ELSEVIER Journal of Econometrics 63 (1994) 37–60

JOURNAL OF Econometrics

Testing for an unstable root in conditional and structural error correction models

H. Peter Boswijk

Department of Actuarial Science and Econometrics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

(Received October 1993)

Abstract

This paper proposes a class of Wald tests for the hypothesis of an unstable root in conditional error correction models. Both single-equation models and simultaneous equations models in structural form are considered. The asymptotic distribution of the test statistics under the null hypothesis is derived in terms of a vector Brownian motion process, and critical values are obtained via Monte Carlo simulation. In an empirical model of the demand for money and the rate of inflation in the UK, the tests reject the instability hypothesis.

Key words: Cointegration; Error correction models; Identification; Money demand; Stability
JEL classification: C32

1. Introduction

The error correction model is characterized by a clear distinction between short-run variation and adjustment towards long-run relationships. For these relationships to be interpretable, a vital requirement is that the model be stable, i.e., that its characteristic polynomial does not have any unit roots. Stability implies that the dependent variables, following a shock to the system, revert to their equilibrium path.

This is a complete revision of the paper 'Testing for Cointegration in Conditional Models', presented at the first $(EC)^2$ conference in Amsterdam. Helpful comments and suggestions by David Hendry, Jan Kiviet, Søren Johansen, Walter Krämer, and three anonymous referees are gratefully acknowledged.

SSDI 0304407693015609

The theory of cointegration (see Granger, 1981; Engle and Granger, 1987) has provided a new interpretation of the error correction model and stability analysis. Cointegration entails the existence of stable relationships between time series that are individually integrated of order one, i.e., time series whose univariate models have a unit root. This implies that a closed autoregressive model of the vector time series is unstable in some directions (the integrated components), but at the same time stable in other directions (the cointegrating relationships). Hence it appears that cointegration may be conveniently analyzed in closed vector autoregressive (VAR) models. Indeed, Johansen (1988, 1991a) proposes to determine the number of cointegrating relationships via a sequence of likelihood ratio tests for unit roots in VAR models (see also Johansen and Juselius, 1990).

In this paper we consider testing for (in)stability of conditional error correction models for cointegrated variables. There are a number of motivations to consider such models next to VAR models. Firstly, long-run economic theory often suggests a large number of explanatory variables in addition to the endogenous variables. Endogenizing all variables may in turn require new explanatory variables, so that the dimension of the system soon becomes unmanageable. Secondly, a VAR model is essentially a statistical model of the series of interest in reduced form. If the analysis of this reduced form suggests the presence of long-run relations, a natural next step is to look for a structure to account for these findings. Specifying such a structural model involves a classification of endogenous and exogenous variables and an identification analysis to separate the long-run relations. Stability is a first requirement for the structural model to explain the observed cointegrating relations. This suggests that cointegration testing in VAR models and stability testing in structural (conditional) models should be seen as complements rather than substitutes.

The plan of the paper is as follows. Section 2 discusses the stability condition in both single-equation and simultaneous error correction models. For the latter type of model, the identification of the long-run as well as the short-run parameters is analyzed. In Section 3, a class of Wald-type test statistics for the null hypothesis of instability is proposed. In the fourth section the asymptotic distribution of the statistics under the null hypothesis is derived in terms of a vector Brownian motion process, and the tests are shown to be consistent. Critical values for the tests are obtained via Monte Carlo simulation. It is also shown in this section that a class of t-tests for an unstable root is in general not asymptotically similar. Section 5 discusses the role of the assumptions, and the final section contains an empirical application of the test to a model of the demand for money and the rate of inflation in the UK.

2. The model

Consider the single-equation error correction model of a time series $\{y_t\}$

conditional upon the $k \times 1$ vector time series $\{z_t\}$, $t = 1, \ldots, T$:

$$\Delta y_t = \beta'_0 \Delta z_t + \lambda(y_{t-1} - \theta' z_{t-1}) + \sum_{j=1}^{p-1} (\gamma_j \Delta y_{t-j} + \beta'_j \Delta z_{t-j}) + v_t, \tag{1}$$

where $\{v_t\}$ is an innovation process relative to $\{z_t, y_{t-j}, z_{t-j},\ j = 1, 2, \ldots\}$ with positive variance ω^2, p denotes the maximal lag length, and where λ and γ_j, $j = 1, \ldots, p-1$, are (scalar) parameters whereas θ and β_j, $j = 0, \ldots, p-1$, are $k \times 1$ parameter vectors. Here θ defines the long-run equilibrium relation $y = \theta' z$, the deviations from which lead to a correction of y_t by a proportion λ, the adjustment or error correction coefficient. The k-vector of conditioning variables $\{z_t\}$ is assumed to be integrated of order 1, or $I(1)$. No parametric model of z_t is specified, although some further conditions will be required in Assumption 1 below.

The conditional model is said to be stable if all roots of the characteristic equation

$$\varphi(\zeta) = (1 - \zeta)\left(1 - \sum_{j=1}^{p-1} \gamma_j \zeta^j\right) - \lambda\zeta = 0 \tag{2}$$

are outside the unit circle. Stability of the model implies that the disequilibrium error $(y_t - \theta' z_t)$ is a stationary process, even though z_t and y_t (if $\theta \neq 0$) are $I(1)$ and hence nonstationary. That is, if the model is stable, then $x_t = (y_t, z'_t)'$ is cointegrated with cointegrating vector $(1, -\theta')'$. The purpose of this paper is to devise a class of tests for the null hypothesis that the characteristic equation has a unit root, so that the model is unstable, against the alternative hypothesis of stability. From (2) it is easily seen that $\varphi(1) = -\lambda$, so that a unit root corresponds to $\lambda = 0$.

The single-equation conditional model (1) can be seen as a special case of a structural error correction model. This is a system of simultaneous error correction equations for a $g \times 1$ vector time series $\{y_t\}$ conditional upon $\{z_t\}$:

$$\Gamma_0 \Delta y_t = \mathrm{B}_0 \Delta z_t + \Lambda(\Gamma y_{t-1} + \mathrm{B} z_{t-1}) + \sum_{j=1}^{p-1} (\Gamma_j \Delta y_{t-j} + \mathrm{B}_j \Delta z_{t-j}) + v_t. \tag{3}$$

Here $\{v_t\}$ is an innovation process relative to $\{z_t, y_{t-j}, z_{t-j},\ j = 1, 2, \ldots\}$ with positive definite covariance matrix Ω, so that (3) is a reparametrization (if $\Gamma_0 \neq I_g$) of a conditional model of y_t given z_t. The parameter matrices Λ, Γ, and Γ_j, $j = 0, \ldots, p-1$, are of order $g \times g$, whereas B and B_j, $j = 0, \ldots, p-1$, are $g \times k$ matrices. Identification of these parameters is considered below. The model (3) implies g cointegrating relationships $\Gamma y + \mathrm{B} z = 0$, provided that it is stable, i.e., that the characteristic equation

$$\varphi(\zeta) = \left|(1 - \zeta)\left(\Gamma_0 - \sum_{j=1}^{p-1} \Gamma_j \zeta^j\right) - \Lambda\Gamma\zeta\right| = 0 \tag{4}$$

has all roots outside the unit circle.

