statistically different from one.

Before finally deciding about the rank, we first check the economic interpretability of the second cointegration relation to see if it contains valuable information for the analysis. It appears from Table 3 that the adjustment coefficients to the second relation α_{21} and α_{22} (item 3) are very small, α_{21} is hardly significant in the first gasoline price equation, whereas α_{22} is more significant in the second price equation. Nevertheless, the coefficients of $\widehat{\boldsymbol{\beta}}_2'\mathbf{x}_t$ (item 6) indicate that it is essentially a unit vector describing price 2. This is confirmed by the graphs of the cointegration vectors (item 5) in Figure 5, where the second graph closely resembles $p_{2,t}$ in Figure 2. Therefore, accepting $r = 2$ is equivalent to saying that the gasoline prices are stationary, albeit subject to substantial persistence. As already discussed in Hendry and Juselius (2000), this might be the case, but choosing $r = 2$ would leave a near unit root in the model, and conventional (t-, F-, χ^2) inference is likely to be misleading. Moreover, asserting a constant long-run mean for nominal gasoline prices does not seem plausible. Finally, since $\boldsymbol{\beta}_2'\mathbf{x}_t$ will not add much valuable information about the co-movements of the two gasoline prices, we conclude that the empirical analysis will not benefit from choosing $r = 2$. We therefore continue the empirical analysis assuming $r = 1$.

10 Identification and hypothesis testing

Given the choice of the number of cointegrating relations, r, the Johansen procedure gives the maximum likelihood estimates of the unrestricted cointegrating relations $\boldsymbol{\beta}'\mathbf{x}_t$. Although the unrestricted $\boldsymbol{\beta}$ is uniquely determined based on the chosen normalization, the latter is not necessarily meaningful from an economic point of view. Therefore, an important part of a long-run cointegration analysis is to impose (over-) identifying restrictions on $\boldsymbol{\beta}$ to achieve economic interpretability. In section 10.1, we discuss a typology of restrictions on $\boldsymbol{\beta}$ and note the problem of calculating the degrees of freedom in the case of over-identifying restrictions. Section 10.2 provides some examples of how to specify hypotheses in a

testable form.

10.1 Restrictions on $\boldsymbol{\beta}$

As an example of just-identifying restrictions, consider the following design matrix $\mathbf{Q} = (\boldsymbol{\beta}_1)$ where $\boldsymbol{\beta}_a$ is a $(r \times r)$ non-singular matrix defined by $\boldsymbol{\beta}' = (\boldsymbol{\beta}_a, \boldsymbol{\beta}_b)$. In this case, $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}(\boldsymbol{\beta}_a \boldsymbol{\beta}_a^{-1\prime}\boldsymbol{\beta}') = \boldsymbol{\alpha}(\mathbf{I}_r, \widetilde{\boldsymbol{\beta}})$ where $\mathbf{I}_r$ is the $(r \times r)$ unit matrix, and $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_a^{-1\prime}\boldsymbol{\beta}_b$ is an $(r \times (p-r))$ full-rank matrix. These just-identifying restrictions have transformed $\boldsymbol{\beta}$ to the long-run 'reduced form'. Because just-identifying restrictions do not change the likelihood function, no tests are involved. In general, just identification can be achieved by imposing one appropriate normalization and $(r-1)$ restrictions on each $\boldsymbol{\beta}_i$. Care is required to ensure that the coefficient which is normalized is non-zero.

The calculation of the degrees of freedom when testing structural hypotheses on the cointegration relations is often quite difficult. It is useful from the outset to distinguish between:

(1) pseudo 'restrictions' that can be obtained by linear manipulations, because it is always possible to impose $r-1$ restrictions and one normalization without changing the value of the likelihood function – no testing is involved for such 'restrictions';
(2) additional testable restrictions on the parameters, which change the value of the likelihood function;
in the latter group there are two kinds of testable restrictions:
   (a) restrictions that are not identifying, for example: (i) the same restrictions on all cointegrating vectors, (ii) one vector assumed known and the remaining vectors unrestricted.
   (b) genuine over-identifying restrictions.

