

# Inference in Cointegrated Models: UK M1 Revisited

Jurgen A. Doornik, David F. Hendry and Bent Nielsen\*  
Nuffield College,  
Oxford.

July 27, 1998

## Abstract

The paper addresses the empirical application of multivariate cointegration analysis to a small model of narrow money, prices, output and interest rates in the UK. Practical determination of cointegration rank is difficult for many reasons: deterministic terms play a crucial role in limiting distributions, and systems may not be formulated to ensure similarity to nuisance parameters; finite-sample critical values may differ from asymptotic equivalents (the latter are usually obtained via simulation, but could be based on response surfaces); dummy variables alter critical values, often greatly; multiple cointegration vectors must be identified to allow inference; the data may be  $I(2)$  rather than  $I(1)$ , altering distributions; and conditioning to partial systems must be done with care.

## Keywords

Cointegration; deterministic terms; indicator variables; weak exogeneity; identification.

*JEL classifications:* E41, C52.

## 1 Introduction

The literature on the formulation, estimation, and testing of models for potentially cointegrated economic time series is truly vast, bordering on a complete discipline in its own right. Even texts which summarize the material comprise many hundreds of pages: see, *inter alia*, Banerjee, Dolado, Galbraith and Hendry (1993), Hamilton (1994), Hendry (1995), Johansen (1995b) (double that length when the companion workbook, Hansen and Johansen, 1998, is added), and Hatanaka (1996). Since, we could only consider a small fraction of the topic, we will focus on one salient problem: determining the cointegration rank of a linear dynamic system for economic time series.

Determining cointegration rank is difficult in practice for many reasons, including:

- the presence or absence of deterministic terms (such as constants and trends) in the generating process and/or the model can greatly alter limiting distributions;
- the system may have been formulated in such a way that the (asymptotic) similarity of key test statistics to nuisance parameters is lost;
- alternative choices of test statistics may deliver apparently conflicting inferences;
- finite-sample critical values can differ notably from asymptotic equivalents;
- the latter are usually approximations, obtained by simulation perhaps summarized by response surfaces;

---

\*Financial support from the UK Economic and Social Research Council under grant R000237500 is gratefully acknowledged. We are grateful to Hans-Martin Krolzig for helpful discussions on the topic. The authors' web pages are respectively: [www.nuff.ox.ac.uk/users/doornik/](http://www.nuff.ox.ac.uk/users/doornik/), [www.nuff.ox.ac.uk/economics/people/hendry.htm](http://www.nuff.ox.ac.uk/economics/people/hendry.htm) [www.nuff.ox.ac.uk/users/nielsen/](http://www.nuff.ox.ac.uk/users/nielsen/).

- indicator variables for ‘blips’ can alter the outcome of the analysis;
- the lag length selected may not remove all residual autocorrelation, or may be too long;
- multiple cointegration vectors must be identifiable to allow coherent inference;
- the data may be  $I(2)$ , or near  $I(2)$ , rather than  $I(1)$ , again altering the relevant limiting distributions; and:
- there may be non-modelled variables, and conditioning to create ‘partial’ systems must be done with care, even under weak exogeneity.

This paper addresses many of these important issues of inference in the empirical application of multivariate cointegration analysis, illustrating the analysis by a (much studied) four-equation model of narrow money (M1), prices, aggregate expenditure, and interest rates in the UK. These data have been the subject of extensive analysis: see, *inter alia*, Hendry (1979), Hendry (1985), Ericsson, Campos and Tran (1990), Hendry and Ericsson (1991), Boswijk (1992), Hendry and Mizon (1993), Hendry and Doornik (1994), and Harris (1995) as well as Johansen (1992b), Paroulo (1996) and Rahbek, Kongsted and Jørgensen (1998) who suggest they are  $I(2)$ . Different investigators have also found varying numbers of cointegrating vectors, though all agree on at least one connecting these four variables.

Part of the difficulty in the empirical analysis is the important role played by structural breaks, including financial liberalization, and both external and domestic shocks. The first of these induced rapid growth in money holdings relative to nominal income, perhaps part of the explanation for the  $I(2)$  hypothesis, and is modelled here by using an appropriate measure of the opportunity cost of holding money. The treatment of the last two (oil shocks and major government budget changes) is problematic, and we consider below how indicator variables affect the empirical analysis. Consequently, the data modelling raises the typical problems empirical researchers regularly confront, and illustrates many of the aspects of the theoretical analysis of cointegration.

The structure of the paper is as follows. To set the scene on the roles of the various forms of non-stationarity, §2.1 discusses the effects of near unit roots in a scalar example, and §2.2 looks at the impacts of blip dummies. Section 3 discusses the statistical problem, beginning with a description of the time-series data we will analyse (§3.1), the economic analysis (§3.2), the notation and formulation of linear dynamic systems (§3.3), and the associated preliminary empirical analysis (§3.4): empirical illustrations occur throughout the text, and are not concentrated in one section. Then §3.5 introduces the cointegration representation, followed by the empirical unrestricted cointegration analysis in §3.6. Next, the impacts of various formulations of intercepts and linear deterministic trends are considered in §4. New Monte Carlo simulation evidence illustrating the importance of the formulation of deterministic effects is presented in §5. Section 6 considers the impacts of impulse dummy variables (indicators) in cointegrated systems. Section 7 reviews other issues arising in determining the cointegration rank of a closed system, so briefly discusses recursive estimation, finite-sample critical values, and approximations to the asymptotic distributions of the tests. This is followed by a discussion of identification issues in §8.1, and the estimation of the restricted empirical cointegration relations. The problems when an  $I(2)$  analysis is needed are noted in §9. Then, §10 turns to the analysis of long-run weak exogeneity and conditional systems, where new problems in determining cointegration rank appear. Finally, §11 concludes. The Appendix gives relevant analytical results.

## 2 Background

This section first illustrates the problems arising when roots of the dynamics are close to unity, as often happens for empirical economic data; the ideas generalize to testing for cointegration in systems.

Secondly, since our model of UK M1 includes two indicator variables to remove the impacts of interventions, we consider that issue in §2.2.

## 2.1 Unit roots

Consider the scalar, mean-zero, first-order autoregressive process:

$$y_t = \rho y_{t-1} + \nu_t \text{ where } \nu_t \sim \text{IN} [0, \sigma_\nu^2]. \quad (1)$$

The  $\nu_t$  are independent and normally distributed with mean zero and variance  $\sigma_\nu^2$ . When  $|\rho| < 1$ , the least-squares estimator  $\hat{\rho}$  is asymptotically distributed as:

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{D} \text{N} [0, (1 - \rho^2)]. \quad (2)$$

When  $\rho \simeq 1$ , but still lies in the stationarity region, then (2) suggests that in large samples,  $\hat{\rho}$  will have a negligibly small variance. This transpires to be a poor approximation in small samples, and not very good in large. Figure 1a reports Monte Carlo sampling standard deviations (MCSDs, which estimate the actual variability in  $\hat{\rho}$ ) and estimated standard errors (MCSEs, based on the usual regression formula) when  $\rho = 0.9999$ , together with their theoretical values  $s = \sqrt{T^{-1}(1 - \rho^2)}$  from (2).<sup>1</sup> The ratios of the MCSDs and MCSEs to  $s$  are shown in panel c and reveal departures as large as 20-fold when  $T \simeq 50$ ; even for  $T = 350$ , the MCSD overestimates  $s$  by 10-fold. However, when  $\rho = 0.8$ , the theory and practice are much closer, as panels b and d show: the maximum departures are about 10%, with the ratios converging on unity.<sup>2</sup>

Approximation (2) works poorly because the form of the limiting distribution changes as  $\rho \rightarrow 1$  and a reformulation is required when  $\rho = 1$ . Write (1) as:

$$\Delta y_t = \gamma y_{t-1} + \nu_t \quad (3)$$

where  $y_0 = 0$  and  $\gamma = \rho - 1$ . When  $\gamma = 0$ , then  $\Delta y_t = \nu_t$ . Now the estimate  $\hat{\gamma}$  of  $\gamma$  needs to be normalized by  $T$  (rather than  $\sqrt{T}$  as in (2)), and does not converge on a normal distribution asymptotically, but to a functional of Brownian motion with a non-zero variance, and a substantial negative bias. This makes discrimination between  $I(1)$  and  $I(0)$  difficult (see, *inter alia*, Hendry, 1995, Ch.4). Economic data are certainly not stationary, and even if exact unit roots were not present in economics, many of the following results would be more useful as approximations to practical behaviour than assuming stationarity.

## 2.2 Indicator variables

Indicators are often added to remove the impacts of ‘outliers’ and thereby obtain a better estimate of the innovation variance. However, as §6 below shows, introducing dummy variables into cointegration analyses raises many issues, even when their existence is based on good historical grounds. This section first illustrates the potential problems in a stationary dynamic model. Let:

$$y_t = \rho y_{t-1} + \psi z_t + \nu_t \text{ where } \nu_t \sim \text{IN} [0, \sigma_\nu^2] \quad (4)$$

<sup>1</sup>1000 replications were used in PcNaive: see Hendry, Neale and Ericsson (1991). In figures,  $2 \times 2$  panels of graphs are notionally labelled  $a, b; c, d$ , such that  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  denotes the location;  $3 \times 3$  panels as  $a, b, c; d, e, f; g, h, i$ ; and so on.

<sup>2</sup>There remained considerable sampling uncertainty in the MCSD at 1000 replications, so 10,000 were used in panels b and d; that the MCSE underestimates the MCSD may be due to averaging estimated standard errors, rather than variances.

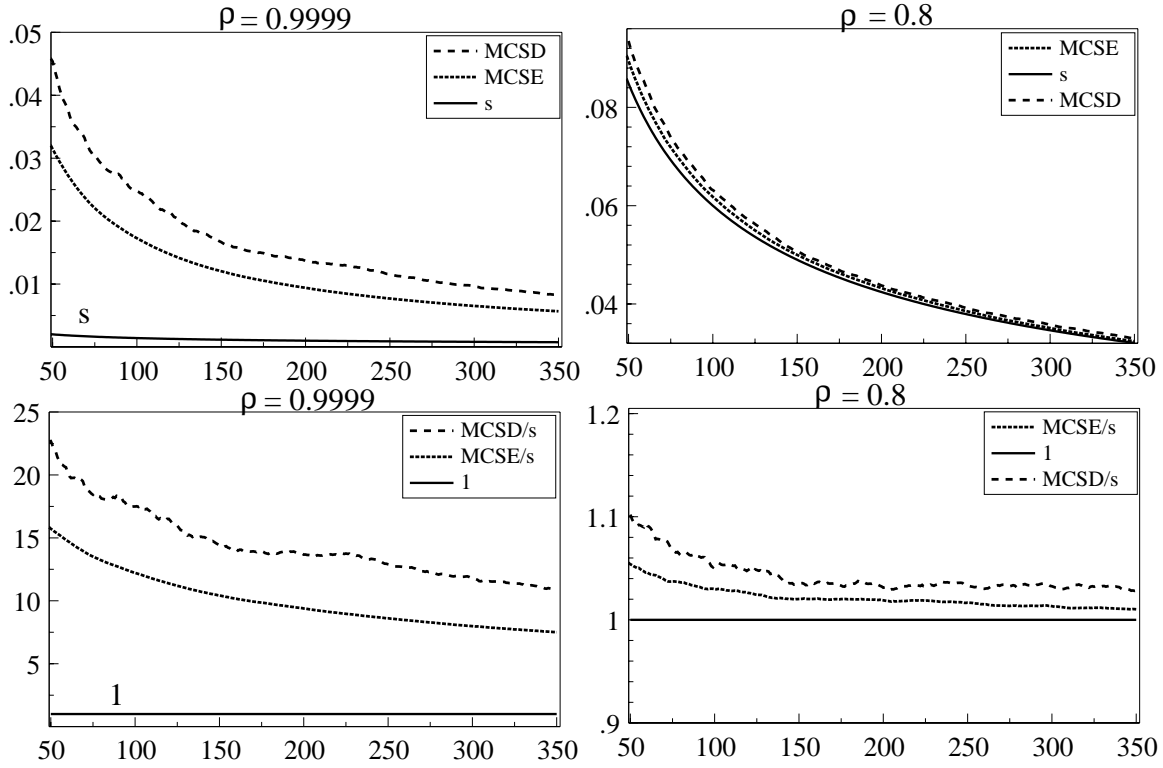


Figure 1 MCSDs and MCSEs for  $\rho = 0.9999$  and  $\rho = 0.8$ .

when  $|\rho| < 1$ . In (4),  $z_t = 1_{\{t=T_b\}}$  is a zero-one indicator ( $1_{\{t=T_b\}} = 0$  except for  $t = T_b$ ). In a static setting, including  $z_t$  is equivalent to dropping the  $T_b^{\text{th}}$  observation. Since:

$$y_t = \sum_{i=0}^{\infty} (\rho^i \psi z_{t-i} + \rho^i \nu_{t-i}),$$

then, letting  $\sigma_u^2 = \sigma_\nu^2 / (1 - \rho^2)$ :

$$T^{-1} \sum_{t=1}^T \mathbb{E} [y_t^2] \simeq \sigma_u^2 + \frac{T^{-1} \psi^2}{1 - \rho^2}. \quad (5)$$

When  $\psi$  is a fixed number, its effect on (5) is negligible for large  $T$ , and hence it has no influence on the limiting distribution. Indeed, least-squares estimation of (4) yields:

$$\begin{pmatrix} \sqrt{T}(\hat{\rho} - \rho) \\ \hat{\psi} - \psi \end{pmatrix} \xrightarrow{\mathcal{D}} \mathbf{N}_2 \left[ 0, \begin{pmatrix} 1 - \rho^2 & 0 \\ 0 & \sigma_\nu^2 \end{pmatrix} \right].$$

The limiting distribution of  $\sqrt{T}(\hat{\rho} - \rho)$  is unaffected by the inclusion of the dummy, and does not depend on the value of  $\psi$ . The residual variance remains a consistent estimator of  $\sigma_\nu^2$ . However, since different scalings are needed on the two estimators to obtain non-degenerate limiting distributions,  $\hat{\psi}$  is inconsistent for  $\psi$ . Omitting the indicator from the estimated model does not affect the limiting distribution, but would bias estimators of  $\rho$  and  $\sigma_\nu^2$  in finite samples.

When the impulse is large in terms of the error relative to the available sample, approximated by  $\psi = \sqrt{T} \delta \sigma_\nu$ , then:

$$T^{-1} \sum_{t=1}^T \mathbb{E} [y_t^2] \simeq \sigma_u^2 (1 + \delta^2). \quad (6)$$

As  $y_{T_b-1}$  does not depend on  $\psi$ :

$$\begin{pmatrix} \hat{\rho} - \rho \\ \hat{\delta} - \delta \end{pmatrix} \underset{\widehat{app}}{\sim} \mathbf{N}_2 \left[ 0, T^{-1} \begin{pmatrix} (1 - \rho^2)(1 + \delta^2)^{-1} & 0 \\ 0 & 1 \end{pmatrix} \right]. \quad (7)$$

Thus, the approximate distribution of  $\hat{\rho}$  in (7) is affected by the size of the break, but is little affected by the inclusion or exclusion of the dummy. This is unlike a static model, where including the indicator would completely remove the effect of the ‘blip’. Again the residual variance is unbiased, but is biased if  $z_t$  is omitted. Finally, the appropriately-scaled dummy  $\delta$  has a variance of  $O(T^{-1})$ : for example, if  $\sigma_v = 0.01$  (1%) and  $T = 100$ , then  $\psi = 0.05$  corresponds to  $\delta = 0.5$  ( $\text{SE} = 0.1$ ).

### 3 Formulation of the statistical problem

The analysis commences from an unrestricted Gaussian vector autoregression, such as that in (10) below. The objective is to find an empirically well-behaved specification as the starting point for cointegration analysis. But first we discuss the data and the economic analysis.

#### 3.1 Data

The data used to illustrate issues in cointegration inference are quarterly, seasonally-adjusted, time series over 1963(1)–1989(2) on  $M$ ,  $I$ ,  $P$  and  $R_n$  for the United Kingdom.<sup>3</sup>

- $M$  nominal  $M_1$ ,
- $I$  real total final expenditure ( $TFE$ ) at 1985 prices,
- $P$  the  $TFE$  deflator,
- $R_a$  the three-month local authority interest rate,
- $R_m$  learning-adjusted interest rate,
- $R_n$   $R_a - R_m$ .