Unless the parameters are restricted in some way, the structural model is not identified. We identify the long-run relations by imposing restrictions of the usual form

$$\Gamma_{ii} = 1, \qquad R_i[\Gamma_{i.} \vdots \mathrm{B}_{i.}]' = 0, \qquad i = 1, \dots, g, \tag{5}$$

where $\Gamma_{i.}$ and $\mathrm{B}_{i.}$ denote the ith row of Γ and B, respectively, and where R_i is a known matrix of appropriate order. The rank condition for identification of the ith long-run relation is

$$\operatorname{rank} R_i[\Gamma \vdots \mathrm{B}]' = g - 1. \tag{6}$$

The remaining parameters are identified by the normalization $\Gamma_{0,ii} = 1$, and the restriction

$$\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_g), \tag{7}$$

i.e., Λ is a diagonal matrix. This means that only the disequilibrium error of the ith long-run relation appears in the ith structural error correction equation. Hence the restrictions (5) are also imposed on the error correction term of the ith equation (except for the normalization):

$$R_i[\Lambda[\Gamma \vdots \mathrm{B}]]'_{i.} = R_i[\Gamma_{i.} \vdots \mathrm{B}_{i.}]'\lambda_i = 0. \tag{8}$$

Thus the ith equation is (over)identified if the ith long-run relation is.

The reason for restricting the error correction matrix Λ to be diagonal is twofold. Firstly, it allows for an interpretation of the separate equations in (3) as representing economic behaviour of a group of agents, whose target consists in a particular long-run relationship, such as a money demand relation or a consumption function. Note that the possibility that all endogenous variables are affected by each disequilibrium error is not excluded, but this is considered a property of the *reduced form* of the system rather than the structural form.

A second motivation for (7) is that it facilitates the implementation and interpretation of a test for instability. From (4) it follows that instability occurs if $\varphi(1) = |-\Lambda\Gamma| = 0$, i.e., if either Λ or Γ has a deficient rank. The joint occurrence of $|\Gamma| = 0$ and $|\Lambda| \neq 0$ would imply that there are g cointegrating relations, at least one (linear combination) of which involves only components of z_t. In that case the normalization of Γ is invalid, so that Γ and B are not identified. If this possibility is excluded, the null hypothesis of instability can be tested by checking whether $|\Lambda| = 0$, i.e., whether $\lambda_i = 0$ for some i.

Under this null hypothesis, there is no error correction in the ith equation of (3), which suggests that the ith row of the system $\Gamma y + \mathrm{B}z = 0$ is not a cointegrating relationship. However, even if $\lambda_i = 0$, the possibility exists that there is an error correction mechanism in the data-generating process of z_t. Specifically, let z_t be generated by

$$\Delta z_t = \alpha(\Gamma y_{t-1} + \mathrm{B}z_{t-1}) + \sum_{j=1}^{p-1} (\mathrm{A}_{1j}\Delta y_{t-j} + \mathrm{A}_{2j}\Delta z_{t-j}) + \varepsilon_t, \tag{9}$$

where $\{\varepsilon_t\}$ is an innovation process relative to the past of y_t and z_t, and where α, A_{1j}, and A_{2j}, $j = 1, \ldots, p-1$, are parameters matrices of appropriate order. Together, (3) and (9) constitute a conditional and marginal error correction model. If the ith column of α differs from zero, then the ith disequilibrium error $(\Gamma_{i.} y_{t-1} + B_{i.} z_{t-1})$ leads to a correction in z_t, so that there is cointegration, even if λ_i is zero. In Section 4, we shall exclude this possibility, i.e., we shall evaluate the asymptotic properties of the instability test under the following assumption:

Assumption 1. The number of stable relationships in (1) or (3) is equal to the number of cointegrating relationships.

Assumption 1 can be interpreted as a particular type of exogeneity assumption, because it essentially states that the cointegration properties of the conditional model carry over to the full VAR system, obtained as a reduced form of (3) and (9). Notice that the possibility of a cointegrating relationship involving only z_t (or stationarity of any of the components of z_t) is hereby excluded. The motivation for making this assumption and the implications of its failure are considered in Section 5. As will be discussed there, the assumption is required to establish the asymptotic null distribution of the test, and to safeguard its consistency. It is difficult to test for the validity of the assumption directly. However, it may be replaced by:

Assumption 1′. In (9), $\alpha = 0$, so that z_t is weakly exogenous for Γ and B.

The weak exogeneity assumption is stronger than Assumption 1, but is easier to interpret and to test (cf. Boswijk, 1992; Johansen, 1992a). If weak exogeneity holds, then the cointegration properties of the full system are determined solely by the conditional model, so that the tests for instability, proposed below, can be interpreted as cointegration tests. Note that Assumption 1 only requires $\alpha = 0$ under the null hypothesis $\lambda = 0$; i.e., it allows the stable relationships in (1) or (3) to enter the marginal model (9).

3. The test statistics

Consider the single-equation model (1). As discussed in Section 2, the null hypothesis of instability corresponds to

$$H_0\colon\ \lambda = 0. \tag{10}$$

This hypothesis is to be tested against the alternative of stability, which implies $\lambda < 0$. Let $x_t = (y_t, z_t')'$ and $\delta = (\delta_1, \delta_2')' = (\lambda, -\lambda\theta')'$, the coefficient vector of x_{t-1} in (1), and let w_t and π denote the vectors of remaining right-hand side variables and parameters, respectively. Hence (1) becomes

$$\Delta y_t = \delta' x_{t-1} + \pi' w_t + v_t. \tag{11}$$

For a given sample $\{x_t,\ t = 1-p,\dots,0,1,\dots,T\}$, let variables without a subscript t denote the corresponding data matrices and vectors, so that $\Delta y = (\Delta y_1,\dots,\Delta y_T)'$, and further $X_{-1} = (x_0,\dots,x_{T-1})' = [y_{-1} \vdots Z_{-1}]$ and $W = (w_1,\dots,w_T)'$.The ordinary least-squares (OLS) estimator of δ is

$$\hat{\delta} = [X'_{-1} M(W) X_{-1}]^{-1} X'_{-1} M(W) \Delta y, \tag{12}$$

where $M(W) = I_T - W(W'W)^{-1}W'$, and its estimated covariance matrix is

$$\hat{V}[\hat{\delta}] = \hat{\omega}^2 [X'_{-1} M(W) X_{-1}]^{-1},$$

$$\hat{\omega}^2 = \frac{1}{T-l} (\Delta y - X_{-1}\hat{\delta})' M(W) (\Delta y - X_{-1}\hat{\delta}), \tag{13}$$

with l the number of regressors in (11).