The first step is to examine whether the restrictions satisfy the rank and order condition for identification: luckily, many available software packages do the checking, and will usually inform the user when identification is violated. Also, note that when normalizing a $\boldsymbol{\beta}_i$ vector by diving through by a non-zero element β_{ij}, the corresponding $\boldsymbol{\alpha}_i$ vector will by multiplied by the same element, so normalization does not change $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$.

10.2 Hypotheses testing

Hypotheses on the cointegration vectors can be formulated in two alternative ways: either by specifying the s_i free parameters in each $\boldsymbol{\beta}$ vector, or by specifying the m_i restrictions on each vector. We consider each in turn. First:

$$\boldsymbol{\beta} = (\mathbf{H}_1\boldsymbol{\kappa}_1, ..., \mathbf{H}_r\boldsymbol{\kappa}_r),$$

where $\boldsymbol{\beta}$ is $(p_1 \times r)$, $\boldsymbol{\kappa}_i$ are $(s_i \times 1)$ coefficient matrices, and $\mathbf{H}_i$ are $(p_1 \times s_i)$ design matrices where p_1 is the dimension of $\mathbf{x}_{t-1}^*$ in (25). In this case, we use the design matrices to determine the s_i free parameters in each cointegration vector.

The other way of formulating restrictions is:

$$\mathbf{R}_1'\boldsymbol{\beta}_1 = \mathbf{0}, ..., \mathbf{R}_r'\boldsymbol{\beta}_r = \mathbf{0}$$

where $\mathbf{R}_i$ is a $p \times m_i$ restrictions matrix. Note that $\mathbf{R}_i = \mathbf{H}_{\perp,i}$, i.e., $\mathbf{R}_i'\mathbf{H}_i = \mathbf{0}$.

In some cases, we may want to test restrictions which are not identifying. Such restrictions could be the same restriction on all cointegration relations, for example long-run price homogeneity on all vectors. In this case, the $\mathbf{H}_i$ (or $\mathbf{R}_i$) are all identical, and we can formulate the hypothesis as $\boldsymbol{\beta} = \mathbf{H}\boldsymbol{\kappa}$, where $\boldsymbol{\kappa}$ is now an $s \times r$ matrix of free parameters. Another possibility is to test a hypothesis on just

one of the cointegration vectors. In this case, we formulate the hypothesis as $\boldsymbol{\beta} = \{\mathbf{H}_1\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, ..., \boldsymbol{\kappa}_r\}$, where $\mathbf{H}_1$ is a $(p \times s_1)$ matrix, $\boldsymbol{\kappa}_1$ is $(s_1 \times 1)$, and the other vectors are unrestricted. All these hypotheses can be tested by a likelihood-ratio procedure described in detail in Johansen and Juselius (1990, 1992, 1994).

Our simple empirical example consists of only two gasoline prices, so the number of interesting hypotheses to test is limited. Here, we will only test one hypothesis as an illustration and refer the interested reader to the many published papers containing a wide variety of testable hypotheses, for example Juselius (1995a, 1995b, 1998). Table 3 showed that the unrestricted coefficients of the first cointegration vector were almost equal with opposite signs, indicating long-run price homogeneity. This hypothesis can be formulated either as $\boldsymbol{\beta} = (\mathbf{H}\boldsymbol{\kappa})$ or, equivalently, as $\mathbf{R}'\boldsymbol{\beta} = \mathbf{0}$, where:

$$\mathbf{H}' = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\kappa} = \begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \end{pmatrix}, \quad \text{and } \mathbf{R}' = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}.$$