$R_m$  is the learning-adjusted interest rate on interest-bearing checking accounts at commercial banks. This type of account was introduced in 1984(3), see Hendry and Ericsson (1991) for details. Money and expenditure are in £ million, the deflator is unity in 1985, and the interest rates are annual, in fractions. Lower-case letters denote logs of the corresponding capitals. After allowing for lags, estimation is usually over 1964(3)–1989(2), which yields 100 observations (Ericsson, Hendry and Tran, 1994, analyse the corresponding raw data, but here we focus on the seasonally-adjusted series: also see Hendry and Mizon, 1993, and Hendry and Doornik, 1994).

The lower-right panel in Figure 2 shows the time series of the observations for  $m - p$ ,  $i$ ,  $\Delta p$ ,  $R_n$  (standardized to facilitate comparison).

#### 3.2 The economic background

The theoretical formulation entails relationships to determine the demand for money, aggregate expenditure, inflation, and the opportunity cost of holding money. We consider these in turn.

Despite being primarily used for transactions, narrow money is also part of financial portfolios, and is held as a liquid reserve for contingencies. Thus, the quantity of nominal money demanded ( $M^d$ ) depends on the price level ( $P$ ), the volume of real transactions to be financed ( $I$ ), the opportunity cost of inter-temporal holding, measured by inflation ( $\dot{p}$ ), and on both the own rate of interest and competing

<sup>3</sup>The data are supplied with the demo version of PcGive, available from Doornik’s web page.

rates of return on alternative liquid assets.<sup>4</sup> Since money demand should be unit free, the relation is usually assumed to be homogeneous of degree one in  $P$  (i.e., real money is demanded), increasing in  $I$  (sometimes homogeneously as well), decreasing in both inflation and  $R_a$ , and increasing in  $R_m$ . Frequently, a log-linear form is assumed linking  $M$ ,  $P$ , and  $I$ , with interest rates entering in levels, and we write this schematically, in steady state, as:

$$m^d - p = \tau_0 + \tau_1 i - \tau_2 \dot{p} - \tau_3 R_a + \tau_4 R_m. \quad (8)$$

The coefficients  $\tau_1 \dots \tau_4$  are anticipated to be positive as written, probably with  $\tau_1 = 1$ , and  $\tau_3 = \tau_4$ . Equation (8) defines the anticipated cointegration relation. However, dynamics are central to many theories of money demand: here economic agents are assumed to have upper and lower targets for their desired real-money holdings, and adjust balances back towards the mean when these bands are exceeded.<sup>5</sup> While the observed money stock also depends on the supply, on the basis of institutional knowledge, we assume that the monetary authority controls the shortest interest rate, and manipulates that to achieve its policy objectives.

Total real expenditure is very schematically modelled here in terms of a trend, approximating growth in human and physical capital, both of which embody technical progress, and demand determinants, represented by the real interest rate  $R_a - 4\dot{p}$  (inflation is re-scaled to annual units), leading to:

$$i = \rho_0 + \rho_1 t - \rho_2 (R_a - 4\dot{p}). \quad (9)$$

Again, dynamic adjustments to the path in (9) need to be modelled empirically. In both (8) and (9), the log form is adopted in the expectation of relatively constant proportional errors. The lack of data on  $R_m$  for most of the sample led us to replace  $-\tau_3 R_a + \tau_4 R_m$  in (8) by  $-\tau_3 R_n$ , and  $-\rho_2 (R_a - 4\dot{p})$  in (9) by  $-\rho_2 R_n + \rho_3 \dot{p}$ . Large policy changes (such as major budget changes) are handled using an indicator variable, called *dout* (see §3.4).

Inflation is a complex phenomenon, and the model lacks many of the variables that might be expected to account for its behaviour, including exchange rates and world prices, government deficits, factor-market conditions, and commodity prices. Thus, the most likely long-run determinants here are the excess demands for money and goods embodied in deviations from (8) and (9). We also introduce a non-modelled indicator for special effects from price shocks (such as the oil crises).

Finally, despite being a net interest rate,  $R_n$  is sufficiently short term to be treated here as being set by the monetary authority, perhaps to stabilize the excess demands in (8) and (9), or achieve a target for inflation.

### 3.3 Linear dynamic systems

We consider a closed, linear dynamic system for  $n$  variables  $\mathbf{x}_t$ , with a maximum lag length of  $s$  periods, and assume normality, thereby postulating a vector autoregression (VAR) with  $m$  deterministic variables  $\mathbf{q}_t$ , over a sample  $t = 1, \dots, T$ , expressed as:

$$\Delta \mathbf{x}_t = \sum_{j=1}^{s-1} \mathbf{\Pi}_j \Delta \mathbf{x}_{t-j} + \boldsymbol{\pi} \mathbf{x}_{t-1} + \boldsymbol{\Phi} \mathbf{q}_t + \boldsymbol{\nu}_t \quad \text{where } \boldsymbol{\nu}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}], \quad (10)$$

when  $\text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}]$  denotes an  $n$ -dimensional independent, normal density with mean zero and covariance matrix  $\boldsymbol{\Omega}$  (symmetric, positive definite). In (10), the parameters  $(\mathbf{\Pi}_1 \dots \mathbf{\Pi}_{s-1}, \boldsymbol{\pi}, \boldsymbol{\Phi}, \boldsymbol{\Omega})$  are assumed

<sup>4</sup>A measure of the volatility of income might reflect precautionary demands, but is omitted here: see e.g., Tobin (1958).

<sup>5</sup>See e.g., Miller and Orr (1966), Akerlof (1973), Milbourne (1983), and Smith (1986).

to be constant and variation free, with sufficient observations to sustain estimation and inference ( $T - s \gg n(s + m + 2)$ ). To allow interpretation of the results, none of the roots of the companion-form polynomial should lie inside the unit circle (see e.g., Banerjee *et al.*, 1993, Ch. 5). We focus on  $s = 2$  for exposition, and denote  $\Pi_1$  by  $\Pi$ . Note that (10) is isomorphic to a VAR in levels.

### 3.4 Preliminary empirical analysis

For coherent inference, it is important that the empirical model satisfies the assumptions of the statistical analysis. We begin by analysing  $m - p$ ,  $i$ ,  $\Delta p$ , and  $R_n$ , thereby imposing long-run price homogeneity; we return in §9 to consider the properties of the levels  $m$  and  $p$ . The four series analysed are treated as potentially  $I(1)$  over this sample. The initial VAR in (10) for  $\{(m - p)_t, i_t, \Delta p_t, R_{n,t}\}$  has 2 lags and includes an intercept unrestrictedly, and a trend. Earlier research suggests that longer lags are not needed. We base our specification closely on Hendry and Doornik (1994), who provide modelling details, so include their dummy variables. These are *dout* for output shifts, zero except for unity in 1972(4), 1973(1), 1979(2), and *doil* for price shocks (such as the oil crises and VAT changes), equal to unity in 1973(3), 1973(4), and 1979(3). These indicators adjust for the largest residuals in the Hendry and Mizon (1993) model. The sample period (after creating lags) is 1964(3)–1989(2), so that  $T = 100$ . Computations and graphics were produced with GiveWin and PcFiml (see Doornik and Hendry, 1996, 1997); simulations were done in Ox (Doornik, 1998b).

Table 1 reports summary evaluation statistics for the estimated VAR:  $\hat{\sigma}$  denotes the standard deviation of the residuals. The diagnostic tests comprise 5<sup>th</sup>-order residual vector serial correlation ( $F_{ar5}^v$ : a multivariate version of Breusch, 1978–Godfrey, 1978, see Doornik, 1995b) and vector heteroscedasticity ( $F_{het}^v$ : a multivariate version of White, 1980, see Doornik, 1995b), as well as a chi-square test for joint normality ( $\chi_{nd}^{2v}(8)$ : see Doornik and Hansen, 1994). Significance at the 5% and 1% levels is denoted by \* and \*\* respectively; p-values are given in square brackets.

Table 1 System evaluation.

	$m - p$	$i$	$\Delta p$	$R_n$
$\hat{\sigma}$	1.65%	1.04%	0.68%	1.31
	statistic	value	p-value	
	$F_{ar5}^v(80, 258)$	1.21	[0.13]	
	$F_{het}^v(180, 548)$	0.97	[0.59]	
	$\chi_{nd}^{2v}(8)$	19.4*	[0.013]	

Apart from the non-normality due to some remaining outliers in the inflation equation, the results are consistent with a congruent system. All first lags and the indicators are significant, but the second lags and the trend appear insignificant (on  $F(4, 85)$ , at 5% or less). The equations for  $\Delta p$  and  $R_n$  show some non-constancy, although the system break-point Chow (1960) test nowhere exceeds the 1% critical value within sample.

Following this preliminary analysis, it is of interest to determine the dynamic properties of the system. The eigenvalues of the long-run matrix  $\pi$  in (10) are  $-0.39$ ,  $-0.17$ , and  $-0.05 \pm 0.05i$  (using  $i$  to denote  $\sqrt{-1}$ ), so the rank seems non-zero, and is unlikely to be greater than two. The eigenvalues  $\lambda$  of the companion form (the inverses of the roots) are shown in Table 2.

Two roots are close to unity, two have moduli near 0.7, and the remainder are small, so overall, this representation appears to be  $I(1)$ , probably with two cointegrating vectors, and two unit roots. We now

Table 2 System dynamics.

$$\begin{bmatrix} \lambda & 0.96 \pm 0.02i & 0.68 \pm 0.03i & -0.34 & 0.32 & -0.21 \pm 0.10i \\ |\lambda| & 0.96, 0.96 & 0.68, 0.68 & 0.34 & 0.32 & 0.24, 0.24 \end{bmatrix}$$

formalize the cointegration analysis.

### 3.5 Cointegration formulation

Equation (10) shows that the matrix  $\pi$  determines how the levels of the process  $\mathbf{x}$  enter the system: for example, when  $\pi = \mathbf{0}$ , the dynamic evolution does not depend on the levels. This indicates the importance of the rank of  $\pi$  in the analysis. The statistical hypothesis of cointegration is:

$$H(r): \text{rank}(\pi) \leq r.$$

Under this hypothesis,  $\pi$  can be written as the product of two matrices:

$$\pi = \alpha\beta',$$

where  $\alpha$  and  $\beta$  have dimension  $n \times r$ , and vary freely. As suggested by Johansen (1988, 1995b), such a restriction can be analysed by maximum likelihood methods. The idea is to find the canonical correlations between the first differences  $\Delta\mathbf{x}_t$  and the lagged levels  $\mathbf{x}_{t-1}$ , having corrected both of these for the other components in (10). It is then possible to test that the  $n - r$  smallest canonical correlations are zero.

Keeping the lag length at two ( $s = 2$ ), and restricting (10) by the hypothesis  $H(r)$ , we obtain:

$$\Delta\mathbf{x}_t = \Pi\Delta\mathbf{x}_{t-1} + \alpha(\beta'\mathbf{x}_{t-1}) + \Phi\mathbf{q}_t + \nu_t. \quad (11)$$

In applications, it is of interest to estimate the rank rather than just finding an upper bound of the form  $\text{rank}(\pi) \leq r$ . When the rank equals  $r$ , the properties of the system can be interpreted using the Granger representation theorem (see Engle and Granger, 1987, and Johansen, 1995b, Theorem 4.2).

When  $\mathbf{x}_t$  is  $I(1)$ , then the first differences  $\Delta\mathbf{x}_t$  and the  $r$  cointegrating relations  $\beta'\mathbf{x}_t$  are  $I(0)$ . To ensure that  $\mathbf{x}_t$  is  $I(1)$  and not  $I(2)$ , we require  $\text{rank}(\alpha'_\perp\Gamma\beta_\perp) = n - r$ . Here  $\Gamma = -(\mathbf{I}_n - \Pi + \pi)$  is the mean-lag matrix,  $\alpha_\perp$  and  $\beta_\perp$  are  $n \times (n - r)$  matrices such that  $\alpha'_\perp\alpha = \mathbf{0}$ ,  $\beta'_\perp\beta = \mathbf{0}$  with  $(\alpha : \alpha_\perp)$  and  $(\beta : \beta_\perp)$  being rank- $n$  matrices. The issue of  $I(2)$ -ness is discussed in §9.

### 3.6 Unrestricted cointegration analysis

Prior to a theoretical analysis of the deterministic terms, our  $I(1)$  cointegration analysis restricts the trend to the cointegration space (coefficients denoted by  $\rho$ ), and enters the constant ( $\phi$ ) and dummies ( $\mathbf{d}_t$ ) unrestrictedly. Equation (11) then becomes:

$$\Delta\mathbf{x}_t = \Pi\Delta\mathbf{x}_{t-1} + \alpha \begin{pmatrix} \beta \\ \rho \end{pmatrix}' \begin{pmatrix} \mathbf{x}_{t-1} \\ t \end{pmatrix} + \phi + \Upsilon\mathbf{d}_t + \nu_t. \quad (12)$$

The first part of Table 3 reports the log-likelihood values ( $\ell$ ), and eigenvalues ( $\mu$ ). The latter are the squared canonical correlations between first differences and lagged levels with trend (corrected for lagged differences, constant and dummies). The remainder of Table 3 reports the trace and maximum eigenvalue (Max) statistics together with the first two estimated cointegrating vectors, and the  $p$ -values for the trace test. Table 4 records the feedback coefficients ( $\hat{\alpha}$ ) and their standard errors when  $r = 2$ .

Before discussing the interpretation of these results, we consider the roles of the constant and trend.



Table 3 Cointegration analysis.

$$\begin{bmatrix} r & 0 & 1 & 2 & 3 & 4 \\ \ell & 1799.7 & 1838.3 & 1845.5 & 1849.8 & 1852.9 \\ \mu & & 0.55 & 0.13 & 0.08 & 0.06 \end{bmatrix}$$

$$\begin{bmatrix} H(r) & r = 0 & r \leq 1 & r \leq 2 & r \leq 3 \\ \text{Trace} & 108.5^{**} & 29.3 & 14.8 & 6.2 \\ & [0.00] & [0.55] & [0.60] & [0.45] \\ \text{Max} & 79.3^{**} & 14.5 & 8.6 & 6.2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}' & m-p & y & \Delta p & R_n & t \\ 1 & 1 & -1.00 & 7.34 & 7.65 & -0.0005 \\ 2 & -0.06 & 1 & -3.38 & 0.86 & -0.0059 \end{bmatrix}$$

Table 4 Feedback coefficients for rank 2.

$$\begin{bmatrix} \hat{\alpha} & 1 & 2 \\ m-p & -0.09 & -0.01 \\ y & -0.02 & -0.10 \\ \Delta p & -0.00 & 0.08 \\ R_n & -0.00 & -0.06 \end{bmatrix}
\quad
\begin{bmatrix} \text{SE}[\hat{\alpha}] & 1 & 2 \\ m-p & 0.012 & 0.074 \\ y & 0.007 & 0.047 \\ \Delta p & 0.005 & 0.031 \\ R_n & 0.009 & 0.060 \end{bmatrix}$$

## 4 Intercepts and linear deterministic trends

Deterministic terms, such as the intercept, linear trend, and indicator variables, play a crucial role in both data behaviour and limiting distributions of estimators and tests in integrated processes: see, for example, Johansen (1994). Depending on their presence or absence, the system may manifest drift, linear trends in cointegration vectors, or even quadratic trends (although the last seems unlikely in economics). Appropriate formulation of the model is important to ensure that cointegrating-rank tests are not too dependent on ‘nuisance parameters’ related to the deterministic terms. Here we consider the intercept and trend; §6 considers dummies.

The impact on the process  $\mathbf{x}_t$  of  $\mathbf{q}_t$  with parameter  $\Phi$  in (11) can be described using Granger’s representation theorem as:

$$\mathbf{C}\Phi \sum_{i=0}^t \mathbf{q}_i + \sum_{i=0}^{\infty} \mathbf{C}_i \Phi \mathbf{q}_{t-i}, \quad (13)$$

where  $\mathbf{C} = \beta_{\perp}(\alpha'_{\perp} \Gamma \beta_{\perp})^{-1} \alpha'_{\perp}$ . Two distinct effects are apparent: a cumulative influence through  $\mathbf{C}\Phi$ , and a distributed lag with coefficients  $\mathbf{C}_i \Phi$ . When  $\Phi = \alpha \mathbf{R}$  (say) the former vanishes since  $\mathbf{C}\alpha = \mathbf{0}$ . It is no surprise that this affects the distributions of test statistics, which we now consider in detail.