From the definition of δ, it is easily seen that $\lambda = 0$ implies $\delta_1 = 0$, but also $\delta_2 = 0$. Therefore, following Boswijk (1989, 1992), this null hypothesis can be tested using a Wald statistic for $\delta = 0$, i.e.,

$$\xi = \hat{\delta}' \hat{V}[\hat{\delta}]^{-1} \hat{\delta}. \tag{14}$$

If the disturbances are normally distributed and the conditioning variables are weakly exogenous, then ξ is also (a monotonic transformation of) the likelihood ratio statistic for $\lambda = 0$. Notice that ξ tests the hypothesis that y_t and z_t only appear in first differences, i.e., that their lag polynomials have a common factor with a unit root. As such, the same statistic was proposed by Hendry and Mizon (1978), although they did not discuss the connection with nonstationarity and nonstandard asymptotic distributions. Alternatively, the Wald statistic may be interpreted as a test for the significance of the estimated error correction term. Define $\hat{\theta} = -\hat{\delta}_2/\hat{\delta}_1$, the traditional long-run multiplier estimator, and $\hat{u}_{-1} = (y_{-1} - Z_{-1}\hat{\theta})$, the estimated vector of error correction terms. Note that

$$X_{-1}\hat{\delta} = y_{-1}\hat{\delta}_1 + Z_{-1}\hat{\delta}_2 = \{y_{-1} - Z_{-1}(-\hat{\delta}_2/\hat{\delta}_1)\}\hat{\delta}_1 = \hat{u}_{-1}\hat{\delta}_1, \tag{15}$$

so that the Wald statistic can be expressed as

$$\xi = \frac{\hat{\delta}_1 [\hat{u}'_{-1} M(W) \hat{u}_{-1}] \hat{\delta}_1}{\hat{\omega}^2} = \tau^2, \tag{16}$$

where (apart from the degrees of freedom correction in $\hat{\omega}^2$) τ is the t-ratio of $\hat{\delta}_1$ in

$$\Delta y = \hat{u}_{-1}\delta_1 + W\pi + v. \tag{17}$$

Observe that if z_t were void, then $\hat{u}_{-1} = y_{-1}$ and $W = [\Delta y_{-1}, \ldots, \Delta y_{-p+1}]$, hence τ reduces to the (augmented) Dickey–Fuller statistic (Dickey and Fuller, 1979).

If the dependence of δ_2 on λ is ignored, then the instability hypothesis may also be tested using a t-statistic for $\delta_1 = 0$ in (11). This alternative generalization of the Dickey–Fuller statistic was suggested as a cointegration test by Banerjee et al. (1986). Recently, Kiviet and Phillips (1992) proposed an exact similar test for cointegration for the case where $p = 1$ and z_t is strictly exogenous, based on the same t-statistic.

The Wald test ξ is related to Dickey and Fuller's (1981) Φ statistics. They consider testing the joint significance of the parameter λ and the constant and/or trend term in univariate autoregressive models, because the constant and trend term vanish under the null hypothesis of a unit root. The Wald test statistic may be generalized to allow for deterministic variables as follows.

Let s denote a T-vector of ones. If a constant term is added to the explanatory variables in (1), then the regression equation is

$$\Delta y = X_{-1}\delta + sb + W\pi + v, \tag{18}$$

where b is the intercept. Two different version of the test statistic may be considered. Firstly, if the constant term is unrestricted under the null, this means that the matrix W is replaced by $[s \vdots W]$ and π is replaced by $(b, \pi')'$ in (12)–(17), leading to the statistic ξ_μ (say). However, by a similar argument as in Johansen and Juselius (1990), it may be reasonable to restrict the constant term to zero under the null hypothesis to avoid a drift in $\{y_t\}$. This means that the constant term appears only in the error correction mechanism. In this case X_{-1} is replaced by $[s \vdots X_{-1}]$ and δ is replaced by (b, δ') in (12)–(17). The resulting statistic is denoted by ξ_μ^*.

If the variables in the system are generated by an integrated process with drift, then a linear trend term should be added to the regressors. This allows for a trend in the long-run relations, but it is also necessary to obtain asymptotically similar tests (see Boswijk, 1989). This yields the statistics ξ_τ if the trend term is unrestricted and ξ_τ^* if the trend term is restricted under H_0; a constant term is maintained in both cases. An overview of the various test statistics is presented in Table 1. Notice that H_0 in Table 1 refers only to the restrictions that are *imposed* under the null hypothesis. However, the asymptotic null distributions, tabulated in Appendix B, shall be derived (in Section 4) under the assumption that the intercept is zero for ξ_μ and that the trend coefficient is zero for ξ_τ.

Consider now the simultaneous error correction model, assumed to be just-identified. The methods proposed here can of course be generalized to overidentified models, but these are not considered here for ease of notation, and because overidentifying restrictions can be tested and imposed at a later stage, after stability of the model has been established. Thus the matrices R_i, $i = 1, \ldots, g$, defining the identifying restrictions, are of order $(g - 1) \times n$. Let $\{H_i,$

Table 1
Constant (c) and trend (t) in Wald stability test statistics

	ξ	ξ_μ^*	ξ_μ	ξ_τ^*	ξ_τ
H_0	–	–	c	c	c, t
H_1	–	c	c	c, t	c, t

$i = 1, \dots, g\}$ denote $n \times (k+1)$ matrices of full column rank such that $R_i H_i = 0$, i.e., the columns of H_i span the null space of R_i. Then, using (8), the vector of coefficients of x_{t-1} in the ith equation satisfies

$$\lambda_i [\Gamma_{i.} \,\vdots\, \mathrm{B}_{i.}]' = H_i \delta_i, \tag{19}$$

for some $(k+1) \times 1$ vector δ_i. Because of the normalization $\Gamma_{ii} = 1$, H_i can be defined in such a way that the first component of δ_i equals the error correction parameter λ_i. Therefore, instability in the ith equation can be tested by checking whether $\lambda_i = 0$ or, because of (19), whether $\delta_i = 0$.

Because the system is just-identified, the test statistics are based on two-stage least-squares (2SLS) estimation of the ith error correction equation. Let Δy_i and v_i denote the T-vectors with components Δy_{it} and v_{it}, and let ΔY_i be the $T \times (g-1)$ matrix of observations on the first difference of the other endogenous variables. Next, let $W_i = [\Delta Y_i \,\vdots\, V]$ be the matrix of regressors in the ith equation other than the error correction term, with π_i its parameter vector (the tth row of V is $(\Delta z_t', \Delta x_{t-1}', \dots, \Delta x_{t-p+1}')'$). Then the ith equation reads

$$\Delta y_i = X_{-1} H_i \delta_i + W_i \pi_i + v_i. \tag{20}$$

Let $\Delta \hat{Y}_i$ denote the matrix of reduced form fitted values, obtained from a multivariate regression of ΔY_i on $[X_{-1} \,\vdots\, V]$ and let $\hat{W}_i = [\Delta \hat{Y} \,\vdots\, V]$. The 2SLS estimator of δ_i is given by

$$\hat{\delta}_i = [H_i' X_{-1}' M(\hat{W}_i) X_{-1} H_i]^{-1} H_i' X_{-1}' M(\hat{W}_i) \Delta y_i. \tag{21}$$

Its covariance matrix estimator is

$$\begin{aligned} \hat{V}[\hat{\delta}_i] &= \hat{\omega}_{ii} [H_i' X_{-1}' M(\hat{W}_i) X_{-1} H_i]^{-1}, \\ \hat{\omega}_{ii} &= \frac{1}{T-l} (\Delta y_i - X_{-1} H_i \hat{\delta}_i - W_i \hat{\pi}_i)' (\Delta y_i - X_{-1} H_i \hat{\delta}_i - W_i \hat{\pi}_i), \end{aligned} \tag{22}$$

where $\hat{\pi}_i = (\hat{W}_i' \hat{W}_i)^{-1} \hat{W}_i' (\Delta y_i - X_{-1} H_i \hat{\delta}_i)$. From these, the Wald statistics ξ, ξ_μ, ξ_μ^*, ξ_τ, and ξ_τ^* can be defined analogously to their OLS counterparts.