The first alternative specifies the hypothesis in terms of the $s_1 = 3$ remaining free parameters; the second in terms of the $m_1 = 1$ imposed restriction. The hypothesis is tested by the LR test procedure described in Johansen and Juselius (1992). The test statistic value was 0.55, distributed as $\chi^2(1)$, and hence accepted with a p-value of 0.46. The restricted cointegration relation became:

$$p_1 - p_2 - \underset{(1.8)}{0.02}\, Ds94.34 - \underset{(0.6)}{0.010}$$

$$\alpha_{11} = \underset{(7.3)}{-0.14}, \quad \alpha_{12} = \underset{(0.8)}{-0.01}$$

The constant term is not significantly different from zero, but for reasons discussed above in Section 6, we prefer to keep it in the cointegration relation. Similarly, the step dummy is only borderline significant (cf. the discussion in Section 9.1), and suggests that the permanent increase in gasoline prices at location 1 might (possibly) have been slightly higher than at location 2. The estimates of the adjustment coefficients α_{1i} show significant adjustment in the first price, but not in the second. This suggests that a price shock leading to an increase in the price differential between the two locations will initiate an adjustment in the first price, but not in the second: in this sense, the second location is a price leader, i.e., is 'driving' the gasoline prices in this market. This takes us to the next issue, which is exogeneity and partial systems.

11 Partial systems and exogeneity

The hypothesis that a variable is influencing the long-run development of the other variables of the system, but is not influenced by them, is called the hypothesis of 'no levels feedback', or long-run weak exogeneity (when the parameters of interest are $\boldsymbol{\beta}$). We test the following hypothesis on $\boldsymbol{\alpha}$:

$$\mathcal{H}_{\boldsymbol{\alpha}}(r) : \boldsymbol{\alpha} = \mathbf{A}\widetilde{\boldsymbol{\alpha}} \tag{29}$$

where $\boldsymbol{\alpha}$ is $p \times r$, $\mathbf{A}$ is a $p \times s$ matrix, $\widetilde{\boldsymbol{\alpha}}$ is an $s \times r$ matrix of nonzero $\boldsymbol{\alpha}$-coefficients and $s \geq r$. (Compare this formulation with the hypothesis of the same restriction on all $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta} = (\mathbf{H}\widetilde{\boldsymbol{\beta}})$.) As with tests on $\boldsymbol{\beta}$, we can express the restriction (29) in the equivalent form:

$$\mathcal{H}_{\boldsymbol{\alpha}}(r) : \mathbf{B}\boldsymbol{\alpha} = \mathbf{0} \tag{30}$$

where $\mathbf{B} = \mathbf{A}_{\perp}$.

The condition $s \geq r$ implies that the number of non-zero rows in $\boldsymbol{\alpha}$ must not be greater than r. This is because a variable that has a zero row in $\boldsymbol{\alpha}$ is not adjusting to the long-run relations and, hence, can be considered as a driving trend in the system, i.e., as a common stochastic trend. Since there can at most be $(p-r)$ common trends, the number of zero-row restrictions can at most be equal to $(p-r)$.

The hypothesis (29) can be expressed as:

$$\begin{pmatrix} \boldsymbol{\alpha}_a \\ \boldsymbol{\alpha}_b \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\alpha}} \\ \mathbf{0} \end{pmatrix}.$$

Under $\mathcal{H}_0$:

$$\Delta \mathbf{x}_t = \boldsymbol{\Phi}_1 \Delta \mathbf{x}_{t-1} + \widetilde{\boldsymbol{\alpha}} \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\pi} + \boldsymbol{\Psi} \mathbf{d}_t + \boldsymbol{\epsilon}_t. \tag{31}$$

The weak exogeneity hypothesis can be tested with a LR test procedure described in Johansen and Juselius (1992). It is asymptotically distributed as χ^2 with the degrees of freedom $\upsilon = s \times r$, i.e., equal to the number of zero restrictions on the α-coefficients.