#### 4.1 Statistical analysis

When determining rank, three models merit consideration. These can be described by the dependence of the expected values of  $\mathbf{x}$  and  $\beta'\mathbf{x}$  on functions of  $t$ :

Hypothesis	$\mathbf{x}$	$\beta'\mathbf{x}$
$H_l(r)$	linear	linear
$H_c(r)$	constant	constant
$H_z(r)$	zero	zero

In these models, the process  $\mathbf{x}$  and the cointegrating relations exhibit the same deterministic pattern. At a later stage, when the rank has been determined, it will be possible to consider further models of the trending behaviour. Note that, under  $H_z(r)$ , it is necessary that  $E[\mathbf{x}_0] = E[\Delta\mathbf{x}_0] = \mathbf{0}$  to ensure that the non-stationary components have zero expectation. Likewise, for the other models, the conditions on the initial values must be such that they preserve the postulated behaviour.

The hypotheses are formalized in Table 5 in terms of the parameters of (11). The parameters  $\alpha, \beta \in \mathbb{R}^{n \times r}$ ,  $\phi, \in \mathbb{R}^n$ , and  $\rho_c, \rho_l \in \mathbb{R}^r$  vary freely when present.

Table 5 Models for rank and deterministic trend.

$H_l(r)$	$\pi = \alpha\beta'$	and	$\Phi\mathbf{q}_t = \phi + \alpha\rho_l t$ ,
$H_c(r)$	$\pi = \alpha\beta'$	and	$\Phi\mathbf{q}_t = \alpha\rho_c$ ,
$H_z(r)$	$\pi = \alpha\beta'$	and	$\Phi\mathbf{q}_t = \mathbf{0}$ .

Let  $H(r)$  be one of  $H_l(r)$ ,  $H_c(r)$ , and  $H_z(r)$ . The hypotheses are nested as follows:

$$H(0) \subset \dots \subset H(r) \subset \dots \subset H(n). \quad (14)$$

The rank can be determined consistently by adopting the procedure given by Johansen (1995b, Section 12.3). Start by testing  $H(0)$  against the general alternative  $H(n)$ . If  $H(0)$  is rejected, test  $H(1)$  against  $H(n)$ , and so on. The rank is estimated as  $r$  if  $H(r)$  is the first hypothesis which cannot be rejected. The test statistic which is used in this procedure is the trace test. As pointed out by Nielsen and Rahbek (1998), these tests are asymptotically similar with respect to the parameters related to the deterministic components (provided that the  $I(1)$  conditions are satisfied).

The maximum eigenvalue test is the likelihood ratio test of  $H(r-1)$  against  $H(r)$ . While this test is sometimes used in practice, a corresponding result of consistency has not been established.

#### 4.2 Test results

There may be economic ground for preferring one specification over another. In particular, for our model, we prefer  $H_l(r)$  based on the following economic considerations:

- While inflation and  $R_n$  should not drift, real  $TFE$  has grown at an annual rate of about 2.5%, precluding a zero intercept. A long-run unit income elasticity of demand for real money then restricts its intercept to equal that for  $TFE$ .
- A linear trend in the cointegration space approximates growth in  $TFE$  from cumulative human and physical capital; its coefficient must therefore match that in a).
- When the analysis uses dummies, as with *doil* and *dout* in our case, similar considerations apply. We return to this issue in §6.

Referring back to Table 3, and noting that it corresponds to  $H_l(r)$ , we can now interpret the results of the test statistics. Following the rank selection procedure described in the previous section, the null of no cointegration is strongly rejected. Although the second cointegrating vector is not significant, we retain it following Hendry and Mizon (1993). The interpretation of the coefficients in relation to the economic issues in §3.2 will be considered in §8 after imposing further restrictions.

### 4.3 Further models

From Table 5, let  $\phi_c = \alpha\rho_c$  and  $\phi_l = \alpha\rho_l$ . Two additional models arise when  $\phi_c$  and  $\phi_l$  are allowed to vary freely, and we embed these in an extension of that table:

	Hypothesis	$\mathbf{x}$	$\beta'\mathbf{x}$
$H_{ql}(r)$	$\pi = \alpha\beta', \Phi\mathbf{q}_t = \phi_c + \phi_l t$	quadratic	linear
$H_l(r)$	$\pi = \alpha\beta', \Phi\mathbf{q}_t = \phi_c + \alpha\rho_l t$	linear	linear
$H_{lc}(r)$	$\pi = \alpha\beta', \Phi\mathbf{q}_t = \phi_c$	linear	constant
$H_c(r)$	$\pi = \alpha\beta', \Phi\mathbf{q}_t = \alpha\rho_c$	constant	constant
$H_z(r)$	$\pi = \alpha\beta', \Phi\mathbf{q}_t = \mathbf{0}$	zero	zero

In terms of the notation used in PcFiml, the hypotheses are:

Hypothesis	trend	constant
$H_{ql}(r)$	unrestricted	unrestricted
$H_l(r)$	restricted	unrestricted
$H_l(r)$	absent	unrestricted
$H_c(r)$	absent	restricted
$H_z(r)$	absent	absent

Likelihood-ratio test statistics for the two additional models have also been derived by Johansen (1995b). The asymptotic distribution under  $H_{ql}(r)$  depends on whether or not  $\alpha'_\perp \phi_l = \mathbf{0}$ , and this complicates the rank determination considerably (*op. cit.*, Theorem 6.2). To develop a consistent test procedure, the idea is to only test  $H_{ql}(r)$  if  $H_l(r)$  has been rejected (*op. cit.*, Ch. 12). In that case, we rule out the possibility that the rank is at most  $r$  as well as  $\alpha'_\perp \phi_l = \mathbf{0}$ , and therefore the assumptions of the asymptotic theory are satisfied. The relevant hypotheses are nested as:

$$\begin{array}{ccccccc}
 H_{ql}(0) & \subset & \cdots & \subset & H_{ql}(r) & \subset & \cdots & \subset & H_{ql}(n) \\
 \cup & & & & \cup & & & & \parallel \\
 H_l(0) & \subset & \cdots & \subset & H_l(r) & \subset & \cdots & \subset & H_l(n).
 \end{array} \tag{15}$$

By testing the hypotheses

$$H_l(0), H_{ql}(0), H_l(1), H_{ql}(1), \dots, H_l(n-1), H_{ql}(n-1),$$

sequentially against the unrestricted alternative and stopping whenever the hypothesis is accepted, a consistent procedure is obtained.

A corresponding complication arises with  $H_{lc}(r)$ . The test procedure is then based on:

$$\begin{array}{ccccccc}
 H_{lc}(0) & \subset & \cdots & \subset & H_{lc}(r) & \subset & \cdots & \subset & H_{lc}(n) \\
 \cup & & & & \cup & & & & \parallel \\
 H_c(0) & \subset & \cdots & \subset & H_c(r) & \subset & \cdots & \subset & H_c(n).
 \end{array} \tag{16}$$

When we allow for a quadratic trend in the UK-M1 model we find for the trace test:

$$\begin{bmatrix} & r = 0 & r \leq 1 & r \leq 2 & r \leq 3 \\ H_{ql}(r) & 97.4^{**} & 22.4 & 8.1 & 1.2 \\ H_l(r) & 108.5^{**} & 29.3 & 14.8 & 6.2 \end{bmatrix}.$$

We encounter  $108.5^{**}$ ,  $97.4^{**}$ ,  $29.3$ , so that the first hypothesis to be accepted is  $H_l(1)$ . Therefore the quadratic trend is rejected, and the conclusion is as before.

It is possible, but not very likely, that an insignificant value is followed by a significant statistic. An example would be: reject  $H_l(0)$ , accept  $H_l(1)$ , and reject  $H_l(2)$ . This could be indicative of more general model mis-specification.

## 5 Illustrating the models of trend behaviour

The practical problem of estimating the cointegrating rank for an appropriate treatment of the deterministic terms can be done by the various sequential testing strategies just discussed. We now illustrate this using an artificial DGP based on the empirical model in this paper. The data exhibit a linear trend, apparent in fig. 2d, so we adopt the model with a restricted trend and two cointegrating vectors to generate the data.

### 5.1 The artificial DGP

The artificial DGP satisfies  $H_l(2)$ . The design of the DGP closely mimics the UK-M1 empirical model, and uses the observations for 1964(1)–(2) as its initial conditions.

$$\begin{aligned} \Delta \mathbf{x}_t = & \begin{pmatrix} -0.102 & 0 \\ 0 & -0.149 \\ 0 & 0.036 \\ 0 & -0.04 \end{pmatrix} \begin{pmatrix} 1 & -1 & 6.41 & 7.16 & 0 & -0.209 \\ 0 & 1 & -2.13 & 1.48 & -0.0063 & -11.186 \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ t \\ 1 \end{pmatrix} \\ & + \begin{pmatrix} 0.0063 \\ 0.0063 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -0.3 & 0 & 0 & -0.06 \\ 0 & 0 & 0 & 0 \\ 0.068 & 0 & -0.26 & 0 \\ 0 & 0 & 0 & 0.17 \end{pmatrix} \Delta \mathbf{x}_{t-1} + \boldsymbol{\nu}_t, \end{aligned} \quad (17)$$

when:

$$\boldsymbol{\nu}_t \sim \text{IN}_4[\mathbf{0}, \boldsymbol{\Sigma}], \text{ where } \boldsymbol{\Sigma}^* = \begin{pmatrix} 1.6\% & & & \\ -0.08 & 1\% & & \\ -0.51 & 0.03 & 0.69\% & \\ -0.49 & 0.09 & 0.31 & 1.3\% \end{pmatrix},$$

using the lower triangle of  $\boldsymbol{\Sigma}^*$  to show the cross correlations of the errors, and:

$$\begin{pmatrix} \mathbf{x}'_{-1} \\ \mathbf{x}'_0 \end{pmatrix} = \begin{pmatrix} 10.9445 & 11.1169 & 0.000779 & 0.048967 \\ 10.9369 & 11.1306 & 0.013567 & 0.050 \end{pmatrix}.$$

Figure 2 records the data generated in three randomly-selected trials, together with the actual empirical outcomes in the fourth panel (all variables are standardized). The outcomes seem representative of the actual data.

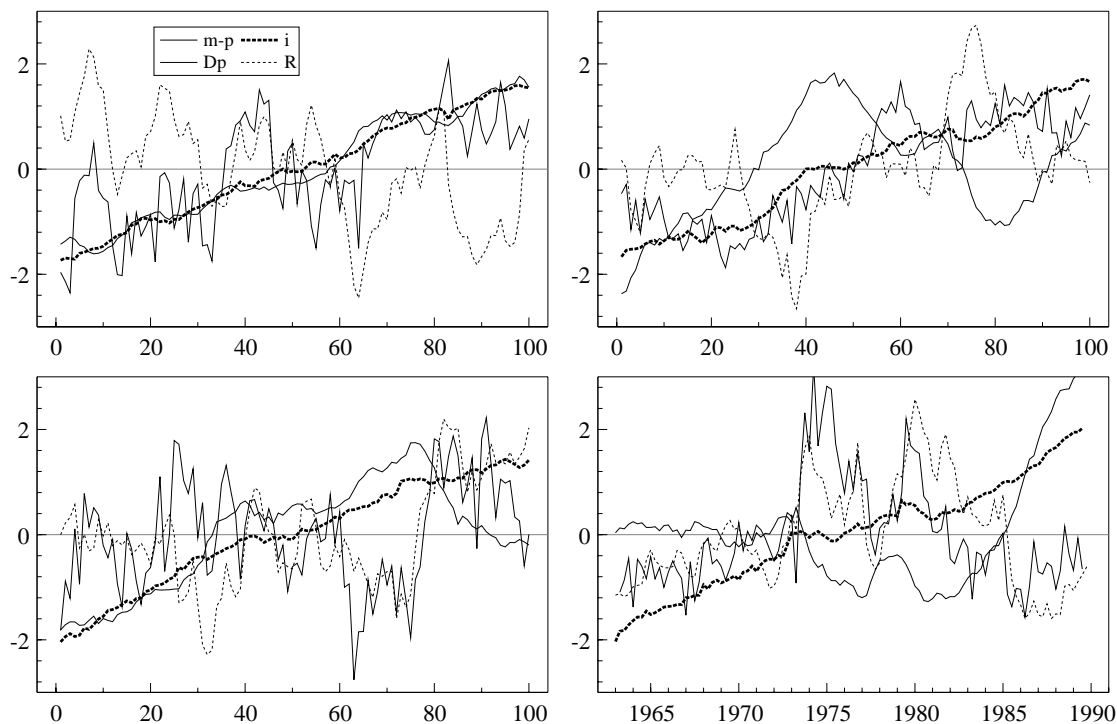


Figure 2 Data generated in three replications, with the actual outcomes.

## 5.2 Monte Carlo analysis

We now simulate the rejection frequencies of the tests  $H_{ql}(r)$ ,  $H_l(r)$ ,  $H_{lc}(r)$ , and  $H_c(r)$  when using data generated by (17). We do not consider  $H_z$ , because it is too far removed from our DGP. The first cointegrating vector is very well determined, with  $H(0)$  rejected virtually 100% of the time at all the sample sizes considered, so we omit this from the graphs.

The results are presented in fig. 3 as a  $4 \times 3$  matrix of graphs. The graphs give recursive Monte Carlo results for sample sizes  $T = 50, 75, 100, 125, 150, 200, 250, 500$ . In all cases, common random numbers were used, so that all graphs were based on the same data, and within each replication, a subset of the 500 data points was analysed. Each row corresponds to one of the four models for testing the rank hypothesis. Each column corresponds to a different rank specification. Rejection frequencies are given at 10% and 5% level based on asymptotic critical values. For example, the graph for  $H_l(1)$  in the second row, (which also formed the basis for the DGP) shows a high probability for rejecting  $r \leq 1$ , which is satisfactory because the actual rank is two.

The second row outcomes for  $H_l$  correspond to our recommended treatment of deterministic factors: high rejection frequency of the null of no second cointegration vector, and somewhat oversized in rejecting the third. So adopting the procedure in (14) to determine the rank, we usually arrive at the correct conclusion. At  $T = 100$ , there is still some uncertainty about the second vector, but this problem gradually vanishes. The probability of rejecting  $r \leq 3$  is virtually zero, as might be hoped when the rank is only two (and is expected from asymptotic theory).

The first row shows that serious problems arise when rank determination is solely based on  $H_{ql}$ . The rejection frequencies for  $r \leq 2$  are considerably higher than 5% and 10% respectively, with even worse results for  $r \leq 3$ . However, when adopting the procedure (15), we reject  $H_l(1)$ ,  $H_{ql}(1)$ , and accept  $H_l(2)$ .

In the remaining two rows, the models are mis-specified relative to the DGP. The fourth row ( $H_c$ ) is

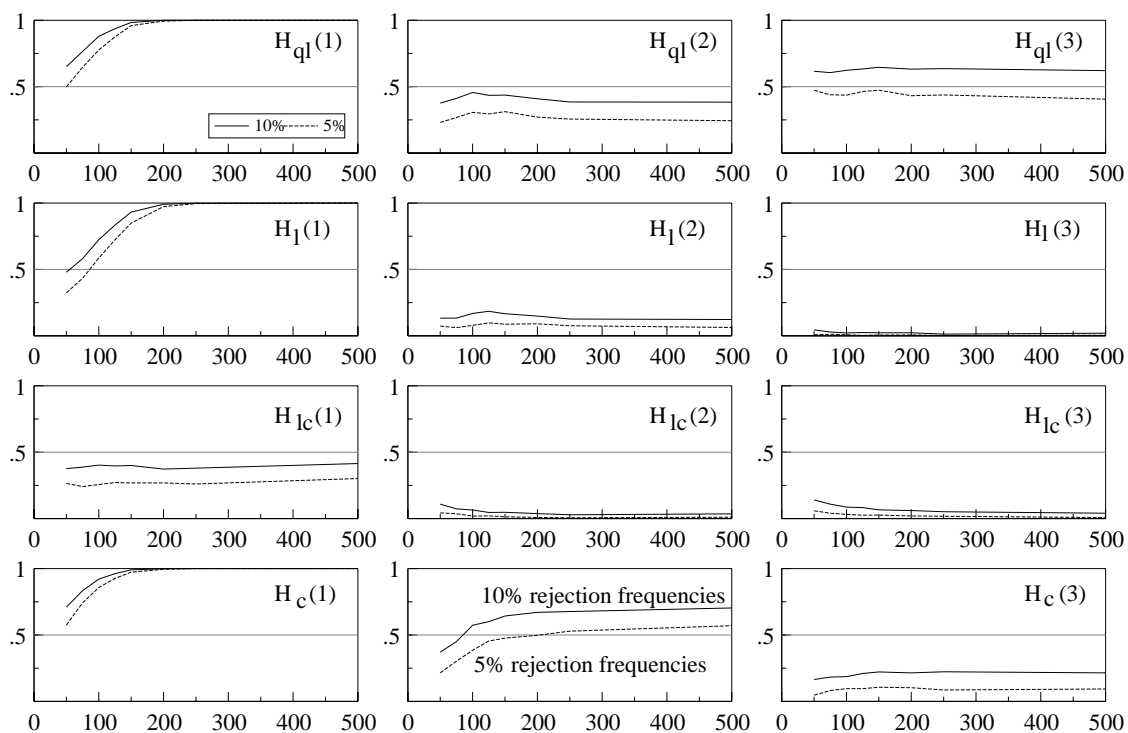


Figure 3 Cointegration rejection frequencies.

certainly unrealistic, in that it enforces no growth on the system. This highlights the severe drawbacks of inappropriate treatment of the deterministic terms: now it is quite possible to accept 3 cointegrating vectors when using the procedure (14).