4. Asymptotic properties

In this section we shall derive the asymptotic distribution of the test statistics under the null hypothesis of instability. First, we make an additional assumption. Let $e_t = (v_t', \Delta z_t')'$ and let S_t denote the partial sum process of e_t, i.e.,

$$S_t = \sum_{s=1}^{t} e_s, \qquad S_0 = 0. \tag{23}$$

Assumption 2. The process $\{e_t\}$ satisfies an invariance principle, so that as $T \to \infty$,

$$\frac{1}{\sqrt{T}} S_{[rT]} \Rightarrow B(r), \qquad r \in [0,1], \tag{24}$$

where $B(r) = (B_1(r)', B_2(r)')'$, a $(g+k)$-vector Brownian motion process with covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \Omega & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{25}$$

Here and below the symbol '$\Rightarrow$' denotes weak convergence of probability measures. Conditions under which Assumption 2 holds are given in Billingsley (1968) and Chan and Wei (1988). Note that the covariance matrix of $B_1(r)$ is equal to that of $\{v_t\}$, because $\{v_t\}$ is an innovation process and hence serially uncorrelated.

The following lemma is obtained as a slight modification of Johansen's version of the Granger representation theorem (see Johansen, 1991a, Theorem 4.1). Let β denote the $n \times r$ matrix of cointegrating vectors, which is void if there is no cointegration ($r = 0$). Consider the reduced form of the simultaneous equations model (3):

$$\Delta y_t = \Pi_0 \Delta z_t + \gamma\beta' x_{t-1} + \sum_{j=1}^{p-1} \Pi_j \Delta x_{t-j} + \eta_t, \tag{26}$$

where $\Pi_0 = \Gamma_0^{-1} \mathrm{B}_0$, $\gamma\beta' = \Gamma_0^{-1} \Lambda [\Gamma \vdots \mathrm{B}]$, $\Pi_j = \Gamma_0^{-1} [\Gamma_j \vdots \mathrm{B}_j]$, $j = 1, \ldots, p-1$, and $\eta_t = \Gamma_0^{-1} v_t$, Note that the single-equation model (1) is already in the format of (26). Define $\gamma_\perp$ and $\beta_\perp$ as matrices of order $g \times (g-r)$ and $n \times (n-r)$, respectively, the columns of which span the null space of γ' and β', i.e., $\beta'\beta_\perp = 0$ and $\gamma'\gamma_\perp = 0$.

Lemma 1. If $\{y_t\}$ is generated by (1) or (3), and Assumption 1 holds, then the vector $x_t = (y_t', z_t')'$ has the following representation:

$$x_t = C \cdot S_t + u_t, \tag{27}$$

where u_t is a stationary process, and where C is given by

$$C = \beta_{\perp} A \begin{bmatrix} \gamma_{\perp}' \Gamma_0^{-1} & \gamma_{\perp}' \Pi_0 \\ 0 & I_k \end{bmatrix}, \tag{28}$$

for some nonsingular $(n - r) \times (n - r)$ matrix A.

Proofs are delegated to Appendix A. In Theorem 1, the asymptotic distribution of the test statistics is expressed in terms of a vector Brownian motion process. To express these functionals concisely, we introduce the following notation. Yet $Y(r)$ and $X(r)$ denote two vector (stochastic or deterministic) processes on [0, 1]. Then we define the residual operator M_X as

$$M_X Y(r) = Y(r) - \int_0^1 Y(r)X(r)' \, \mathrm{d}r \left[\int_0^1 X(r)X(r)' \, \mathrm{d}r \right]^{-1} X(r). \tag{29}$$

This operator generalizes the familiar projection matrix $M(X) = I_T - X(X'X)^{-1}X'$ to continuous time regressions. This is because $M_X Y(r) = Y(r) - \hat{A}X(r)$, where $\hat{A}$ minimizes the integral of squared residuals $\int \| Y(r) - AX(r) \|^2 \, \mathrm{d}r$. For example, $M_1 Y(r) = Y(r) - \int Y(s) \, \mathrm{d}s$, the deviation of $Y(r)$ from its sample mean.

Theorem 1. Let y_t be generated by (1) or (3), and let Assumptions 1 and 2 hold. Then as $T \to \infty$ and under H_0:

$$\xi, \xi_\mu, \xi_\mu^*, \xi_\tau, \xi_\tau^* \Rightarrow \int_0^1 \mathrm{d}W_1(r) U(r)' \left[\int_0^1 U(r)U(r)' \, \mathrm{d}r \right]^{-1} \times \int_0^1 U(r) \, \mathrm{d}W_1(r), \tag{30}$$

where $W(r) = (W_1(r), W_2(r)')'$ is a standard $(1 + k)$-vector Brownian motion process on [0, 1], and where $U(r)$ is a vector process on [0, 1] defined by:

(i) If there is no constant term in (1) or (3) and if x_t does not have a drift, then for ξ, $U(r) = W(r)$.

(ii) If x_t does not have a drift, then for ξ_μ, $U(r) = M_1 W(r)$, whereas for ξ_μ^*, $U(r) = (W(r)', 1)'$.

(iii) For ξ_τ, $U(r) = M_f W(r)$ with $f(r) = (1, r)'$, whereas for ξ_τ^*, $U(r) = M_1(W(r)', r)'$.

The functional (30) generalizes the asymptotic distributions of the Dickey–Fuller statistics to the case where stability is tested in a model with conditioning variables. Thus if $k = 0$, so that $W(r) = W_1(r)$, then the asymptotic distribution of ξ, ξ_μ, and ξ_τ is equal to the distribution of the square of the Dickey–Fuller statistics $\hat{\tau}$, $\hat{\tau}_\mu$, and $\hat{\tau}_\tau$, respectively, whereas ξ_μ^* and ξ_τ^* correspond to the F-statistics Φ_1 and Φ_3 in Dickey and Fuller (1981). The cumulative distribution function of the functionals can be obtained via Monte Carlo simulation, where a vector Brownian motion is replaced by a vector of Gaussian random walks and integrals are replaced by sums. Critical values thus obtained are given in Tables B.1 through B.5 in Appendix B.

We now consider the asymptotic behaviour under the null hypothesis of the t-ratio of $\hat{\delta}_1$ in (11), to be denoted τ_1. In Theorem 2 it is shown that the asymptotic distribution of τ_1 in general depends upon a nuisance parameter, so that this test is not asymptotically similar. We only give the result for the case where $g = 1$ and no deterministic components are included in the regression, as the derivation for the other cases will be entirely analogous.