If the zero-row restriction on $\boldsymbol{\alpha}$ is accepted, we can partition the p variables into $(p-s)$ variables which exhibit levels feedback, and s variables with no levels feedback. We say that the s variables are weakly exogenous when the parameters of interest are $\boldsymbol{\beta}$. Because the s weakly-exogenous variables do not contain information about the long-run parameters, we can obtain fully-efficient estimates of $\boldsymbol{\beta}$ from the $(p-s)$ adjustment equations, conditional on the marginal models of the s weakly-exogenous variables (see Engle, Hendry and Richard, 1983, Johansen, 1992b, and Hendry, 1995b). This gives the condition for when partial models can be used to estimate $\boldsymbol{\beta}$ without losing information. More formally, this can be stated as: let $\{\mathbf{x}_t\} = \{\mathbf{x}_{a,t}, \mathbf{x}_{b,t}\}$ where $\mathbf{x}_{b,t}$ is weakly exogenous when $\boldsymbol{\beta}$ is the parameter of interest; then a fully efficient estimate of $\boldsymbol{\beta}$ can be obtained from the partial model:

$$\Delta \mathbf{x}_{a,t} = \mathbf{A}_0 \Delta \mathbf{x}_{b,t} + \boldsymbol{\Phi}_{11} \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha}_1 \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\pi}_1 + \boldsymbol{\Psi}_1 \mathbf{d}_t + \mathbf{v}_{1,t}. \tag{32}$$

In our simple bivariate model of the two gasoline prices, weak exogeneity can be tested as a hypothesis on one $\boldsymbol{\alpha}$ coefficient. Because $p = 2$ and $r = 1$, there can at most be one weakly exogenous variable. It appeared from Table 3 that $\widehat{\alpha}_{11}$ is statistically significant, whereas $\widehat{\alpha}_{21}$ is not. Hence, in this case $p_{2,t}$ is weakly exogenous based on a t-test. In the bivariate system, the t-test and the χ^2-test are asymptotically equivalent. In a larger system with several cointegrating relations, the hypothesis that a variable is long-run weakly exogenous is that a row of α-coefficients is zero, i.e., $\mathbf{R}'\boldsymbol{\alpha} = \mathbf{0}$. In our simple example, $\mathbf{R} = (0, 1)$. The test statistic value, distributed as $\chi^2(1)$, was 0.48 and, hence, is accepted.

Long-run weak exogeneity does not imply short-run weak-exogeneity. Hence, it does not exclude the possibility that $p_{1,t}$ is reacting in the short run to changes in $p_{2,t}$. But in many cases, the economic interest is in the long-run effects, and establishing long-run weak exogeneity for a variable is often used as a justification for continuing the model analysis conditional on the weakly exogenous variable(s). Although a full system analysis is usually needed to test for weak exogeneity (see Harbo, Johansen, Nielsen and Rahbek, 1998, for a test procedure based on a partial system), there can be some advantages to continuing the analysis in a partial system. For example, if the weakly-exogenous variable has been subject to many interventions (large shocks) and current changes in this variable have significantly affected the other variables in system, so $\mathbf{A}_0 \neq \mathbf{0}$ in (32), then conditioning on the weakly-exogenous variable is likely to reduce the need for intervention dummies in the model, and possibly lead to more constant parameter estimates. For example, if the model had additionally included the price of crude oil and it had been found to be weakly exogenous, conditioning on the current change of the price of crude oil would probably have taken care of many of the large changes in gasoline prices that in the present

model had to be accounted for by dummy variables. Moreover, the residual variance in the conditional model is often much reduced relative to the marginal model, hence improving the precision of statistical inference.