When the deterministic trend is erroneously omitted as in the third row, there is a tendency to replace it by a stochastic trend. Of course, the correct procedure to use here is (16), but the mis-specification prevents us from reaching the right rank.

### 5.3 Over-specified trend

It is interesting to consider what happens when  $H_l$  is used to analyze data that do not in fact have a deterministic trend. To investigate this, we used two versions of the DGP. The first omitted the trend from the cointegrating vector (so was a member of  $H_{lc}(2)$ ), and the second only maintained the intercept in the cointegrating vectors (i.e., all three values 0.0063 were replaced with zero, placing it in  $H_c(2)$ ).

The resulting graphs were visually impossible to distinguish from the second row of fig. 3. This is explained by the (asymptotic) invariance to the value of the trend coefficient in the model  $H_l$ . Thus, adopting a model that includes a trend in the cointegration space has low cost even when the DGP does not have one. Such a beneficial outcome contrasts markedly with the costs of adding an unnecessary unrestricted trend.

### 5.4 Asymptotic analysis

The asymptotic background for the results reported in fig. 3 is now discussed. Further details are given in the appendix.

*Test for  $H_l(r)$  against  $H_l(n)$  when the DGP is  $H_l(s)$*

This is the standard case (second row), and the results correspond to the asymptotic analysis.

*Test for  $H_{ql}(r)$  against  $H_{ql}(n)$  when the DGP is  $H_l(s)$*

The first row of fig. 3 shows oversized tests, as expected from the asymptotic analysis.

Two asymptotic distributions are of relevance in the analysis of  $H_{ql}(r)$ . The standard distribution applies when the condition  $\alpha'_\perp \phi_l \neq \mathbf{0}$  is satisfied. However, when the DGP, as here, satisfies  $H_l(r)$  the condition is not satisfied since  $\alpha'_\perp \phi_l = \mathbf{0}$  and a different asymptotic distribution is relevant. The latter distribution is not tabulated in the literature. Simulations indicate that it dominates the standard distribution, and consequently the considered test is oversized.

In fig. 3, the test for  $H_{ql}(2)$  is based on the critical values 15.9 and 18.2, the 10% and 5% quantiles when  $\alpha'_\perp \phi_l \neq \mathbf{0}$ . The asymptotic p-values for these values are approximately 35% and 22% respectively when the rank is two as specified, but  $\alpha'_\perp \phi_l = \mathbf{0}$ . Similarly, the asymptotic p-values for the test for  $H_{ql}(3)$  are approximately 60% and 38% when  $\alpha'_\perp \phi_l = \mathbf{0}$  and the rank is two.

*Test for  $H_{lc}(r)$  against  $H_{lc}(n)$  when the DGP is  $H_l(s)$*

The first panels of the first two rows of fig. 3 show that the power of a test for  $\text{rank}(\boldsymbol{\pi}) \leq (r - 1)$  tends to one when  $\text{rank}(\boldsymbol{\pi}) = r$  in a well-specified model. This is not necessarily the case for a mis-specified model, as indicated by the first panel of the third row. Actually the rejection frequency converges to a constant less than unity when testing  $H_{lc}(1)$  using the DGP. This convergence is illustrated analytically in the Appendix.

*Test for  $H_c(r)$  against  $H_c(n)$  when the DGP is  $H_l(s)$*

For this test, the rejection frequency converges to one when testing that the rank is at most  $r$  and  $r < \text{rank}(\boldsymbol{\pi})$ . However, the rejection frequency of the test for  $H_c(\text{rank}(\boldsymbol{\pi}))$  converges to a constant less than unity, explaining the excess rejections in the second panel of row four. Again, the Appendix provides an analytic derivation.

## 6 Dummy variables

Two impulse dummies were included unrestrictedly in the analysis. Since the inclusion of dummies can greatly alter the distribution approximating the rank test, we consider the empirical and economic background to the *dout* and *doil* dummies in more detail, noting some specific problems related to the UK M1 analysis. The effects of dummies are studied using a Monte Carlo experiment; analytical results are presented in the Appendix.

### 6.1 Dummies in the UK M1 model

The indicators *dout* and *doil* serve two purposes. The first is to improve the fit of the model. The graphs of output  $i$ , and inflation  $\Delta p$ , indicate two transient shocks to the economy around 1973 and 1979 – see fig. 3. Also, the residuals of a VAR fitted without dummies have large outliers in these periods. This problem is removed by the inclusion of dummies. Initially, six separate impulse dummies were included, but, as a special feature of this data set, it was found that they could be collected into just two dummies as discussed by Hendry and Mizon (1993). The second feature of the dummies is that they describe shocks to the economy which can be attributed to specific events in the UK's economic history: the 'Heath-Barber' boom and the first effects of the Thatcher government for output, and the two oil crises for inflation.

There are several possible strategies for handling impulse dummies: they could be ignored, entered unrestrictedly, restricted to the cointegration space, or a mixture of the last two. It transpires that the size

of their effect matters, potentially even asymptotically. We now discuss the impact on the asymptotic distribution of the rank test of including small (really ‘not too big’) and large impulse dummies.

## 6.2 The rank test in the presence of small dummies

The impact of  $\mathbf{q}_t$  on  $\mathbf{x}_t$  was described in (13) above. The usual asymptotic results concerning the rank test hold as long as:

$$\frac{1}{\sqrt{T}} \Phi \sum_{i=0}^t \mathbf{q}_i \rightarrow \mathbf{0} \text{ for } T \rightarrow \infty,$$

so that the effect of the dummy is negligible as compared with a random walk (see Johansen, 1995a, §5.8). A break in the slope of a linear trend, corresponding to a change in the growth rate, would lead to new asymptotic tables. New tables also apply when a level shift is included in the cointegrating vector. These different asymptotic distribution could be simulated using the program DisCo (Johansen and Nielsen, 1993).

A dummy which is unity at a few points and zero otherwise may give a persistent shock to the non-stationary components of the process, but is usually asymptotically negligible. However, if the parameter of such a dummy is big in relation to  $\sqrt{T}$ , then this is not necessarily the case.

## 6.3 Distribution of the rank test in the presence of large dummies

In the analysis of Hendry and Doornik (1994), the impact of the dummies is quite dramatic. When the dummy for the output shocks, *dout*, was entered unrestrictedly in the model, one cointegration relation was found, whereas two were found when *dout* was restricted to the cointegrating vector. In this particular example, the estimated coefficient for the dummy is large, so the standard asymptotic distribution – which is derived by ignoring the dummies – is misleading. In their final model (Table 5), the coefficient for the *dout* dummy is 0.046. The dummy is unity for three out of 100 observations, and zero otherwise, hence, the cumulated effect of the dummy is 0.138. This number is approximately 10, or  $\sqrt{100}$ , times the standard deviation of the innovations, which is of order 0.014. The impact of such a big dummy is illustrated by the following example, related to that in §2.2; details are given in the Appendix.

Consider a data generating process given by:

$$\Delta \mathbf{x}_t = \rho \sqrt{T} 1_{\{t=T_b\}} + \boldsymbol{\nu}_t \text{ where } \boldsymbol{\nu}_t \sim \text{IN}[0, 1], \quad (18)$$

and  $1 < T_b < T$ . This is a unit-root process with a broken constant level. Assuming  $\mathbf{x}_0 = \mathbf{0}$ , then:

$$\mathbf{x}_t = \rho \sqrt{T} 1_{\{t \geq T_b\}} + \sum_{i=1}^t \boldsymbol{\nu}_i.$$

When these data are analysed using a univariate, first-order model, where the dummy is entered unrestrictedly as in Table 3 of Hendry and Doornik (1994), then the likelihood-ratio test for a unit root converges to a distribution which depends on the nuisance parameter  $\rho$ . However, when a univariate first-order model with a dummy restricted to the cointegration space is applied, as in Table 4 of Hendry and Doornik (1994), the eigenvalue converges to a non-zero distribution. Accordingly, the size of the likelihood-ratio test for a unit root converges to unity, leading to the conclusion that the process may be stationary. Thus, the two results for the UK money data could be explained by a DGP with one cointegrating vector and a shock generated by unrestricted dummies.



### 6.4 A Monte Carlo study of impulse dummies

To investigate the impact of dummies, we use the artificial DGP from §5.1. We start by investigating the impact of adding dummies to the model when they are not present in the DGP. So the baseline is a DGP under  $H_l(2)$  as in (17). In the experiments we restrict ourselves to two dummy variables. Let  $s_{49}$  and  $s_{55}$  denote step dummies with value zero before  $T = 49, 55$  respectively, and unity after. The impulse dummies are  $d_{49} = \Delta s_{49}$  and  $d_{55} = \Delta s_{55}$ . Estimation (after allowing for lags) was from  $T = 3$  onwards.

In the first experiment, there were no breaks in that DGP, but impulses were included in the statistical models (i.e., small dummies). Concerning trend and intercept, the model conforms to  $H_l$ . Four specifications were considered for the dummy variables:

- $M_1$ :  $d_{49}$  and  $d_{55}$  unrestricted
- $M_2$ :  $d_{49}$  and  $d_{55}$  restricted
- $M_3$ :  $d_{49}$  and  $d_{55}$  restricted,  $\Delta d_{49}$  and  $\Delta d_{55}$  unrestricted
- $M_4$ :  $s_{49}$  and  $s_{55}$  restricted,  $d_{49}$  and  $d_{55}$  unrestricted

Figure 4 shows the rejection frequencies for  $H_l(1)$  and  $H_l(2)$  for the four models. The graphs reveal that the specifications of the dummies had little impact on the test for  $r \leq 1$ , but could dramatically alter the outcome of testing  $r = 2$  when  $\text{rank}(\pi) = 2$ . Now only  $M_1$  produces reasonable sizes, which are essentially unchanged from fig. 3 (the first two graphs in the second row).

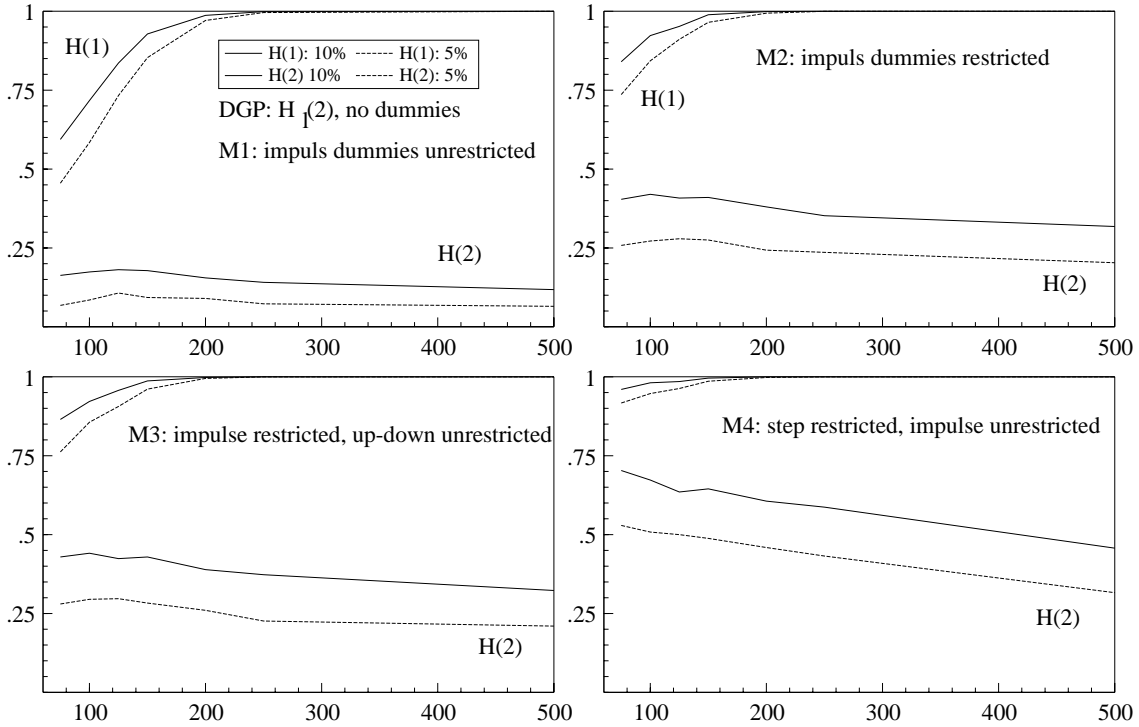


Figure 4 Impacts of impulse dummies on cointegration-rank test.

Changing the DGP to  $H_c(2)$  hardly alters the outcomes from fig. 4. However, changing the coefficients of the impulses in the DGP from zero can significantly alter the results. We consider two further versions of (17) under  $H_l(2)$ , where we add the two dummies unrestrictedly to the DGP:

- DGP<sub>1</sub>: small break:  $0.05d_{49}$  in  $x_2$ ,  $0.02d_{55}$  in  $x_3$ ,
- DGP<sub>2</sub>: large break:  $0.5d_{49}$  in  $x_2$ ,  $0.2d_{55}$  in  $x_3$ .

The two new DGPs are analyzed by  $M_0$  (no dummies in model), and  $M_1$ – $M_3$  from the previous experiment. This generates the eight graphs shown in fig. 5. The two columns correspond to small and large dummies, the four rows to the models. Panel a reveals that there is little problem in simply omitting the impulses when they are small, and b shows this finding extends approximately to large dummies. Panel c confirms the results in fig. 4a, and d shows that large dummies entered unrestrictedly, if anything, enhance that finding. The remaining four panels emphasise the disastrous nature of restricting the dummies to the cointegration space and adopting conventional critical values.

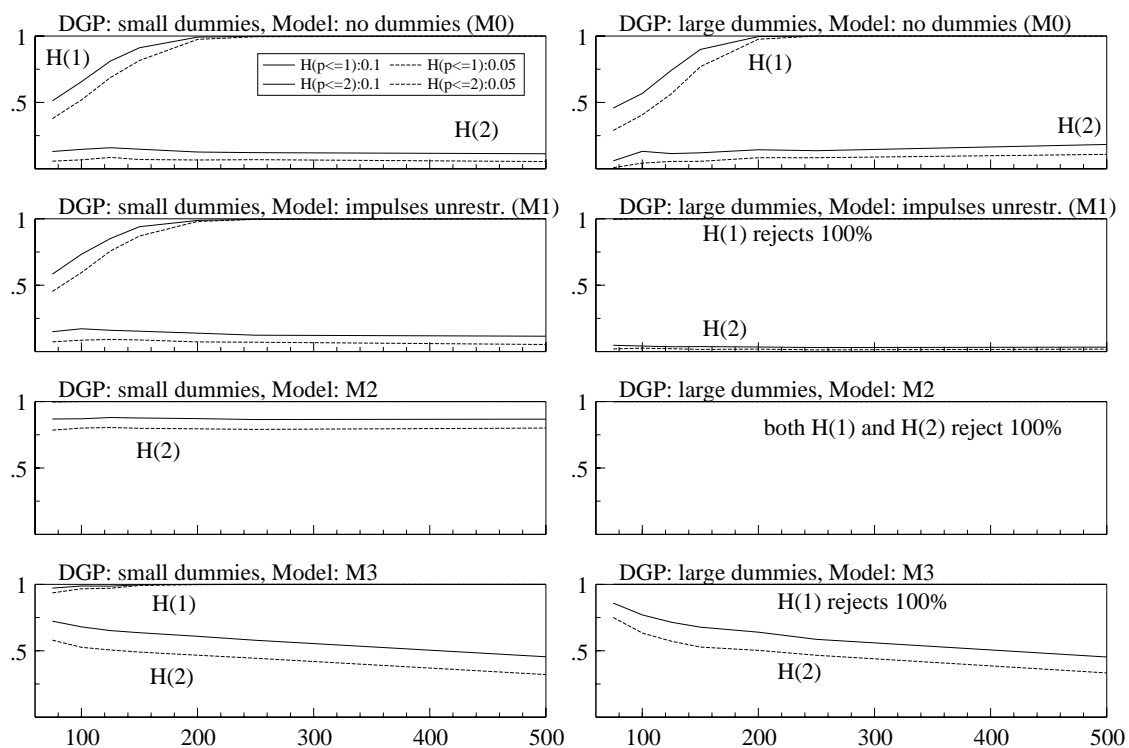


Figure 5 Breaks in cointegrating vectors and data.