Theorem 2. Let y_t be generated by (1), and let Assumptions 1 and 2 hold. If there is no constant term in (1) and x_t does not have a drift, then as $T \to \infty$ and under H_0:

$$\tau_1 \Rightarrow \left[\int_0^1 U(r)^2 \,\mathrm{d}r \right]^{-1/2} \times \left[\sqrt{(1-\rho^2)} \int_0^1 U(r)\,\mathrm{d}W_1(r) + \rho \int_0^1 U(r)\,\mathrm{d}\bar{W}_2(r) \right], \tag{31}$$

where $W(r) = (W_1(r), W_2(r)')'$ is a standard $(1+k)$-vector Brownian motion process on $[0,1]$, where $U(r) = M_{W_2} W_1(r)$, where $\bar{W}_2(r) = a' W_2(r)$ for any vector a such that $a'a = 1$ and where ρ is the multiple correlation coefficient of $B_1(r)$ and $B_2(r)$:

$$\rho^2 = \frac{\sigma_{12} \Sigma_{22}^{-1} \sigma_{21}}{\sigma_{11}}. \tag{32}$$

The nuisance parameter dependency in (31) leads us to investigate the special case when $\sigma_{12} = 0$ so that $\rho = 0$. From, e.g., Park and Phillips (1989), we have

$$\sigma_{12} = \mathrm{E}[v_t \Delta z_t'] + \sum_{j=1}^{\infty} \mathrm{E}[v_t \Delta z_{t-j}'] + \sum_{j=1}^{\infty} \mathrm{E}[v_{t-j} \Delta z_t']. \tag{33}$$

Because $\{v_t\}$ is an innovation relative to z_t and the past, the first two terms are equal to zero. A sufficient condition for the third term to vanish as well is that z_t be strictly exogenous in (1). If that is the case, one could use the functional in

Theorem 2 with $\rho = 0$ to construct critical values for the t-test statistic τ_1^1.[1] However, for strictly exogenous z_t the exact approach of Kiviet and Phillips (1992) seems more worthwhile, at least if the autoregressive order p equals 1.

The consistency of the Wald tests is analyzed in the final theorem of this section.

Theorem 3. Let y_t be generated by (1) or (3), and let Assumptions 1 and 2 hold. Then as $T \to \infty$ and under H_1,

$$\xi,\ \xi_\mu,\ \xi_\mu^*,\ \xi_\tau,\ \xi_\tau^* = O_p(T), \tag{34}$$

so that the tests are consistent.

5. Discussion

All asymptotic results in Section 4 require that Assumption 1 holds. In order to analyze the role of this assumption and the implications of its absence, consider the single-equation case (1). Here violation of Assumption 1 may occur either because there is a cointegrating relation $y = \theta' z$ which only appears in the marginal model (9), or because there is only cointegration between the components of z_t, i.e., the normalization of the cointegrating vector is invalid.

In the first case, the coefficients δ_1 and δ_2 in (11) are equal to zero, but because there exists a stationary linear combination of y_{t-1} and z_{t-1}, the Wald statistic will not have the asymptotic null distribution of Theorem 1. It can be derived from Park and Phillips (1989) that ξ now has the same distribution as the sum of the functional in (30), but with k replaced by $k - 1$, and a chi-square variate with one degree of freedom. Because there are less unit roots in the system, the critical values of this distribution are expected to be smaller than those corresponding to Theorem 1, in which case the asymptotic size (i.e., the maximum significance level) can still be controlled using the critical values from Table B.1. In practice, this case will show up if the estimated cointegrating rank from the VAR is larger than the number of significant stable relationships in the conditional model.

The second case has more serious effects on the performance of the test. Now δ_1 still equals zero, but because there exists a stationary linear combination of z_{t-1}, its coefficient vector δ_2 may differ from zero. If that happens, the Wald test will reject the null hypothesis, even though the model is unstable. The underlying reason for this problem is that the Wald test exploits the identifying information in the structural model; if this information is invalid then the test becomes inconsistent. For the specific case where some of the components of z_{t-1} are known to be stationary, this problem may be avoided by adding these variables to w_t instead of x_{t-1} in (11), i.e., by letting their coefficients be unrestricted under

[1] These critical values can be obtained from the author upon request.

the null hypothesis. Then it can be shown that the results of Section 4 will continue to hold, but with k now equal to the number of nonstationary components in z_t.

In both of the above cases the chosen structure is not satisfactory. In the first case the choice of the conditioning variables is inadequate, whereas in the second case the identifying restrictions are invalidated. The fact that we cannot expect Assumption 1 or Assumption 1′ to hold *a priori* means that the number of cointegrating relationships cannot be determined on the basis of the conditional model only. Therefore it is recommended to use Johansen's LR tests for this purpose, before the structural model is analyzed.

If Assumption 1 does hold, and in addition weak exogeneity of z_t (Assumption 1′) is assumed, then the Wald test can be interpreted as a test for cointegration, and therefore it can be compared to other approaches. In Boswijk and Franses (1992), the size and power properties of the Wald test are compared to those of the residual augmented Dickey–Fuller (ADF) test (see Engle and Granger, 1987) and Johansen's maximal eigenvalue test. They consider two data-generating processes, in both of which the conditioning variables are weakly exogenous, and find that the Wald test dominates the other two tests in both size and power performance. For the ADF test, this may be explained by the analysis of Kremers et al. (1992), who demonstrate that the implicit common factor restriction employed in the ADF test leads to its relatively low power for general dynamic structures. The power superiority over the maximal eigenvalue test should come as no surprise, because if the weak exogeneity restriction is known to hold, then it should also be imposed on the likelihood, in which case (and if $g = 1$) the likelihood ratio statistic will be a monotonic function of the Wald statistic.

6. An empirical application

In this section the Wald test is applied to a model of the demand for M1 money and the rate of inflation in the UK. We use Hendry and Ericsson's (1991) quarterly seasonally adjusted data from 1963(1) through 1989(2) on the monetary aggregate M1 (M), total final expenditure at 1985 prices (Y), the implicit price deflator for $Y(P)$, and the interest rate differential (R^*). For details on the sources and adjustments of the data, we refer to Hendry and Ericsson (1991, App.). In Johansen (1992b) it has been shown that $m = \ln(M)$ and $p = \ln(P)$ are integrated of order 2, whereas the log of real money $(m - p)$ is $I(1)$. Therefore we analyze $x_t = ((m-p)_t, \Delta p_t, y_t, R_t^*)'$, where $y = \ln(Y)$. After the creation of lags, the available sample size is 100 [1964(3)–1989(2)], the last 12 observations of which are preserved for post-sample prediction testing (hence $T = 88$).

For a slightly different data set and estimation period, Hendry and Mizon (1993) found two cointegrating relationships between real money, expenditure, and the inflation and interest rates. The first was interpreted as a long-run

Table 2
Likelihood ratio cointegration test statistics for UK data

H_0	Trace	λ_{max}
$r \leqslant 3$	2.802	2.802
$r \leqslant 2$	13.259	10.457
$r \leqslant 1$	39.932[a]	26.673[b]
$r = 0$	86.333[c]	46.401[c]

The 'trace' and 'λ_{max}' statistics are defined in Johansen and Juselius (1990).
[a] Significant at the 10% level.
[b] Significant at the 5% level.
[c] Significant at the 1% level.

money demand relationship and the second as a relationship between the rate of inflation and excess expenditure, where the latter is represented by y_t in deviation from a linear trend. However, these two relationships were obtained by taking *ad hoc* linear combinations of the two cointegrating vectors, without any formal identification analysis. This will be considered below.