In our simple bivariate model, $p_{2,t}$ was found to be weakly exogenous and we present the estimation results when conditioning on it in (33). Note that this model is the equivalent of the dynamic, single-equation EqCM in Hendry and Juselius (2000):

$$\begin{array}{rcl} \Delta p_{1,t} & = & \underset{(14.8)}{0.48}\ \Delta p_{1,t-1} \quad \underset{(25.1)}{0.86}\ \Delta p_{2,t} \quad - \underset{(7.2)}{0.33}\ \Delta p_{2,t-1} \quad \underset{(8.7)}{-0.12}\ ecm_{t-1} \quad + \cdots \\ & & \widehat{\sigma}_\epsilon = 0.016, \qquad \mathsf{R}^2 = 0.77 \end{array} \tag{33}$$

As discussed above, the need for dummy variables changed in this model version: five of the seven dummy variables became insignificant after conditioning on $\Delta p_{2,t}$. But, because the residual standard error dropped from 0.023 to 0.016, some of the larger price changes no longer passed the outlier criterion $|\widehat{\epsilon}_{i,t}| > 3.3$, so in the final model we needed the same number of outlier dummies as before, albeit at different dates.

12 Forecasting in cointegrated processes

The existence of cointegration between series importantly affects forecasts of them, since some linear combinations remain 'linked', whereas others drift apart. It seems less important whether or not such cointegration links are imposed in estimation: the additional 'errors' from not doing so are $O(1/T)$. The topic is treated at length in Clements and Hendry (1998, 1999); here we consider the simplest case, where there are no unmodeled structural changes in the data generation process.

Consider an h-period ahead forecast based on a dynamic system for the vector of p variables $\mathbf{x}_t$, using the simplest EqCM:

$$\mathbf{x}_t = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \ \text{ where } \ \boldsymbol{\epsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]. \tag{34}$$

Then, h-periods ahead, from an end-of-sample point T (the forecast origin), the outcome is:

$$\mathbf{x}_{T+h} = \boldsymbol{\Pi}^h \mathbf{x}_T + \sum_{i=0}^{h-1} \boldsymbol{\Pi}^i \left(\boldsymbol{\pi} + \boldsymbol{\epsilon}_{T+h-i}\right). \tag{35}$$

The VAR forecast is:

$$\widehat{\mathbf{x}}_{T+h} = \widehat{\boldsymbol{\Pi}}^h \mathbf{x}_T + \sum_{i=0}^{h-1} \widehat{\boldsymbol{\Pi}}^i \widehat{\boldsymbol{\pi}},$$

which can be rather uncertain due to powering up $\widehat{\boldsymbol{\Pi}}$ (which contains estimated unit roots), and not partitioning the components of $\widehat{\boldsymbol{\pi}}$ between equilibrium means and growth rates.

When there is cointegration, we can write (34) as the EqCM:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t. \tag{36}$$

It can be shown that when forecasting increasingly further ahead, forecasts of cointegrating combinations converge on their equilibrium means, as is appropriate for stationary combinations: $\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{x}}_{T+h} \to \boldsymbol{\mu}$. Consequently, forecasts of changes converge on $\boldsymbol{\gamma}$, so any EqCM effect on $\Delta\mathbf{x}_t$ dies out: $\widehat{\Delta\mathbf{x}}_{T+h} \to \boldsymbol{\gamma}$. However, for levels ($\widehat{\mathbf{x}}_{T+h}$), the EqCM coefficient increases: thus, a small feedback coefficient $\boldsymbol{\alpha}$ may nevertheless entail a large long-run effect from the disequilibrium.

Imposing valid cointegration restrictions can improve forecast accuracy somewhat (see Clements and Hendry, 1995), but for short-horizon forecasts, any potential benefit can be overwhelmed by changes in the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ (surprisingly, changes to the other parameters are less pernicious: see Hendry, 2000). Imagine an unmodeled change in $\boldsymbol{\mu}$ to $\boldsymbol{\mu}^*$ (say): the model treats the discrepancy $\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right)$ as a disequilibrium, so forecasts of changes go in the *opposite* direction to the data – if growth actually increases, the model will forecast a fall. For example, if the price differential between the two gasoline prices changed permanently at the outbreak of the Kuwait war (i.e., relative gasoline prices moved to a new equilibrium position), without us modeling it explicitly, then the model will interpret the increased price differential as a disequilibrium, so (wrongly) forecast a reduction back to the old equilibrium. This is why it is important to view the model as 'equilibrium correcting': it always adjusts to the imposed equilibrium (whether that is correct or not), and certainly does not 'error correct' after a shift to a new equilibrium position. However, Clements and Hendry (1999) consider a variety of solutions to this problem.