## 7 Further issues in determining cointegration rank

Three further issues are noted in this section: recursive estimation of the eigenvalues  $\mu_i$ ; computationally convenient approximations to the asymptotic distribution of the rank test; and the small-sample properties of the trace test.

### 7.1 Recursive estimation

Graphs of the recursive estimates of the eigenvalues may be of help in revealing which eigenvalues stay systematically away from zero (see Hansen and Johansen, 1992, and Hansen and Johansen, 1998). It must be remembered that the eigenvalues are ordered at each sample size, and hence cannot ‘cross over’ by construction. Figure 8 below records the time series of the first two recursively-computed eigenvalues in panels b and d. They are shown for models with unrestricted dummies and with no impulses included. Both sets of eigenvalues are relatively constantly estimated, and the omission of the dummies has not worsened the constancy.

When plotting the recursive eigenvalues, there is a choice between re-estimating the complete cointegration analysis for every sample size, or estimating the short-run only once for the full sample, and

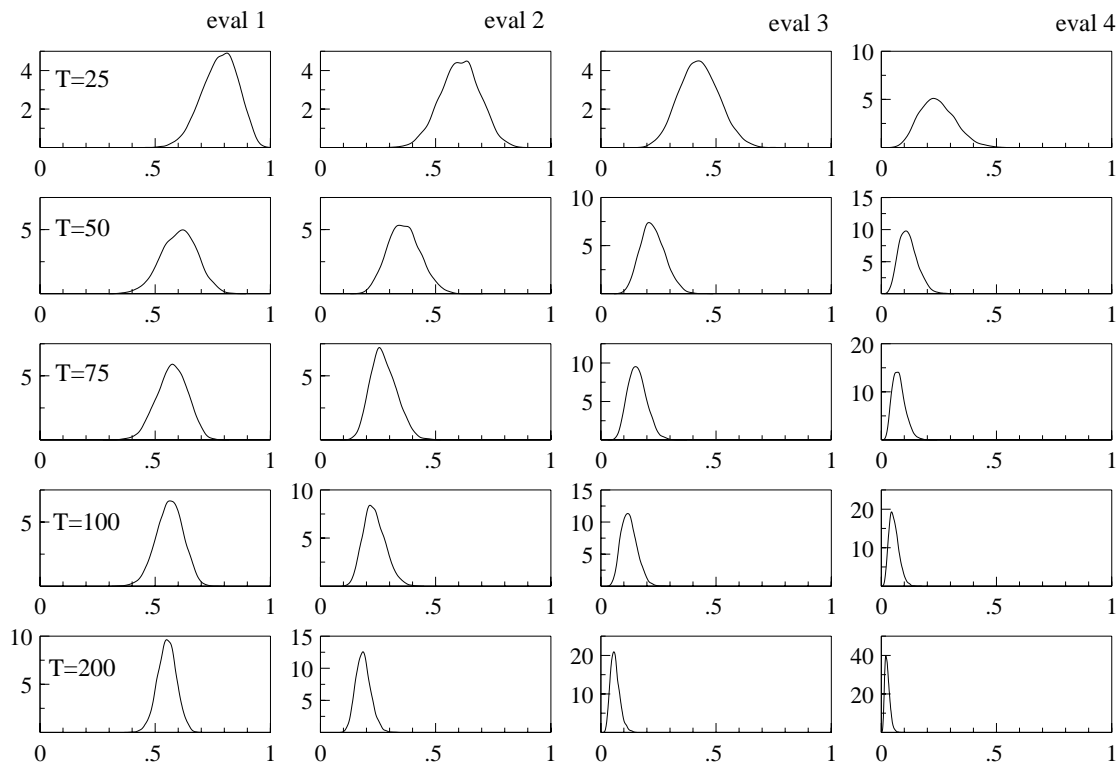


Figure 6 Recursive eigenvalues, re-estimating the short-run parameters.

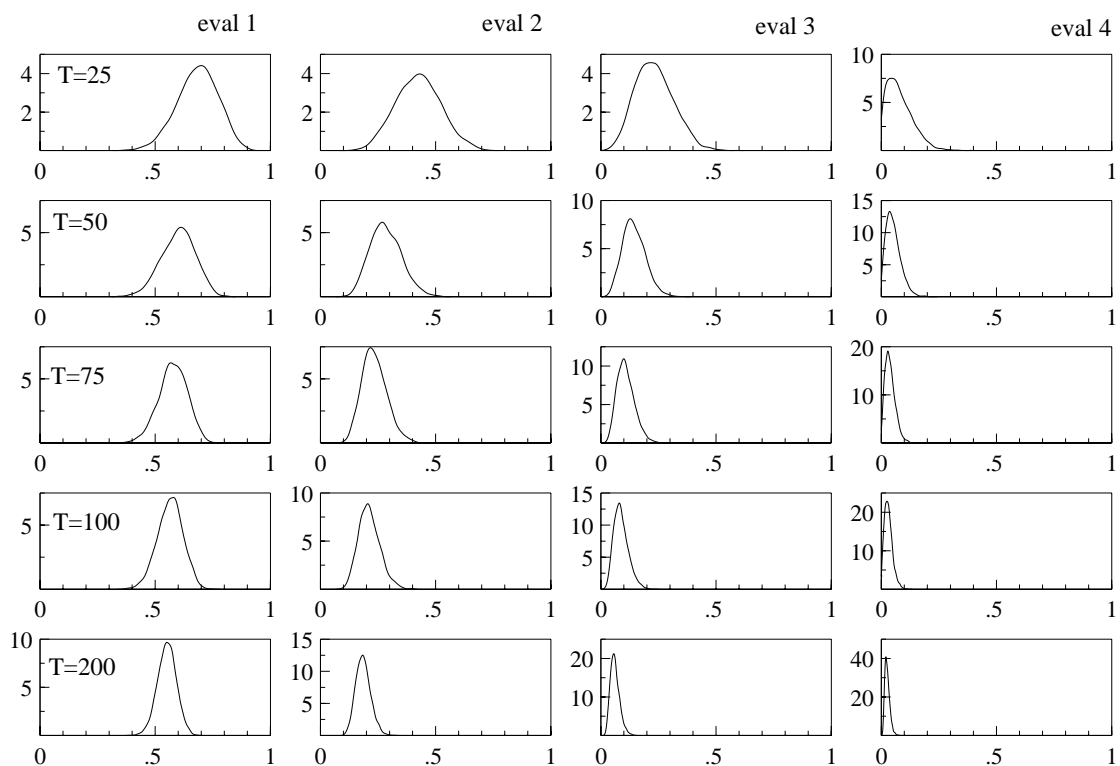


Figure 7 Recursive eigenvalues, conditioning on short-run parameters.

then doing the cointegration analysis conditional on the estimated short-run. Figures 6 and 7 plot the distributions of the eigenvalues for sample sizes  $T = 25, 50, 75, 100, 200$ , based on 5000 replications. In fig. 7 the eigenvalues are estimated conditionally on the short run at  $T = 200$ . In this case, the collapsing of the distribution of  $\mu_3$  and  $\mu_4$  on zero is already more pronounced at smaller sample sizes.

## 7.2 Approximating the asymptotic distribution

The asymptotic distributions of tests for cointegration rank involve integrals of Brownian motions for which no closed form expressions are available. To allow testing, tables have been based on simulation experiments.

Until recently, the most widely used tables were those constructed by Osterwald-Lenum (1992). Updated versions of these tables are in Johansen (1995b), computed with the DisCo program (Johansen and Nielsen, 1993). MacKinnon, Haug and Michelis (1996) note that a sample size of 400, on which these tables are based, is not so accurate for higher dimensions. These approaches are inconvenient for use in a computer program, either because they only list a few quantiles, or because they require a large amount of data.

Doornik (1998a) approximates the asymptotic distribution of cointegration tests using the Gamma distribution. Formulae for the parameters of the Gamma distributions are derived from response surfaces involving terms of  $O(T^{-1})$  (although this remains an asymptotic and not a finite-sample approximation). The resulting approximation works sufficiently well to replace the standard tables, and can provide quantiles as well as p-values. It is also easy to implement, requiring only 150 numbers to summarize the distributions for all test statistics (both for the  $I(1)$  and the  $I(2)$  model). All rejection frequencies and p-values in this paper are based on this approximation. Ox code implementing  $I(1)$  and  $I(2)$  cointegration tests and their p-values is available from the first author's web page.

## 7.3 Small-sample properties of the trace test

Whereas the previous section dealt with tabulating the asymptotic distribution, there has been some concern in the literature about the appropriateness of the asymptotic distribution in small samples. Figure 3 showed that in simulations the asymptotic distributions worked well in samples  $\geq 150$ , although there is some lack of power to find the second vector for the sample size we use ( $T \leq 100$ ).

Somewhat ad hoc small-sample corrections have been suggested by Reinsel and Ahn (1992) and Reimers (1992). This involves a so-called degrees-of-freedom correction, which entails scaling the test statistic by  $(1 - ns/T)$  where  $n$  is the dimension of the time series and  $s$  is the lag length. The effect of such a correction is that the acceptance region for the hypothesis  $H(r)$  is enlarged and therefore more cointegrating relations are found than when the correction is not used.

The degrees-of-freedom correction is not theoretically founded, and sometimes misleading. This is, for instance, the case for the very simplest situation: testing for a unit root in a univariate first-order autoregressive model. The exact distribution of the trace statistic is indistinguishable from the asymptotic distribution whenever the sample has more than seven observations, see Nielsen (1997a). In this situation the degrees of freedom correction  $(1 - 1/T)$  introduces a size distortion. When testing the hypothesis of no cointegration,  $H(0)$ , in a first-order model, the degrees-of-freedom correction is similarly over-correcting (Nielsen, 1997b).

In most practical situations, nuisance parameters are involved in the asymptotic distribution of the trace test. If the rank of  $\pi$  is  $r$  and the  $I(1)$  conditions are satisfied, then only one asymptotic distribution applies when testing  $H(r)$ . However, this does not ensure asymptotic similarity. On the boundary of the set given by these restrictions different distributions apply whenever the process has extra unit roots. The simplest example is the test for 'no cointegration',  $H(0)$ , in a univariate second-order model,  $n = 1$ ,  $s = 2$ . Under  $H(0)$ , the time series is given by:

$$\Delta x_t = \pi_1^* \Delta x_{t-1} + \varepsilon_t.$$

The  $I(1)$  conditions are satisfied if the first difference of the process is  $I(0)$ , that is  $|\pi_1^*| < 1$ . However, if  $\pi_1^* = 1$ , the first difference is  $I(1)$  and the process itself is  $I(2)$ . In this case, a different asymptotic distribution applies (see Pantula, 1989). Asymptotically, this is obviously not a problem; however, the small-sample distribution is some sort of a weighted average of the various asymptotic distributions. Nielsen (1997b) suggests a new asymptotic theory which includes the nuisance parameters continuously. This would give rather accurate approximations to the exact distribution at the expense of a more complicated asymptotic distribution.

## 8 Cointegration in the UK M1 model

### 8.1 Identification

When  $r > 1$ , as  $\alpha\beta' = \alpha\mathbf{H}\mathbf{H}^{-1}\beta'$  for all  $r \times r$  non-singular matrices  $\mathbf{H}$ , the cointegration vectors and feedback coefficients are not uniquely determined. One cannot compute standard errors of  $\hat{\beta}$  until it is identified, although by fixing  $\hat{\beta}$ , standard errors of  $\hat{\alpha}$  can be obtained, as shown in Table 4 for a fixed rank of two. Consequently, we next consider the use of restricted estimation of the parameters of  $\alpha\beta'$ ; this allows us to test over-identification and obtain standard errors for the over-identified parameters. The number of restrictions imposed in doing so may not match the degrees of freedom of the resulting tests as some restrictions are not binding, or can be ‘absorbed’ by changes elsewhere. Johansen (1995a) and Boswijk (1994) consider the identification conditions applicable to cointegration vectors and tests thereof. Doornik (1995a) considers identification under general, possibly non-linear, restrictions on the cointegration space.

### 8.2 Restricted cointegration analysis

To uniquely determine the two cointegration vectors, given their possible interpretations as excess demands for money and goods respectively as discussed in §3.2, we removed the trend from the first, and  $m - p$  from the second. Then we restricted the income coefficient to  $-1$  in the first vector (converting it to an inverse-velocity relation), and the trend coefficient in the second to the mean value of  $\Delta i$  (namely, 0.0063, approximately 2.5% p.a.). Finally, we set the feedbacks to zero for the second vector on the first equation, and the first on the last three equations (related to long-run weak exogeneity) which yields the results shown in Table 6, with the test of the restrictions being  $\chi^2(6) = 5.36$  [ $p = 0.5$ ].

Table 6 Restricted cointegration analysis.

$$\begin{bmatrix} \hat{\alpha} & 1 & 2 \\ m-p & -0.102 & 0 \\ & (0.011) & (-) \\ i & 0 & -0.149 \\ & (-) & (0.038) \\ \Delta p & 0 & 0.036 \\ & (-) & (0.023) \\ R_n & 0 & -0.040 \\ & (-) & (0.043) \end{bmatrix}, \begin{bmatrix} \hat{\beta}' & m-p & i & \Delta p & R_n & t \\ 1 & 1 & -1 & 6.41 & 7.16 & 0 \\ & (-) & (-) & (1.37) & (0.53) & (-) \\ 2 & 0 & 1 & -2.13 & 1.48 & -0.0063 \\ & (-) & (-) & (0.69) & (0.27) & (-) \end{bmatrix}$$

The first cointegration vector relates the ratio of money to expenditure ( $m - p - i$ ) negatively

to inflation and interest rates, corresponding an excess demand for transactions money. The second cointegration vector is interpretable as the excess demand for goods and services, being the deviation of expenditure from trend, negatively related to interest rates and positively to inflation; as its main influence is onto the  $i$  equation, we retain these two long-run relations.

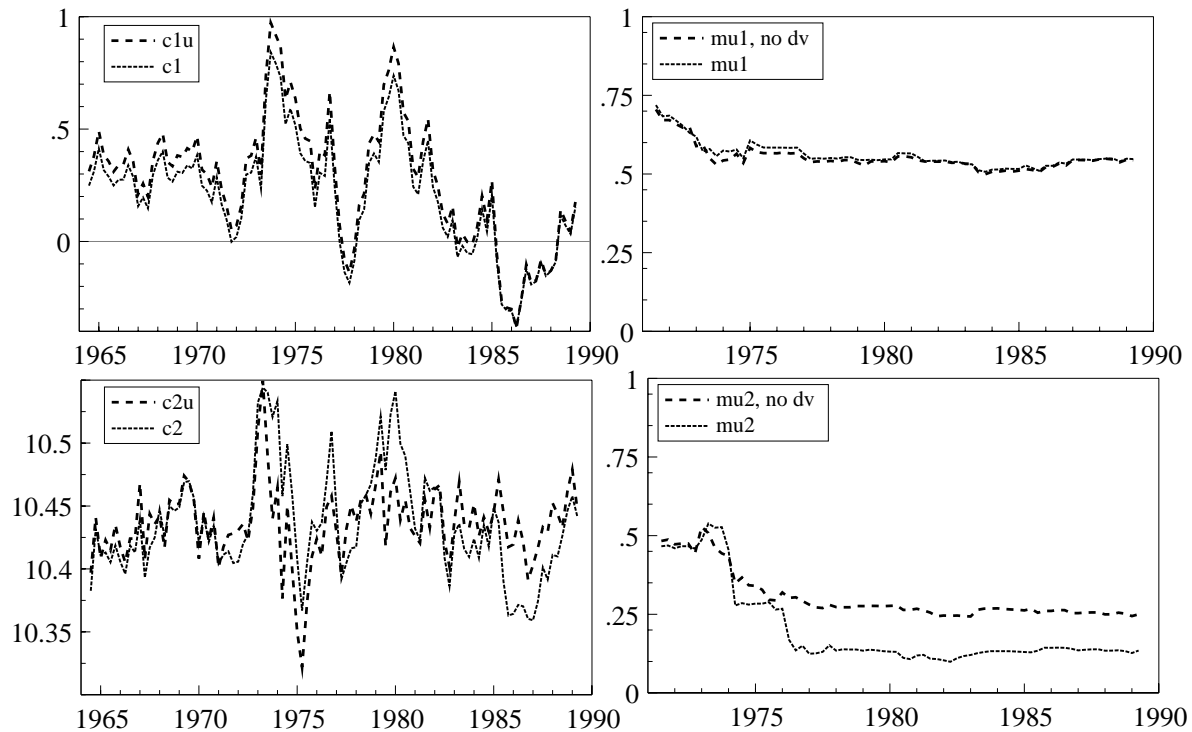


Figure 8 Cointegration vectors and recursively-computed eigenvalues.