Before analyzing a structural error correction model, we apply Johansen's procedure to test for the number cointegrating relationships in the system. A VAR model of order 5 with a constant and linear trend term appears to be well-specified according to diagnostics for first- and fourth-order serial correlation. To exclude the possibility of quadratic trends under the hypothesis of one or more unit roots, the linear trend is restricted to appear only in the error correction mechanism, similar to the ξ_τ^* statistic (cf. Johansen, 1991b). In Table 2 the likelihood ratio statistics for cointegration are given, which may be compared to critical values from Johansen (1991b, Table V). It appears that both tests do not reject the hypothesis of at most two cointegrating relations, whereas the null hypothesis of no cointegration is rejected. At the significance level of 10% the hypothesis of at most one long-run relationship is rejected as well, so that we corroborate Hendry and Mizon's (1993) conclusion of two cointegrating relationships for these data.

As indicated in the introduction, the finding of two cointegrating relations may be accounted for by a stable structural error correction model, which we consider next. As argued by Spanos (1990), the statistical adequacy of a structural model should be assessed via misspecification tests of the unrestricted reduced form of the model. Therefore, we start the analysis with an unrestricted model of $(m - p)$ and Δp, conditional upon y and R^*.[2] As with the VAR model,

[2] Because of their natural lower bound, the possibility of a unit root in interest rates is often questioned. However, these series usually display very slow mean-reversion, so that their behaviour may be better described by an integrated process than by a stationary process.

Table 3
Diagnostics for the unrestricted reduced form model

Test statistic	Null distribution	$\Delta(m-p)$	$\Delta^2 p$
AR1	$F(1,63)$	0.560	0.820
AR2	$F(4,60)$	0.305	0.233
ARCH	$F(4,56)$	0.645	2.174[a]
RESET	$F(3,61)$	0.260	0.632
CHOW	$F(12,64)$	0.831	0.674
N	$\chi^2(2)$	5.758[a]	0.758

The test statistics are *ARp*: LM test for AR(p)/MA(p) disturbances; *ARCH*: LM test for fourth-order ARCH disturbances; *RESET*: functional form misspecification test; *CHOW*: Chow test for post-sample predictive failure; *N*: LM test for nonnormality.

[a] Significant at the 10% level.

the maximal lag length p is set to 5, and a constant and trend term are included. Table 3 lists some statistics for the two reduced form equations. None of the diagnostics indicate misspecification at the 5% significance level, although there is some indication of nonnormality in the real money equation and of ARCH effects in the inflation equation.

The next step in the analysis is the specification of a just-identified structural model. In addition to the normalization restrictions and the diagonality of the error correction matrix, we impose a unit long-run income elasticity in the money demand equation and we exclude real money from the inflation rate equation. This just-identified model is estimated by 2SLS. For notational economy, we do not report the short-run parameters, but only the long-run equilibrium relationships with their error correction parameters. The long-run parameters are obtained by normalizing $\hat{\delta}_i$ with respect to the first component, and the corresponding standard errors are calculated using the methods given in Bårdsen (1989). The estimated money demand relationship is (standard errors between parentheses)

$$(m-p-y)_t = \underset{(1.80)}{-\,7.27}\Delta p_t \underset{(1.85)}{-\,8.35}R_t^* + \underset{(0.10)}{0.33} + \underset{(0.22)}{0.17}t/100 + \hat{u}_{1t}, \tag{35}$$

where the estimated adjustment parameter equals $\hat{\lambda}_1 = -\,0.14$ (with a standard error of 0.04) and the Wald stability test statistic equals $\xi_\tau = 35.75$. Comparing this statistic with the critical values from Table B.4 ($k = 2$), we see that the null hypothesis of instability is rejected at the 1% level. This allows (35) to be interpreted as a cointegrating money demand relationship. Notice that the linear trend term is not significant in this relation.

The estimated inflation rate relationship is given by

$$\underset{}{\Delta p_t} = \underset{(0.08)}{0.34 y_t} + \underset{(0.27)}{0.31 R_t^*} - \underset{(0.85)}{3.78} - \underset{(0.07)}{0.21 t/100} + \hat{u}_{2t}, \tag{36}$$

where $\hat{\lambda}_2 = -0.53$ (standard error 0.13). Because in this case the trend term is expected to play a significant role in the long-run relationship, we have calculated the ξ_τ^* statistic as well as ξ_τ; they are given by $\xi_\tau = 19.98$ and $\xi_\tau^* = 20.00$. Both tests reject the null hypothesis of instability at the 5% significance level. The interest rate does not appear to be significant in this relation, so that (36) may be interpreted as a long-run relationship between the rate of inflation, the level of expenditure, and a trend term (representing aggregate supply).

In summary, the two cointegrating relations that are indicated by the LR tests in a VAR model correspond to two stable structural equations for the demand for money and the rate of inflation. Therefore, it appears that Assumption 1 is not violated and that this structural model may be used as a basis for subsequent analysis, including testing for weak exogeneity of the conditioning variables y and R^* and testing overidentifying restrictions. These topics are analyzed in Boswijk (1992).

Appendix A: Proofs

Proof of Lemma 1. The system can be expressed as

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} \gamma \\ 0 \end{pmatrix} \beta' x_{t-1} + \sum_{j=1}^{p-1} \begin{bmatrix} \Pi_j \\ 0 \end{bmatrix} \Delta x_{t-j} + \begin{bmatrix} \Gamma_0^{-1} & \Pi_0 \\ 0 & I_k \end{bmatrix} \begin{pmatrix} v_t \\ \Delta z_t \end{pmatrix}. \tag{A.1}$$

This is a vector autoregressive model in ECM format, even though the errors are not uncorrelated. Letting $\alpha' = (\gamma', 0)$, so that the columns of

$$\alpha_\perp = \begin{bmatrix} \gamma_\perp & 0 \\ 0 & I_k \end{bmatrix} \tag{A.2}$$

span the null space of α', we can apply Theorem 4.1 of Johansen (1991a), leading to

$$x_t = x_0 + \beta_\perp (\alpha_\perp' \Psi \beta_\perp)^{-1} \alpha_\perp' \begin{bmatrix} \Gamma_0^{-1} & \Pi_0 \\ 0 & I_k \end{bmatrix} S_t + C_1(L) e_t, \tag{A.3}$$

where

$$\Psi = I_n + \alpha\beta' - \sum_{j=1}^{p-1} \begin{bmatrix} \Pi_j \\ 0 \end{bmatrix}. \tag{A.4}$$

Letting $u_t = x_0 + C_1(L)e_t$ and $\mathrm{A} = (\alpha_\perp' \Psi \beta_\perp)^{-1}$, the required expression obtains. □

Proof of Theorem 1. Consider first the single-equation case, i.e., $g = 1$. Under H_0, $\hat{\delta}$ satisfies [cf. (12)]:

$$\hat{\delta} = [X'_{-1}M(W)X_{-1}]^{-1}X'_{-1}M(W)v. \tag{A.5}$$

Let S_{-1} and U denote the $T \times n$ matrices of observations on S_{t-1} and u_t from Lemma 1, respectively. Because x_t does not have a drift, W contains only stationary mean-zero regressors, whereas U is stationary, but possibly with a nonzero mean because of x_0. Now Lemma 2.1 of Park and Phillips (1989), together with Assumption 2 and the assumption that $\{v_t\}$ is an innovation process relative to z_t and the past, implies the following results:

$$(T^{-2}S'_{-1}S_{-1},\ T^{-1}S'_{-1}v) \Rightarrow \left(\int_0^1 B(r)B(r)'\,\mathrm{d}r, \int_0^1 B(r)\,\mathrm{d}B_1(r) \right), \tag{A.6}$$

$$(T^{-1}W'S_{-1}, T^{-1}W'W, T^{-1/2}W'v, T^{-3/2}U'S_{-1}, T^{-1/2}U'W,$$
$$T^{-1/2}U'v) = \mathrm{O}_p(1).$$