13 Parameter constancy and policy

Constancy of the parameters of models is a fundamental attribute for inference, theory, forecasting and policy, and is no less relevant in cointegrated systems. Many tests exist for that hypothesis (see e.g., Hendry, 1995a, and Doornik and Hendry, 1999), and it has been established that the conventional distributions of diagnostic test statistics remain valid in integrated-cointegrated systems (see e.g., Wooldridge, 1999), so the usual tests remain applicable.

We remarked in the previous section on the crucial role of changes in deterministic terms in inducing forecast failure, and the consequential ease of detection, as against the difficulty of detecting changes in (say) $\boldsymbol{\alpha}$. Unfortunately, some procedures for analyzing policy in cointegrated VARs, such as impulse-response analyses, are heavily dependent on the absence of changes in the parameters of the dynamics – the very feature that is difficult to detect. Thus, their implications must remain tentative pending better tests for mean-zero parameter changes (for additional critiques, see Ericsson, Hendry and Mizon, 1998, and Hendry and Mizon, 1998: Banerjee, Hendry and Mizon, 1996, provide an overview).

We also remarked on the potential advantages of 'open' or partial systems which validly conditioned on weakly-exogenous variables. At first sight, it would seem unlikely that policy variables could be weakly exogenous, since the disequilibria in an economy are a major determinant of policy actions, and cointegration deviations comprise an important class of disequilibria. However, provided the cointegration analysis is appropriately undertaken in the full (closed) system, and only then are the cointegration vectors computed, weak exogeneity (necessary for efficient conditional inference) need not be invoked. Once the $\boldsymbol{\beta}'\mathbf{x}_t$ are estimated, they become variables, the presence of which in several equations, including those equations determining conditional policy variables, is inconsequential. Thus, conditional policy models can be constructed. This is invaluable, since most practical policy involves changes in variables with non-zero means, and hence unmodeled changes would be easily detected. Thus, one can learn of any mistakes quickly, and rapidly adapt to them, rather than persist with an incorrect policy.

14 Conclusions

The recognition that economic time series are non-stationary has profoundly altered the technology of econometrics, introducing the concepts and tools associated with integrated-cointegrated data that have been the focus of our two-part exposition. In part, 'plus ça change, plus c'est la meme chose' rules:

many inferences can be conducted as in a stationary world, since unit roots can be removed by the linear transformations of cointegration and differencing. But, additional sources of non-stationarity still remain problematic, especially structural change; changes in the covariances of the data violate one of the underlying assumptions of the VAR model. Nevertheless, we can understand, and resolve other aspects of unit-root data such as 'nonsense regressions' (dating back to the century before last), how to avoid them (as well as deliberately create them, as in Hendry, 1980). In part I, we formalized an appropriate canonical model, the equilibrium-correction model, relevant for non-stationarity (unit-root) economic data, which resolves the 'nonsense regression' problem. Here we have analyzed the properties of this model in further detail, demonstrated the changes needed to validate inference procedures, and illustrated the powerful new modeling procedures with gasoline price series relevant for energy economics.

Application of cointegration analysis requires careful thought about model specification and interpretation, and an increased emphasis on the appropriate treatment of deterministic terms. How they enter a model, and whether the same terms also enter the DGP both affects parameter inference and the behavior of the model for such purposes as policy and forecasting. Recent developments in the theory of forecasting, summarized in the two volumes Clements and Hendry (1998, 1999), have highlighted the key role in forecast failure of shifts in the deterministic terms, so more attention should be paid to the theory and practice of modeling deterministic terms.