Figure 8 records the time series of the two unrestricted ( $c_{iu}$ ) and restricted ( $c_i$ ) cointegration vectors, and the associated recursively-computed eigenvalues discussed above. The unrestricted and restricted vectors for money are similar, whereas the restrictions have somewhat altered the second. Thus, subtracting their in-sample means, the two, zero-mean,  $l(0)$  linear combinations defining the equilibrium-correction mechanisms (EqCMs) are:

$$c_{1,t} = m_t - p_t - i_t + 6.41\Delta p_t + 7.16R_{n,t} - 0.209 \quad (19)$$

and:

$$c_{2,t} = (i_t - 0.0063t) - 2.13\Delta p_t + 1.48R_{n,t} - 11.186. \quad (20)$$

The definitions in (19) and (20) are required for multi-step forecasts when formulating the model in terms of the differences ( $\Delta(m-p)_t$ ,  $\Delta i_t$ ,  $\Delta^2 p_t$ ,  $\Delta R_{n,t}$ ) of the original variables.

## 9 $l(2)$ analysis

The UK money data have been analysed in  $l(2)$  models by Johansen (1992a), Paroulo (1996) and Rahbek *et al.* (1998). The data  $m$ ,  $p$ ,  $i$ ,  $R_n$  were analysed in levels using a fifth-order VAR. The last two papers conclude that there are two  $l(2)$  trends and one stationary polynomial cointegrating relation. To compare that result with the  $l(1)$  analysis reported above, we now return to analyse the original measures,  $m$ ,  $p$ ,  $i$ , and  $R_n$ . Should the analysis commence in  $l(2)$  space, then  $\alpha'_\perp \Gamma \beta_\perp = \gamma \delta'$  is also reduced rank, so some

linear combinations first cointegrate from  $I(2)$  to  $I(1)$  and then  $r$  others (perhaps with  $I(1)$  differences of  $I(2)$  variables) cointegrate to  $I(0)$ . Thus, both  $I(2)$  and  $I(1)$  impose reduced rank restrictions on the initial formulation in (10), and the former imposes restrictions on (11).

Empirically, Rahbek *et al.* (1998) find that the cointegration relation has the form:

$$C_1^* = C_1 + 11.63\Delta(m - p), \quad (21)$$

where:

$$C_1 = m - p - i + 6.13\Delta p + 7.01R.$$

Thus, the change in real money is needed to establish cointegration in their analysis, so real money is  $I(2)$ . Hence, either the nominal magnitudes are one degree of integration higher, or nominal money and prices do not cointegrate. Neither implication is very palatable, nor can we interpret (21) easily in terms of the possible plans of economic agents. Conversely, adding  $11.63\Delta(m - p)$  to  $C_1$  makes it look considerably more stationary as fig. 9b shows. Note that the cointegrating relations found in Hendry and Ericsson (1991), Hendry and Doornik (1994),  $C_1$ , and  $c_1$  in (19) are all very similar. Figure 9a records the original and the negative of the additional correction (without any graphical matching), and fig. 9d records the cross plot: as can be seen, there is an almost perfect offset. Figure 9c shows the first difference of  $C_1$ , which would be  $I(0)$  if the original was  $I(1)$ .

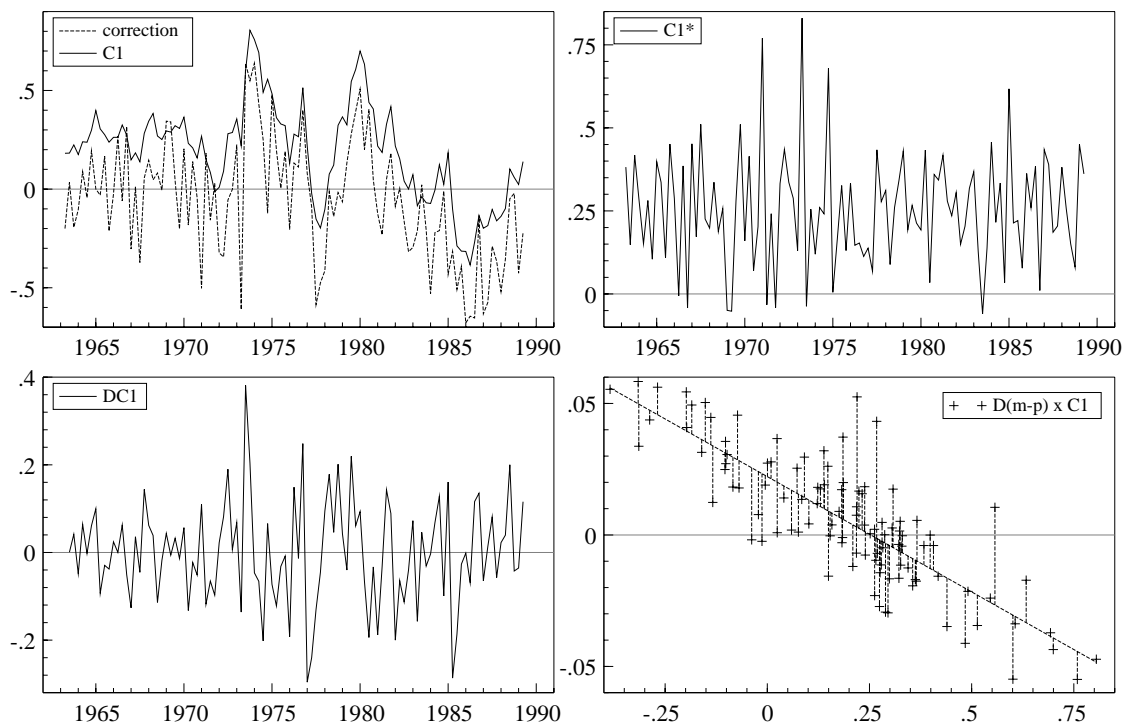


Figure 9 Comparisons of alternative cointegration measures.

The coefficient of 11.63 times a feedback of 0.093 is almost precisely unity, hence in equation (6) of Hendry and Ericsson (1991):

$$\Delta(m - p)_t = -0.17\Delta(m - p)_{t-1} - 0.093C_{1,t-1} + \dots, \quad (22)$$

or:

$$\begin{aligned} \Delta^2(m - p)_t &= -1.17\Delta(m - p)_{t-1} - 0.093C_{1,t-1} + \dots \\ &= -0.09\Delta(m - p)_{t-1} - 0.093C_{1,t-1}^* + \dots, \end{aligned} \quad (23)$$

so the  $I(1)$  versus  $I(2)$  decision turns on the significance of  $\Delta(m-p)_{t-1}$ . Estimating (23) by OLS reveals a very small effect, even though it was significant in (22). Interestingly, on ADF tests  $\Delta(m-p)_t$ ,  $\Delta R_{n,t}$  and  $\Delta^2 p_t$  are all  $I(0)$  (i.e., the tests reject unit roots at 1% or better), but do not reject for  $C_{1,t}$ . As (23) shows, if one keeps  $C_{1,t}$ , then the coefficient on  $\Delta(m-p)_{t-1}$  is less than  $-1$ , and if  $\Delta(m-p)_t$  is  $I(0)$ , then  $\Delta^2(m-p)_t$  is  $I(-1)$ , so  $\Delta(m-p)_t$  cannot matter for cointegration.

In the  $I(2)$  analysis, there is also some evidence for a second cointegrating relation. Rahbek *et al.* (1998) find that the asymptotic probability of rejecting the hypothesis of at most one cointegrating relation, and thereby accepting two such relations, is around 15%. Their second relation has the form:

$$C_2^* = C_2 + 0.38\Delta(m-p),$$

where the coefficients of  $C_2$  are very similar to those reported in (20). The coefficient of the (possibly)  $I(1)$  component, the first differences of real money, is very small as compared with that in the money relation (21): however, the second relation,  $C_2$ , appears to be more stable than the first,  $C_1$ ,

The economic analysis in §3.2 suggested that it was more convenient to base an econometric analysis on the variables  $m-p$ ,  $\Delta p$ ,  $i$ , and  $R_n$ . An  $I(2)$  analysis of these data confirms this. There is one  $I(2)$  trend and one (possibly two) cointegrating relations. The  $p$ -values and coefficients are hardly changed from the  $I(1)$  analysis. When dummies are included, as discussed in §6, the conclusions are similar to those in Hendry and Doornik (1994). This analysis was based on the extension by Jørgensen (1996) to the program CATS by Hansen and Juselius (1994), and also using an Ox routine for testing  $I(2)$ -ness (Doornik, 1998b).

## 10 Long-run weak exogeneity and conditional systems

In a conditional analysis, namely of a system with unmodelled variables, there is a tendency to find more cointegrating relations than in the corresponding closed-system analysis. For the UK money data, this was found in the analysis of Harbo, Johansen, Nielsen and Rahbek (1998).

The joint likelihood of  $\mathbf{x}_t = (\mathbf{y}_t, \mathbf{z}_t)$  can always be factorised into the product of the conditional likelihood of  $\mathbf{y}_t$  given  $\mathbf{z}_t$ , and the marginal likelihood of  $\mathbf{z}_t$ . When the parameters of the two likelihood functions vary freely,  $\mathbf{z}_t$  is said to be weakly exogenous for the parameters of the conditional likelihood, and, in particular, it is equivalent to maximize the two likelihood functions jointly or separately (see Engle, Hendry and Richard, 1983). For cointegration analysis, Johansen (1995b, Ch. 8) finds that  $\mathbf{z}_t$  is weakly exogenous for the cointegration parameter if the adjustment parameter has the form  $\alpha' = (\alpha'_y, \mathbf{0})$ .

The conditional likelihood can be analysed for cointegration using the ‘partial’ systems approach of Harbo *et al.* (1998). They consider an analysis of  $m-p$  conditional on  $(i, \Delta p, R_n)$  assuming the rank is at most unity, as well as  $(m-p, \Delta p)$  conditional on  $(i, R_n)$  when the rank is at most 2. The dummies are omitted in their analysis, and a lag length of five is chosen. The results are therefore slightly different from those of this paper and Hendry and Doornik (1994).

To test for long-run weak exogeneity in the two-variable conditional model requires testing if  $\alpha$  is of the form:

$$\begin{matrix} m-p \\ i \\ \Delta p \\ R_n \end{matrix} \begin{pmatrix} * & * \\ 0 & 0 \\ * & * \\ 0 & 0 \end{pmatrix}.$$

Referring back to Table 6, we see a highly significant feedback coefficient for the second cointegrating vector in the  $i$  equation, rejecting this hypothesis in the restricted model. In the unrestricted rank-two



model of Tables 3–4, we also reject this form of long-run weak exogeneity:  $\chi^2(4) = 10.9^*$  [0.03]. The same result (to the reported accuracy) is obtained when we apply the mis-specification test suggested in Harbo *et al.* (1998). In that case, the significance of the cointegration vectors is tested in a regression of  $\Delta i_t$  and  $\Delta R_{nt}$  on  $\Delta(m-p)_{t-1}$ ,  $\Delta^2 p_{t-1}$ ,  $\Delta i_{t-1}$ ,  $\Delta R_{n,t-1}$ ,  $c_{1,t-1}$ ,  $c_{2,t-1}$ ,  $1, t, doil, dout$ . When using five lags and without dummies (as in Harbo *et al.*, 1998), the p-value of the test increases to about 30%, and the hypothesis is not rejected.

Long-run weak exogeneity is easily accepted for  $R_n$ . The results for the model conditional on  $R_n$  are given in Table 7. Now the critical values are changed (the p-values are derived using the Gamma approximation of Doornik, 1998a), and there is slightly stronger support for the hypothesis of two cointegrating vectors.

Table 7 Cointegration analysis, conditional on  $R_n$ .

$r$	0	1	2	3
$\ell$	1363.0	1401.7	1408.6	1412.7
$\mu$		0.55	0.13	0.08
$H_l(r)$	$r = 0$	$r \leq 1$	$r \leq 2$	
Trace	101.4**	22.2	8.4	
	[0.00]	[0.35]	[0.43]	

A valid conditional cointegration analysis should give higher power in determining the cointegration rank. A power analysis can explain this finding. To simplify the argument, consider a three-dimensional DGP such that :

$$\begin{aligned}\Delta x_{1,t} &= -\frac{b}{T}x_{1,t-1} + \nu_{1,t}, \\ \Delta x_{i,t} &= \nu_{i,t} \quad \text{for } i = 2, 3.\end{aligned}$$

The asymptotic power of the test for no cointegration in the full system is derived by Johansen (1995b, Chapter 14). The power for the conditional analysis of the first component, given the last two, can be derived correspondingly: also see Phillips (1988). Figure 10 shows that the asymptotic power of the test in a conditional system is considerably higher than for the closed-system test, which explains the tendency to find more cointegrating relations. Table 6 gives the adjustment coefficient for the first cointegrating vector as approximately 0.1. The corresponding value of  $b$  is 10 when  $T = 100$ .

## 11 Conclusions

The determination of cointegration rank remains a subtle task, dependent on a number of modelling considerations. The introduction listed many of these, and the paper concentrated on seven issues.

First, the treatment of the constant and a deterministic trend in the cointegration formulation. A range of possible models was evaluated analytically and by Monte Carlo simulation. We found that including an unrestricted trend was problematic, and could lead to excess rejection for ranks above that in the DGP. However, a restricted trend in the cointegration space with an unrestricted constant produced good power, and reasonable size. Moreover, the known asymptotic similarity of this case extended to finite samples in our Monte Carlo, and the same formulation worked well even when the DGP was a special case with no trend, and even no drift. Thus, we recommend commencing the analysis with

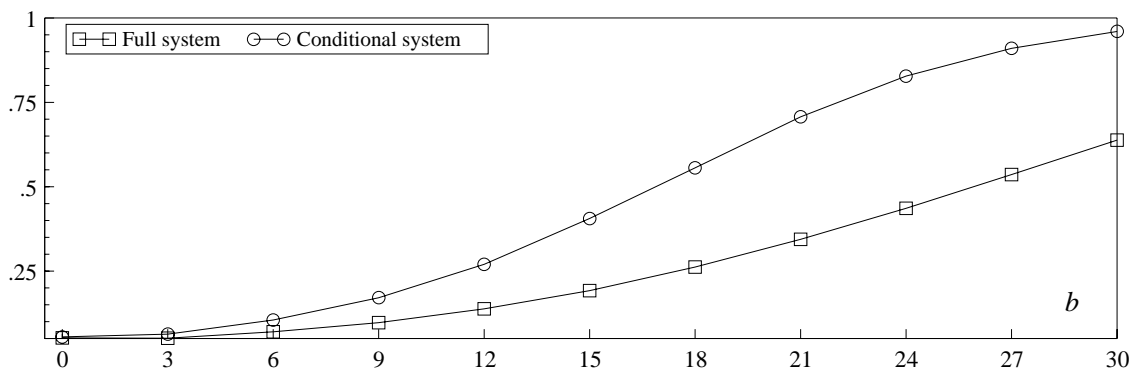


Figure 10 Asymptotic power functions of the test for no cointegration at 5% in a closed and conditional system,  $T = 400$ ,  $M = 10\,000$ .

a linear trend restricted to the cointegration space and an unrestricted constant to ensure (asymptotic) similarity to the nuisance parameters of these effects.