Combining these results yields

$$\begin{aligned} T^{-2}X'_{-1}M(W)X_{-1} &= T^{-2}CS'_{-1}S_{-1}C' + \mathrm{o}_p(1) \\ &\Rightarrow C\int_0^1 B(r)B(r)'\,\mathrm{d}r\,C', \end{aligned} \tag{A.7}$$

$$T^{-1}X'_{-1}M(W)v = T^{-1}CS'_{-1}v + \mathrm{o}_p(1) \Rightarrow C\int_0^1 B(r)\,\mathrm{d}B_1(r). \tag{A.8}$$

Let L denote an $n \times n$ lower triangular matrix such that the Cholesky decomposition of Σ is LL', and let l_{11} denote its first diagonal element. Because L is lower triangular, it follows that $l_{11} = \omega$. Define the n-vector standard Brownian motion process (i.e., with covariance matrix I_n) as $W(r) = L^{-1}B(r)$, so that $B(r) = LW(r)$. Note that $B_1(r) = \omega W_1(r)$, where $W_1(r)$ is the first component of $W(r)$. The asymptotic distribution of $\hat{\delta}$ now follows from

$$T\hat{\delta} \Rightarrow \left[CL\int_0^1 W(r)W(r)'\,\mathrm{d}r\,L'C' \right]^{-1} CL\int_0^1 W(r)\,\mathrm{d}W_1(r)\omega. \tag{A.9}$$

The consistency proof of the error variance estimator $\hat{\omega}^2$ proceeds along the usual lines, and will not be discussed here. Then we have for the OLS covariance matrix of $\hat{\delta}$ in (13):

$$T^2\hat{V}[\hat{\delta}] \Rightarrow \omega^2\left[CL\int_0^1 W(r)W(r)'\,\mathrm{d}r(CL)' \right]^{-1}. \tag{A.10}$$

Combining (A.9) with (A.10) leads to

$$\xi \Rightarrow \int_0^1 \mathrm{d}W_1(r)W(r)'(CL)'\left[CL\int_0^1 W(r)W(r)'\,\mathrm{d}r(CL)'\right]^{-1}(CL)$$
$$\times \int_0^1 W(r)\,\mathrm{d}W_1(r)$$
$$= \int_0^1 \mathrm{d}W_1(r)W(r)'\left[\int_0^1 W(r)W(r)'\,\mathrm{d}r\right]^{-1}\int_0^1 W(r)\,\mathrm{d}W_1(r), \tag{A.11}$$

where the second equality follows from the fact that under H_0 there is no cointegration, so that C is nonsingular.

Let s denote a T-vector of ones, and consider the regression (18). Under the null hypothesis, $b = 0$, as otherwise y_t would have a drift. The appropriately scaled least-squares estimators of δ and b satisfy under H_0:

$$\begin{pmatrix} T\hat{\delta} \\ \sqrt{T}\hat{b} \end{pmatrix} = \begin{bmatrix} T^{-2}X'_{-1}M(W)X_{-1} & T^{-3/2}X'_{-1}M(W)s \\ T^{-3/2}s'M(W)X_{-1} & T^{-1}s'M(W)s \end{bmatrix}^{-1}$$
$$\times \begin{pmatrix} T^{-1}X'_{-1}M(W)v \\ T^{-1/2}s'M(W)v \end{pmatrix}. \tag{A.12}$$

Using Lemma 2.1 from Park and Phillips (1989) it can be shown that, in addition to (A.6),

$$(T^{-3/2}s'S_{-1}, T^{-1}s'v) \Rightarrow \left(\int_0^1 B(r)'\,\mathrm{d}r, \int_0^1 \mathrm{d}B_1(r)\right),$$
$$(T^{-1/2}s'W, T^{-1}s'U) = \mathrm{O}_p(1), \tag{A.13}$$

so that

$$\begin{bmatrix} T^{-2}X'_{-1}M(W)X_{-1} & T^{-3/2}X'_{-1}M(W)s \\ T^{-3/2}s'M(W)X_{-1} & T^{-1}s'M(W)s \end{bmatrix}$$
$$\Rightarrow \int_0^1 \begin{pmatrix} CB(r) \\ 1 \end{pmatrix}(B(r)'C', 1)'\,\mathrm{d}r, \tag{A.14}$$

$$\begin{pmatrix} T^{-1}X'_{-1}M(W)v \\ T^{-1/2}s'M(W)v \end{pmatrix} \Rightarrow \int_0^1 \begin{pmatrix} CB(r) \\ 1 \end{pmatrix}\mathrm{d}B_1(r). \tag{A.15}$$

Using again $B(r) = LW(r)$ and $B_1(r) = \omega W_1(r)$, the required result for ξ_μ^* is obtained [note that the inverse of (A.14), multiplied by ω^2, yields the covariance matrix of $(T\hat{\delta}', \sqrt{T}\hat{b})'$]. The expression for ξ_μ follows from the partitioned inverse of (A.14). The proof for the case were a trend term is added is analogous.

Now consider the first equation from a simultaneous system:

$$\Delta y_1 = X_{-1} H_1 \delta_1 + W_1 \pi_1 + v_1. \tag{A.16}$$

Because W_1 is a mean-zero stationary process, the same applies to $\hat{W}_1$. By the same derivations as for the single-equation case, we obtain

$$T\hat{\delta}_1 \Rightarrow \left[H_1' C \int_0^1 B(r)B(r)' \mathrm{d}r\, C' H_1 \right]^{-1} H_1' C \int_0^1 B(r)\,\mathrm{d}B_{11}(r), \tag{A.17}$$

where $B_{11}(r)$ is the first element of $B_1(r)$. Because the alternative hypothesis entails that the system is stable, we assume that under the null hypothesis $\lambda_1 = 0$, (A.16) is the only unstable equation in the system. Therefore, the number of cointegrating relations is $g - 1$ and the rank of C equals $(n - (g - 1)) = k + 1$. Using (28), we have

$$H_1' C = H_1' \beta_\perp \mathrm{A} \begin{bmatrix} \gamma_\perp' \Gamma_0^{-1} & \gamma_\perp' \Pi_0 \\ 0 & I_k \end{bmatrix}. \tag{A.18}$$

Because under the null hypothesis no cointegrating vector $\beta_1 = H_1 \delta_1$ exists, i.e., no δ_1 exists such that $\delta_1' H_1' \beta_\perp = 0$, other than $\delta_1 = 0$, we have that $H_1' \beta_\perp \mathrm{A}$ is a nonsingular $(k + 1) \times (k + 1)$ matrix. Let e_1 be the first $g \times 1$ unit vector, i.e., $e_1 = (1, 0, \ldots, 0)'$. Because $e_1' \Gamma_0 \gamma \beta' = e_1' \Lambda [\Gamma \,\vdots\, \mathrm{B}] = 0$ and β has full column rank, we have that $\gamma_\perp' \Gamma_0^{-1} = e_1'$. Therefore,

$$H_1' C B(r) = D \begin{pmatrix} B_{11}(r) \\ B_2(r) \end{pmatrix} = D \tilde{B}(r), \tag{A.19}$$

for some nonsingular matrix D and $(1 + k)$-vector Brownian motion $\tilde{B}(r)$, the first component of which is $B_{11}(r)$. The covariance matrix of $\tilde{B}(r)$ can again be decomposed as LL' with L lower triangular and $l_{11} = \omega_{11}$, which leads to the expression for ξ, analogously to the single-equation case. Likewise, the inclusion of a constant and linear trend term have the same effects as before. □