The reader who has followed both parts of our overview in their entirety has traveled a long route. We have explained the many concepts, and the associated language of the 'unit-root' revolution in econometrics, which admittedly has been a slow revolution! Regrettably we suspect that the worry expressed in Working (1934) that 'Economic theory has fallen far short of recognizing the full implications of the resemblance of many economic time series to random-difference series' still remains. Modeling cointegrated series is difficult because of the need to model systems of equations in which one has to simultaneously specify the deterministic terms and how they enter, determine the lag length, and ensure a congruent representation. Nevertheless, powerful software facilitates the task for those wishing to undertake their own analyses, including the programs *PcGive* and CATS in RATS that we have utilized. We also hope that publications on the topic will be more comprehensible in the future.

To those who found the material hard going, we also draw on an old quote, by a discussant of Yule (1926), who complained about the mathematical difficulty of that paper – yet school mathematics today easily suffices to understand it. There is no going back....

15 Acknowledgments

All the computations reported in this paper were carried out using the *PcGive* suite (see especially Doornik and Hendry, 1996, 1999), and CATS in RATS. We are pleased to acknowledge financial support from the Danish Social Sciences Research Council, and the UK Economic and Social Research Council under grant L116251015. We are grateful to Campbell Watkins for helpful comments on a previous draft.

References

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.

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

Clements, M. P., and Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.

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

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

Doornik, J. A., and Hendry, D. F. (1996). *GiveWin: An Interactive Empirical Modelling Program*. London: Timberlake Consultants Press.

Doornik, J. A., and Hendry, D. F. (1999). *Modelling Dynamic Systems using PcFiml 9 for Windows*. London: Timberlake Consultants Press.

Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, **12**, 533–572.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *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., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

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

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, **60**, 203–233.

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

Harbo, I., Johansen, S., Nielsen, B., and Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business and Economic Statistics*, **16**, 388–399.

Hendry, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, **4**, 51–88. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F. (1980). Econometrics: Alchemy or science?. *Economica*, **47**, 387–406. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

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

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

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

Hendry, D. F., and Doornik, J. A. (1996). *Empirical Econometric Modelling using PcGive for Windows*. London: Timberlake Consultants Press.

Hendry, D. F., and Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, **21**, 1–42.

Hendry, D. F., and Mizon, G. E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics*, **23**, 267–294.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254. Reprinted in R.F. Engle and C.W.J. Granger (eds), Long-Run Economic Relationships, Oxford: Oxford University Press, 1991, 131–52.

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.

Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, **13**, 205–229.

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

Johansen, S. (2000). A small sample correction to the test for cointegrating rank in the vector autoregressive model. Mimeo, European University of Institute, Florence.

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration – With application to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Johansen, S., and Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*, **53**, 211–244.

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

Johansen, S., Mosconi, R., and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Journal of Econometrics*, forthcoming.

Jørgensen, C. (2000). On the I(2) cointegration model. Chapter 3 in Ph.D. Thesis, 65, University of Copenhagen, Institute of Economics, Copenhagen.

Juselius, K. (1995a). Do purchasing power parity and uncovered interest rate parity hold in the long run? An example of likelihood inference in a multivariate time-series model. *Journal of Econometrics*, **69**, 211–240.

Juselius, K. (1995b). An empirical analysis of the changing role of the German Bundesbank after 1983. *Oxford Bulletin of Economics and Statistics*, **58**, 791–819.

Juselius, K. (1998). A structured VAR for Denmark under changing monetary regimes. *Journal of Business and Economic Statistics*, **16**, 400–412.

Juselius, K. (2000). On the role of deterministic components in the cointegrated VAR model. Unpublished report, University of Copenhagen, Institute of Economics, Copenhagen.

Nielsen, B., and Rahbek, A. (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics*, **62**, 5–22.

Wooldridge, J. M. (1999). Asymptotic properties of some specification tests in linear models with integrated processes. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 366–384. Oxford: Oxford University Press.

Working, H. (1934). A random difference series for use in the analysis of time series. *Journal of the American Statistical Association*, **29**, 11–24.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in

sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society*, **89**, 1–64.