Next, concerning the choice of test statistic, there is a consistent rank-selection procedure for the trace test, even for the extended set of models in §4.3. However, for the maximum eigenvalue test, the consistency of the rank-selection procedure has not been established. For the model formulations  $H_l(r)$  in Table 5, trace tests are asymptotically similar with respect to the values taken by the parameters of the deterministic components (assuming the  $l(1)$  conditions are satisfied): see Nielsen and Rahbek (1998). As this does not apply to the extended models, we recommend the sequential trace test based on  $H_l(r)$ .

Third, the asymptotic distributions of the test statistics under the null can be well approximated by response surfaces for most case of practical concern. However, their adequacy as a guide to finite samples depends on the unknown values of various nuisance parameters, particularly the presence of roots close to making the process  $l(2)$ . Existing finite-sample corrections based on degrees of freedom need not work well, and often over-correct.

Concerning the treatment of impulse indicator variables, we strongly recommend that these be entered unrestrictedly if they are used to establish an estimate of the innovation variance. We certainly advise against their restriction to the cointegration space, and the Appendix suggested an explanation for the results in Hendry and Doornik (1994). However, our analysis is incomplete, and requires extension to other forms of dummy (such as step changes). Also, further analysis is needed of including the integral of a dummy in the cointegration space with the original unrestricted, as some forms of invariance, or similarity, may be present, although new critical values will be needed.

We did not address the selection of lag length in this paper.

Restricted cointegration analysis to estimate identified cointegration relations is straightforward to implement, and yielded interpretable results for our empirical example of UK M1. This allowed standard errors of the cointegration and the feedback coefficients to be calculated. Nevertheless, formal identification depends on having an adequate number of restrictions to impose.

Considerable doubt remains as to whether the UK M1 data are  $l(1)$  or  $l(2)$ . We first analyzed them as  $l(1)$  in the space of real money, real expenditure, inflation and nominal net interest rates, and in earlier research developed a congruent model of the four series. However, when  $(m, p, i, R_n)$  are treated as  $l(2)$ , two  $l(2)$  components have been found, and consistent with that result, when  $(m-p, \Delta p, i, R_n)$  are treated as  $l(2)$ , one  $l(2)$  component remains. Thus, reduction to  $l(0)$  involves polynomial cointegration, with double differencing of  $m-p$ . Nevertheless, the resulting model is essentially isomorphic to the earlier specification – a plus for focusing on developing congruent representations.

Finally, we considered recent research on modelling conditional, or ‘partial’ systems, and noted the care needed to ensure coherent inference. If possible, it seems advisable to first model the complete system, test for long-run weak exogeneity, and only model the conditional system if long-run weak exogeneity is not rejected. This may increase the power of the trace test.

## References

- Akerlof, G. A. (1973). The demand for money: A general equilibrium inventory-theoretic approach. *Review of Economic Studies*, **40**, 115–130.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.
- Boswijk, H. P. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.
- Boswijk, H. P. (1994). Identifiability of cointegrated systems. Mimeo, Department of Actuarial Sciences and Econometrics, University of Amsterdam.
- Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, **17**, 334–355.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.
- Doornik, J. A. (1995a). Testing general restrictions on the cointegrating space. Mimeo, Nuffield College.
- Doornik, J. A. (1995b). Testing vector autocorrelation and heteroscedasticity in dynamic models. Mimeo, Nuffield College.
- Doornik, J. A. (1998a). Approximations to the asymptotic distribution of cointegration tests. Mimeo, Nuffield College, University of Oxford.
- Doornik, J. A. (1998b). *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Press.
- Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.
- 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. (1997). *Modelling Dynamic Systems using PcFiml 9 for Windows*. London: Timberlake Consultants Press.
- 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. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Ericsson, N. R., Campos, J., and Tran, H.-A. (1990). PC-GIVE and David Hendry’s econometric methodology. *Revista De Econometria*, **10**, 7–117.
- Ericsson, N. R., Hendry, D. F., and Tran, H.-A. (1994). Cointegration, seasonality, encompassing and the demand for money in the United Kingdom. In Hargreaves, C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 179–224. Oxford: Oxford University Press.
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.

- Hamilton, J. D. (1994). *Time series Analysis*. Princeton: Princeton University Press.
- Hansen, H., and Johansen, S. (1992). Recursive estimation in cointegrated VAR-models. Discussion paper, Institute of Mathematical Statistics, University of Copenhagen.
- Hansen, H., and Johansen, S. (1998). Some tests for parameter constancy in cointegrated VAR-models. Mimeo, Economics Department, European University Institute.
- Hansen, H., and Juselius, K. (1994). Manual to cointegration analysis of time series CATS in RATS. Discussion paper, Institute of Economics, University of Copenhagen.
- Hansen, P. R., and Johansen, S. (1998). *Workbook on Cointegration*. Oxford: Oxford University Press.
- Harbo, I., Johansen, S., Nielsen, B., and Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. Discussion paper. Forthcoming.
- Harris, R. I. D. (1995). *Using Cointegration Analysis in Econometric Modelling*. London: Prentice Hall.
- Hatanaka, M. (1996). *Time-Series-Based Econometrics: Unit Roots and Cointegration*. Oxford: Oxford University Press.
- Hendry, D. F. (1979). Predictive failure and econometric modelling in macro-economics: The transactions demand for money. In Ormerod, P. (ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F. (1985). Monetary economic myth and econometric reality. *Oxford Review of Economic Policy*, **1**, 72–84. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F., and Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.
- Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.
- Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.
- Hendry, D. F., Neale, A. J., and Ericsson, N. R. (1991). *PC-NAIVE, An Interactive Program for Monte Carlo Experimentation in Econometrics. Version 6.0*. Oxford: Institute of Economics and Statistics, University of Oxford.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1992a). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.
- Johansen, S. (1992b). Testing weak exogeneity and the order of cointegration in UK money demand. *Journal of Policy Modeling*, **14**, 313–334.
- Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, **13**, 205–229.
- Johansen, S. (1995a). Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of Econometrics*, **69**, 111–132.

- Johansen, S. (1995b). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Johansen, S., and Nielsen, B. (1993). Asymptotics for cointegration rank tests in the presence of intervention dummies - manual for the simulation program DisCo. <http://www.math.ku.dk/sjo>, University of Copenhagen, Institute of Mathematical Statistics, Copenhagen.
- Jørgensen, C. (1996). I(2) cointegration procedure for CATS. <http://www.estima.com/procs/i2index.htm>, University of Copenhagen, Institute of Mathematical Statistics, Copenhagen.
- MacKinnon, J. G., Haug, A. A., and Michelis, L. (1996). Numerical distribution functions of likelihood ratio tests for cointegration. Mimeo, Queens University, Canada.
- Milbourne, R. (1983). Optimal money holding under uncertainty. *International Economic Review*, **24**, 685–698.
- Miller, M. H., and Orr, D. (1966). A model of the demand for money by firms. *Quarterly Journal of Economics*, **80**, 735–759.
- Nielsen, B. (1997a). Bartlett correction of the unit root test in autoregressive models. *Biometrika*, **84**, 500–504.
- Nielsen, B. (1997b). On the distribution of tests for cointegration rank. Discussion paper, Nuffield College, Oxford.
- Nielsen, B., and Rahbek, A. (1998). Similarity issues in cointegration analysis. Mimeo, Nuffield College, Oxford.
- Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the ML cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, **54**, 461–472.
- Pantula, S. G. (1989). Testing for unit roots in time series data. *Econometric Theory*, **5**, 256–271.
- Paroulo, P. (1996). On the determination of integration indices in I(2) systems. *Journal of Econometrics*, **72**, 313–356.
- Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, **56**, 1021–1043.
- Phillips, P. C. B., and Solo, V. (1992). Asymptotics of linear processes. *Annals of Statistics*, **20**, 971–1001.
- Rahbek, A., Kongsted, H. C., and Jørgensen, C. (1998). Trend-stationarity in the I(2) cointegration model. Forthcoming, *Journal of Econometrics*.
- Rahbek, A., and Mosconi, R. (1998). The role of stationary regressors in the cointegration test. Mimeo 1998, University of Copenhagen, Institute of Mathematical Statistics, Copenhagen.
- Reimers, H.-E. (1992). Comparisons of tests for multivariate cointegration. *Statistical Papers*, **33**, 335–359.
- Reinsel, G. C., and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure, estimation, likelihood ratio tests, and forecasting. *Journal of Time Series Analysis*, **13**, 353–375.
- Smith, G. W. (1986). A dynamic Baumol-Tobin model of money demand. *Review of Economic Studies*, **53**, 465–469.
- Tobin, J. (1958). Liquidity preference as behavior toward risk. *Review of Economic Studies*, **25**, 65–86.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

## 12 Appendix: Some Analytical Results

### 12.1 Illustration of the models of trend behaviour

Test for  $H_{ql}(r)$  against  $H_{ql}(n)$  when the DGP is  $H_l(s)$

When  $\alpha'_\perp \phi = \mathbf{0}$ , so the assumption  $\alpha'_\perp \phi \neq \mathbf{0}$  is not satisfied, the standard asymptotic distribution does not apply. Instead, if  $r < \text{rank}(\boldsymbol{\pi})$ , then the power converges to unity; otherwise ( $r \geq \text{rank}(\boldsymbol{\pi})$ ), the limit distribution of the trace test is given by the distribution of the sum of the  $n - r$  smallest eigenvalues of:

$$\int_0^1 d\mathbf{B}_u \mathbf{F}'_u \left( \int_0^1 \mathbf{F}_u \mathbf{F}'_u du \right) \int_0^1 \mathbf{F}_u d\mathbf{B}'_u. \quad (24)$$

Here  $\mathbf{B}_u$  is an  $n - \text{rank}(\boldsymbol{\pi})$  dimensional standard Brownian motion, and  $\mathbf{F}_u$  is  $\mathbf{B}_u$  corrected for a linear trend and a constant.

Simulations indicate that when  $r = \text{rank}(\boldsymbol{\pi})$ , the distribution in (24) is stochastically larger than that for  $\alpha'_\perp \phi \neq \mathbf{0}$  (given in Johansen, 1995b, Theorem 11.1 and Table 15.5). On the other hand, the distribution is smaller than that which applies when testing  $H_l(r)$  in  $H_l(n)$  (*op. cit.*, Table 15.4).

Test for  $H_{lc}(r)$  against  $H_{lc}(n)$  when the DGP is  $H_l(s)$

The power of this mis-specified test converges to unity if  $r < \text{rank}(\boldsymbol{\pi}) - 1$ , and to a constant less than unity if  $r \geq \text{rank}(\boldsymbol{\pi}) - 1$ , provided the  $I(1)$  conditions are satisfied. The constant depends on a number of nuisance parameters, which we illustrate for a bivariate case.

Consider a bivariate DGP with one cointegrating vector and a restricted trend:

$$\Delta \mathbf{x}_t = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} \phi \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \rho \end{pmatrix} t + \boldsymbol{\nu}_t \quad \text{where } \boldsymbol{\nu}_t \sim \text{IN}_2[\mathbf{0}, \mathbf{I}].$$

The model  $H_{lc}$  is analysed by a two-step procedure. First,  $\Delta \mathbf{x}_t$  and  $\mathbf{x}_{t-1}$  are corrected for the intercept, giving the residuals  $\mathbf{R}_0$  and  $\mathbf{R}_1$ :

$$\begin{aligned} \mathbf{R}_0 &= \Delta \mathbf{x}_t |_1 = \begin{cases} \nu_{1t} - \bar{\nu}_{1t}, \\ \Delta \nu_{2t} - \Delta \bar{\nu}_{2t}, \end{cases} \\ \mathbf{R}_1 &= \mathbf{x}_{t-1} |_1 = \begin{cases} \phi(t - \bar{t}) + \varsigma_{t-1} - \bar{\varsigma}, \\ \rho(t - \bar{t}) + \nu_{2t} - \bar{\nu}_{2t}. \end{cases} \end{aligned}$$

where  $\varsigma_t = \sum_{i=1}^t \nu_{1t}$ . Next, the empirical canonical correlations of  $\mathbf{R}_0$  and  $\mathbf{R}_1$  are found as the solutions to the eigenvalue problem:

$$|\lambda \mathbf{S}_{00} - \mathbf{S}_{01} \mathbf{S}_{11}^{-1} \mathbf{S}_{10}| = 0, \quad (25)$$

where  $\mathbf{S}_{ij} = T^{-1} \mathbf{R}'_i \mathbf{R}_j$  (the product-moment matrices of the residuals).

For the asymptotic analysis, let  $B_u$  be a univariate standard Brownian motion and let:

$$\bar{B} = \int_0^1 B_u du.$$

Further, for  $\rho \neq 0$ , define:

$$\mathbf{A} = \begin{pmatrix} (\phi^2 + \rho^2) \sqrt{T} & 0 \\ 0 & -\rho \end{pmatrix}^{-1} \begin{pmatrix} \phi & \rho \\ -\rho & \phi \end{pmatrix},$$

so that:

$$\mathbf{A} \mathbf{R}_1 = \begin{cases} \frac{1}{\sqrt{T}} (t - \bar{t}) + \frac{\phi}{\sqrt{T}(\phi^2 + \rho^2)} (\varsigma_{t-1} - \bar{\varsigma}) + \frac{\rho}{\sqrt{T}(\phi^2 + \rho^2)} (\nu_{2t} - \bar{\nu}_{2t}), \\ (\varsigma_{t-1} - \bar{\varsigma}) - \frac{\phi}{\rho} (\nu_{2t} - \bar{\nu}_{2t}). \end{cases}$$

and similarly for  $\mathbf{AR}_0$ .

Then, by the analysis of linear processes in Phillips and Solo (1992):

$$\begin{aligned} T^{-1} \mathbf{AS}_{11} \mathbf{A}' &\xrightarrow{\mathcal{D}} \int_0^1 \begin{pmatrix} (u - \frac{1}{2}) \\ (B_u - \bar{B}) \end{pmatrix} \begin{pmatrix} (u - \frac{1}{2}) \\ (B_u - \bar{B}) \end{pmatrix}' du, \\ \mathbf{AS}_{10} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 (u - \frac{1}{2}) dB_u & 0 \\ \int_0^1 (B_u - \bar{B}) dB_u & \phi/\rho \end{pmatrix}, \\ \mathbf{S}_{00} &\xrightarrow{\mathcal{P}} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \end{aligned}$$

since:

$$\begin{aligned} T^{-3} \sum (t - \bar{t})^2 &\rightarrow \int (u - \frac{1}{2})^2 du, & \sum (t - \bar{t}) (\nu_{2t} - \bar{\nu}_{2t}) &= O_p(T^{3/2}), \\ T^{-5/2} \sum (t - \bar{t}) (\varsigma_{t-1} - \bar{\varsigma}) &\rightarrow \int (u - \frac{1}{2}) (B_u - \bar{B}) du, & \sum (\varsigma_{t-1} - \bar{\varsigma}) (\nu_{2t} - \bar{\nu}_{2t}) &= O_p(T), \\ T^{-2} \sum (\varsigma_{t-1} - \bar{\varsigma})^2 &\rightarrow \int (B_u - \bar{B})^2 du, & \sum (\nu_{2t} - \bar{\nu}_{2t})^2 &= O_p(T). \end{aligned}$$

and

$$\begin{aligned} T^{-3/2} \sum (t - \bar{t}) \nu_{1t} &\rightarrow \int (u - \frac{1}{2}) dB_u, & \sum (t - \bar{t}) \Delta \nu_{2t} &= O_p(T), \\ T^{-1} \sum (\varsigma_{t-1} - \bar{\varsigma}) \nu_{1t} &\rightarrow \int (B_u - \bar{B}) dB_u, & \sum (\varsigma_{t-1} - \bar{\varsigma}) \Delta \nu_{2t} &= O_p(T^{1/2}), \\ \sum (\nu_{2t} - \bar{\nu}_{2t}) \nu_{1t} &= O_p(T^{1/2}), & \sum (\nu_{2t} - \bar{\nu}_{2t}) \Delta \nu_{2t} &= -T + O_p(T^{1/2}). \end{aligned}$$

By substituting these asymptotic results into the eigenvalue problem (25), it can be shown that both eigenvalues converge to zero at a rate of  $1/T$ . So when testing  $\mathbf{H}_{lc}(0)$ , the power does not converge to unity, but to a constant which is smaller than unity: compare with the power function for the test of  $\mathbf{H}_{lc}(1)$  in fig. 3. Moreover, the asymptotic distribution of  $T\lambda_1, T\lambda_2$  depends on the ratio  $\phi/\rho$ .

Figure 3 illustrates that, for certain values of  $\phi$ , the rejection frequency of the test of  $\mathbf{H}_{lc}(\text{rank}\pi - 1)$  could converge to a level which is larger than the nominal size. This is not generally the case for all values of  $\phi$ . If actually  $\phi = 0$  and  $\rho \neq 0$ , then the rejection frequency for the test of  $\mathbf{H}_{lc}(0)$  against  $\mathbf{H}_{lc}(2)$  converges to zero. The reason is that the smallest eigenvalue converges to zero at rate  $1/T^2$ , and the largest eigenvalue, normalized by  $T$ , converges to the asymptotic distribution of the test for  $\mathbf{H}_c(0)$  against  $\mathbf{H}_c(1)$  in a well-specified model. So, the test for  $\mathbf{H}_{lc}(0)$  against  $\mathbf{H}_{lc}(2)$  for the above DGP has the same asymptotic distribution, which is tabulated as the entry for one degree of freedom in Johansen (1995b, Table 15.1). The postulated result is then found by comparison of this distribution with the standard asymptotic distribution for the test for  $\mathbf{H}_{lc}(0)$  against  $\mathbf{H}_{lc}(2)$  (*op. cit.*, Table 15.3).

*Test for  $\mathbf{H}_c(r)$  against  $\mathbf{H}_c(n)$  when the DGP is  $\mathbf{H}_l(s)$*

The first asymptotic observation concerning this test is that the power converges to unity if  $r < \text{rank}(\pi)$  and to a constant less than unity otherwise, provided that the  $l(1)$  conditions are satisfied. However, for certain values of the nuisance parameter, the convergence of the test for  $\mathbf{H}_c(\text{rank}\pi)$  is rather slow. This feature is seen in fig. 3 for the test of  $\mathbf{H}_c(1)$ . The DGP of the previous argument is considered again.

The statistical analysis is based on the squared empirical canonical correlations of  $\Delta \mathbf{x}_t$  and the vector of  $\mathbf{x}_{t-1}$  and 1. Because of invariance, the latter vector can be replaced by that of  $\mathbf{x}_{t-1}$  corrected for 1, and 1. The new residual matrices are denoted  $\mathbf{S}_{ij}^*$ . Now, define  $\mathbf{s}_0$  as:

$$\mathbf{s}_0 = \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{x}_t = \begin{pmatrix} \phi \\ \rho \end{pmatrix} + \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \nu_{1,t} \\ \Delta \nu_{2,t} \end{pmatrix} = \begin{pmatrix} \phi \\ \rho \end{pmatrix} + O_p(1/\sqrt{T}),$$

then:

$$\mathbf{S}_{11}^* = \begin{pmatrix} \mathbf{S}_{11} & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{S}_{10}^* = \begin{pmatrix} \mathbf{S}_{10} \\ \mathbf{s}'_0 \end{pmatrix}, \quad \mathbf{S}_{00}^* = \mathbf{S}_{00} + \mathbf{s}_0 \mathbf{s}'_0,$$

so the eigenvalue problem is given by:

$$\begin{aligned} 0 &= |\lambda \mathbf{S}_{00}^* - \mathbf{S}_{01}^* \mathbf{S}_{11}^{*-1} \mathbf{S}_{10}^*| \\ &= |(\lambda - 1) \mathbf{s}_0 \mathbf{s}'_0 + \lambda \mathbf{S}_{00} - \mathbf{S}_{01} \mathbf{S}_{11}^{-1} \mathbf{S}_{10}|. \end{aligned}$$

Thus, one eigenvalue converges to a non-zero constant, whereas the smallest, normalized by  $T$ , converges to a distribution involving various nuisance parameters. It also follows that, if  $\mathbf{s}_0 \mathbf{s}'_0$  is small relative to  $\mathbf{S}_{00}$ , then the convergence of the largest eigenvalue could be rather slow.

A corresponding result holds for general processes as in (10). A more detailed representation of the process than that given by Johansen (1995b, Theorem 4.2) yields:

$$\mathbf{s}_0 \xrightarrow{\mathcal{P}} \mathbf{C} \boldsymbol{\phi} - (\mathbf{C} \boldsymbol{\Gamma} + \mathbf{I}_r) \boldsymbol{\beta} (\boldsymbol{\beta}' \boldsymbol{\beta})^{-1} \boldsymbol{\rho},$$

where  $\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}'_\perp$ , and, for a process with two lags,  $\boldsymbol{\Gamma} = -(\mathbf{I}_n - \boldsymbol{\Pi} + \boldsymbol{\pi})$ : see Appendix A of Rahbek and Mosconi (1998). For the DGP (17), both  $\boldsymbol{\rho}_l$  and  $\boldsymbol{\alpha}'_\perp \boldsymbol{\phi}_c$  are rather small.

## 12.2 Dummies

### 12.2.1 Unit-root test in a model with an unrestricted dummy

It is of interest to find out whether the distribution of the test depends on the parameter of the dummy. From the following analysis, it is concluded that the asymptotic distribution does not depend on a nuisance parameter unless this parameter is of a size proportional to the square root of the sample size. Since dummies are often used in connection with big outliers, this indicates that the finite-sample distribution would be affected.

Consider the univariate DGP:

$$\Delta x_t = \psi 1_{\{t=T_b\}} + \nu_t \quad \text{where } \nu_t \sim \text{IN}[0, \sigma_\nu^2] \quad (26)$$

when  $t = 1, \dots, T$  and  $1 < T_b < T$ . The unit-root test is based on the squared correlation of  $\Delta x_t$  and  $x_{t-1}$  where both are corrected for  $1_{\{t=T_b\}}$ . We set  $\sigma_\nu^2 = 1$ .

The case of large dummies is considered first. Introduce the scaled parameter  $\delta = \psi/\sqrt{T}$  and the univariate Brownian motion  $B_u$ . Then the asymptotic behaviour of the residual matrix, for fixed  $\delta$ , is given by:

$$\begin{aligned} S_{00} &= \frac{1}{T} \sum_{t \neq T_b} \Delta x_t^2 \quad \xrightarrow{\mathcal{P}} 1 \\ S_{10} &= \frac{1}{T} \sum_{t \neq T_b} x_{t-1} \Delta x_t \quad \xrightarrow{\mathcal{D}} \int_0^1 B_u dB_u + \delta (B_1 - B_b) = G \text{ (say)}, \\ S_{11} &= \frac{1}{T^2} \sum_{t \neq T_b} x_{t-1}^2 \quad \xrightarrow{\mathcal{D}} \int_0^1 (B_u + \delta 1_{\{u \geq b\}})^2 du = H \text{ (say)}. \end{aligned} \quad (27)$$

In (27),  $B_1$  and  $B_b$  are the values at full sample and the proportion  $b$ . Therefore the distribution of the unit-root test is asymptotically equivalent to

$$T\lambda = T S_{00}^{-1} S_{01} S_{11}^{-1} S_{10} \xrightarrow{\mathcal{D}} \frac{G^2}{H}.$$

This also indicates that for small dummies, for fixed  $\psi$ , the usual asymptotic distribution without nuisance parameters applies.



### 12.2.2 Unit-root test in a model with a restricted dummy

A well-specified model of this type may lead to similar tests, although these are based on non-standard distributions. A general model to consider in this case is given by:

$$\Delta x_t = \pi' \begin{pmatrix} x_{t-1} \\ 1 \\ 1_{\{t=T_b\}} \end{pmatrix} + \psi \Delta 1_{\{t=T_b\}} + \nu_t.$$

This would allow for processes with a ‘blip’ in all components at time  $T_b$ . However, to simplify the algebra, we consider a simpler situation, where the conclusion is basically the same.

Consider the DGP:

$$\Delta x_t = \nu_t.$$

This will be analysed using the squared multiple correlation,  $\lambda$ , of  $\Delta x_t$  and the vector  $(x_{t-1}, 1_{\{t=T_b\}})$ . Due to invariance,  $x_{t-1}$  can be replaced by  $x_{t-1}$  corrected for the dummy. The residual matrices are:

$$\begin{aligned} S_{00} &= \frac{1}{T} \sum_{t \neq T_b} \Delta x_t^2 \xrightarrow{\mathcal{P}} 1, \\ \begin{pmatrix} 1 & 0 \\ 1 & T \end{pmatrix} \mathbf{S}_{10} &= \begin{pmatrix} \frac{1}{T} \sum_{t \neq T_b} x_{t-1} \Delta x_t \\ \Delta x_{T_b} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 B_u dB_u \\ \nu_{T_b} \end{pmatrix} \\ \frac{1}{T} \begin{pmatrix} 1 & 0 \\ 1 & T \end{pmatrix} \mathbf{S}_{11} \begin{pmatrix} 1 & 0 \\ 1 & T \end{pmatrix} &= \begin{pmatrix} \frac{1}{T^2} \sum_{t \neq T_b} x_{t-1}^2 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 B_u^2 du & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Consequently:

$$T\lambda \xrightarrow{\mathcal{D}} \frac{\left( \int_0^1 B_u dB_u \right)^2}{\int_0^1 B_u^2 du} + \nu_{T_b}^2. \quad (28)$$

This is the usual asymptotic unit-root statistic plus an extra term, which has an expected value of unity. There are three points to be made.

First, the two asymptotic terms in (28) are independent. This is seen by noting that  $x_t = \sum_{i=1}^t \nu_i$  and introducing  $x_t^* = x_t - \nu_{T_b} 1_{\{t \geq T_b\}}$ . Then:

$$\frac{1}{T} \sum_{t \neq T_b} x_{t-1} \Delta x_t = \frac{1}{T} \sum_{t \neq T_b} x_{t-1}^* \Delta x_t + \frac{\nu_{T_b}}{\sqrt{T}} T^{-1/2} \sum_{t > T_b} \nu_t.$$

The first term is independent of  $\nu_{T_b}$ , whereas the second term is asymptotically negligible. Correspondingly:

$$\frac{1}{T^2} \sum_{t \neq T_b} x_{t-1}^2 = \frac{1}{T^2} \sum_{t \neq T_b} x_{t-1}^{*2} + \frac{2\nu_{T_b}}{\sqrt{T}} T^{-3/2} \sum_{t > T_b} x_{t-1} + \frac{\nu_{T_b}^2}{T} T^{-1} \sum_{t > T_b} 1.$$

Again the first term is independent of  $\nu_{T_b}$ , and the last two terms are asymptotically negligible.

Secondly, a model with (say) two dummies would result in the addition of an extra independent term in (28), again with an expectation of unity. We checked this result using the Monte Carlo in §6.4 when there were no breaks in the DGP. Figure 11 shows the rejection frequencies when restricting dummies to the cointegration space, and increasing the number of impulses added to the model. The size of the test of  $r = 2$  when  $\text{rank}(\pi) = 2$  grows towards unity in this four-variable system. Indeed, the means of the test statistics increased essentially linearly in  $n - r$  as Table 8 shows. The columns correspond roughly to  $n$ ,  $n - 1$  and  $n - 2$ , as anticipated from (28).

Finally, the asymptotic distribution does not depend on the location in the sample of the dummy.

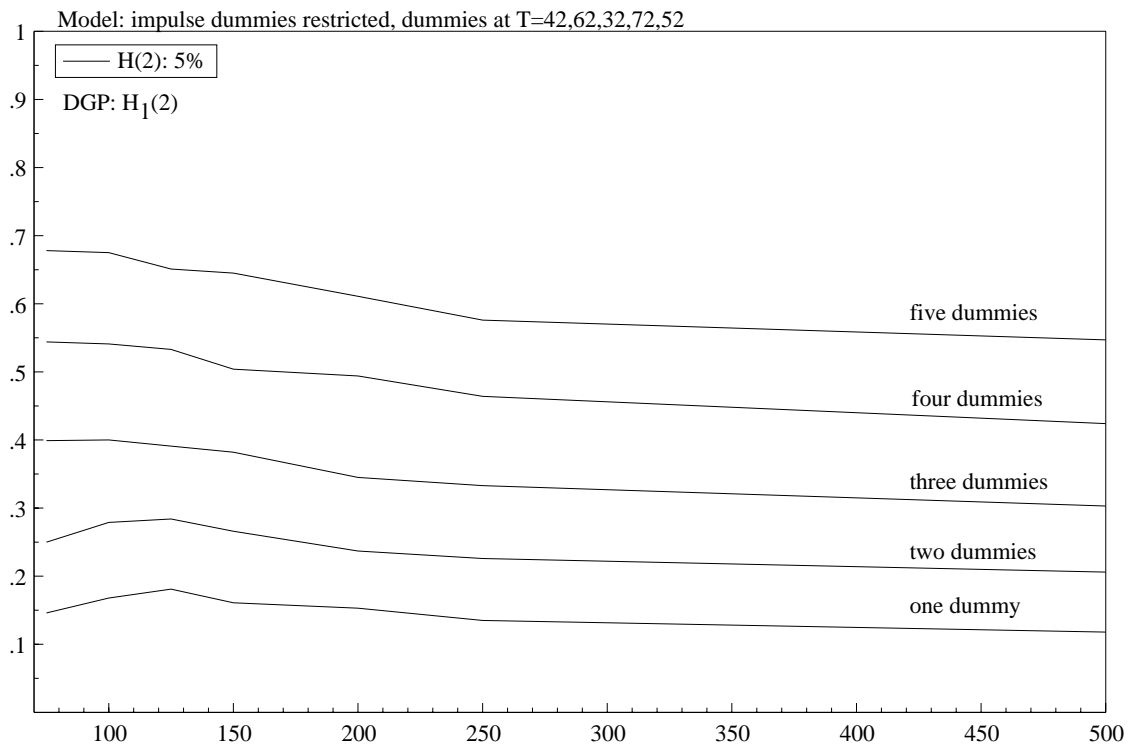


Figure 11 Rejection frequencies with restricted dummies in the model but none in the DGP.

Table 8 Increments in means of test statistics.

dummies	$H_l(0)$	$H_l(1)$	$H_l(2)$
0	0.00	0.00	0.00
1	4.24	3.18	2.08
2	4.03	3.04	1.96
3	4.07	2.95	1.92
4	4.18	3.17	2.06
5	4.03	3.05	1.99

### 12.2.3 Unit-root test in a mis-specified model with a restricted dummy

This example illustrates a possible explanation for the results reported in Hendry and Doornik (1994, Table 4): a dummy in the non-stationary component is modelled using a dummy which only occurs in the cointegration space. The conclusion from the analysis below is quite dramatic. The asymptotic distribution of the test depends on nuisance parameters, and in case of large dummies, the test is not consistent.

Consider the DGP (26). This is analyzed using the squared multiple correlation,  $\lambda$ , of  $\Delta x_t$  and of the vector  $(x_{t-1}, 1_{\{t=T_b\}})$ . The results of §12.2.1 are therefore of relevance again. The residual matrices are given by:

$$S_{00}^* = \frac{1}{T} \sum_{t=T_b} \Delta x_t^2 \xrightarrow{P} 1 + \rho^2, \quad \mathbf{S}_{10}^* = \begin{pmatrix} S_{10} \\ \Delta x_{T_b}/T \end{pmatrix}, \quad \mathbf{S}_{11}^* = \begin{pmatrix} S_{11} & 0 \\ 0 & 1 \end{pmatrix}.$$

Using the scaling matrix:

$$\mathbf{M}_T = \begin{pmatrix} 1/\sqrt{T} & 0 \\ 0 & \sqrt{T} \end{pmatrix},$$

for large values of  $\psi$ , i.e., fixed values of  $\delta = \psi/\sqrt{T}$ :

$$\begin{aligned} \lambda &= \mathbf{S}_{00}^{*-1} \mathbf{S}_{01}^* \mathbf{S}_{11}^{*-1} \mathbf{S}_{10}^* \\ &= \mathbf{S}_{00}^{*-1} \mathbf{S}_{01}^* \mathbf{M}_T (\mathbf{M}_T \mathbf{S}_{11}^* \mathbf{M}_T)^{-1} \mathbf{M}_T \mathbf{S}_{10}^* \\ &\xrightarrow{\mathcal{P}} (1 + \delta^2)^{-1} \begin{pmatrix} 0 \\ \delta \end{pmatrix}' \begin{pmatrix} G & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \delta \end{pmatrix} = \frac{\delta^2}{1 + \delta^2}, \end{aligned}$$

so that the test statistic, which is asymptotically equivalent to  $T\lambda$ , diverges to infinity.

It can be seen that for small values of the dummy parameter, the asymptotic distribution of the test involves a nuisance parameter.