Proof of Theorem 2. Let $V(r) = CB(r)$, so that (A. 7) and (A.8) lead to

$$T\hat{\delta} \Rightarrow \left[\int_0^1 V(r) V(r)' \mathrm{d}r \right]^{-1} \int_0^1 V(r)\,\mathrm{d}B_1(r). \tag{A.20}$$

If $V(r)$ is partitioned as $(V_1(r), V_2(r)')'$, then partitioned regression leads to

$$\tau_1 \Rightarrow \left[\int_0^1 [M_{V_2} V_1(r)]^2 \mathrm{d}r \right]^{-1/2} \int_0^1 [M_{V_2} V_1(r)]\,\mathrm{d}B_1(r) \omega^{-1}. \tag{A.21}$$

Because $x_t = (y_t, z_t')'$ and $e_t = (v_t, \Delta z_t')'$, it is clear from (27) that C can be represented as

$$C = \begin{bmatrix} c_{11} & c_{12}' \\ 0 & I_k \end{bmatrix}, \tag{A.22}$$

where c_{11} and c_{12} follow from (28). Thus $V_1(r) = c_{11}B_1(r) + c'_{12}B_2(r)$ and $V_2(r) = B_2(r)$, and

$$M_{V_2}V_1(r) = M_{B_2}(c_{11}B_1(r) + c'_{12}B_2(r)) = c_{11}M_{B_2}B_1(r). \tag{A.23}$$

Define

$$P = \begin{bmatrix} p_{11} & p'_{12} \\ p_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} (\sigma_{11} - \sigma_{12}\Sigma_{22}^{-1}\sigma_{21})^{1/2} & \sigma_{12}\Sigma_{22}^{-1/2} \\ 0 & \Sigma_{22}^{1/2} \end{bmatrix}. \tag{A.24}$$

Because $PP' = \Sigma$, $B(r) = PW(r)$ for some n-vector standard Brownian motion $W(r) = (W_1(r), W_2(r)')'$. Note that $B_1(r) = p_{11}W_1(r) + p'_{12}W_2(r)$ and $B_2(r) = P_{22}W_2(r)$. This can be used to show that (cf. Phillips and Ouliaris, 1990, Lemma 2.2)

$$M_{B_2}B_1(r) = p_{11}M_{W_2}W_1(r) = p_{11}U(r). \tag{A.25}$$

Combining (A.21), (A.23), and (A.25) yields

$$\tau_1 \Rightarrow \left[\int_0^1 U(r)^2\,\mathrm{d}r\right]^{-1/2} \int_0^1 U(r)\,\mathrm{d}\left(\frac{p_{11}}{\omega}W_1(r) + \frac{1}{\omega}p'_{12}W_2(r)\right). \tag{A.26}$$

From the definition of p_{11} and because $\omega = \sqrt{\sigma_{11}}$, we have $(p_{11}/\omega) = (1-\rho^2)^{1/2}$. Because $(p'_{12}p_{12}/\omega^2) = \rho^2$, we can express p'_{12}/ω as $\rho a'$ with $a'a = 1$. This results in (31). Let S be an orthogonal $k \times k$ matrix, i.e., $S'S = I_k$. Because we may always replace $W_2(r)$ by $W_2^*(r) = SW_2(\mathrm{r})$ without changing the distribution in (A.26), we may also replace a by $a^* = Sa$, where $a^{*\prime}a^* = 1$. In other words, we may choose a to be any unit length vector. Hence the distribution only depends on the single nuisance parameter ρ. □

Proof of Theorem 3. We shall only proof the consistency for the single-equation case; the proof for the simultaneous-equations case is entirely analogous. We use $\xi = \tau^2$, with τ the t-ratio of δ_1 in (17). If $\delta_1 \neq 0$ and there is cointegration, then $\hat{\theta}$ is superconsistent, i.e., $(\hat{\theta} - \theta) = \mathrm{O}_p(T^{-1})$ (see, e.g., Boswijk, 1992). As shown by Engle and Granger (1987), this implies that $\hat{\delta}_1$ and $\hat{\pi}$ are asymptotically equivalent to the OLS estimators in

$$\Delta y = \delta_1 u_{-1} + W\pi + v. \tag{A.27}$$

Thus τ is asymptotically equivalent with the t-ratio of $\hat{\delta}_1$ in (A.27). Because u_{-1} is stationary under the alternative, $\hat{\delta}_1$ converges to δ_1 at a rate of $\mathrm{O}_p(T^{-1/2})$, see Park and Phillips (1989), so that $\hat{V}[\hat{\delta}_1] = \mathrm{O}_p(T^{-1})$. Letting $AV[\hat{\delta}_1]$ denote the asymptotic variance of $\sqrt{T}(\hat{\delta}_1 - \delta_1)$, we have

$$T^{-1/2}\tau \xrightarrow[P]{} (AV[\hat{\delta}_1])^{-1/2}\delta_1, \tag{A.28}$$

where $\rightarrow_P$ denotes convergence in probability, so that $\tau = \mathrm{O}_p(T^{1/2})$ and $\xi = \mathrm{O}_p(T)$. Adding a constant or a linear trend term to the regressors does not change the above results. The rate of divergence of the t- and Wald statistics only depend upon the consistency and convergence rate of the coefficient $\hat{\delta}_1$ of u_{-1}, which does not change. □

Appendix B: Critical values

In Tables B.1 through B.5 the asymptotic critical values for the Wald statistics ξ, ξ_μ, ξ_μ^*, ξ_τ, and ξ_τ^* are given, with k, the number of conditioning variables, ranging from 1 to 10. The quantiles are obtained via Monte Carlo simulation (with 10,000 replications) of the functionals of Theorem 1, where the Brownian motion processes are approximated by a Gaussian random walk with 500 observations.

Table B.1
Asymptotic critical values for the test statistic ξ

k	20%	10%	5%	2.5%	1%
1	4.73	6.35	7.95	9.64	11.83
2	7.37	9.40	11.37	13.34	15.87
3	9.89	12.12	14.31	16.42	18.78
4	12.28	14.79	17.13	19.04	21.51
5	14.71	17.42	19.78	21.88	24.17
6	16.97	19.89	22.41	24.63	27.18
7	19.20	22.29	24.77	27.16	30.30
8	21.51	24.53	27.40	29.90	32.62
9	23.67	26.82	29.67	32.16	35.07
10	25.73	29.12	32.08	34.65	38.25

Table B.2
Asymptotic critical values for the test statistic ξ_μ

k	20%	10%	5%	2.5%	1%
1	7.52	9.54	11.41	12.99	15.22
2	9.94	12.22	14.38	16.35	18.68
3	12.38	14.93	17.18	19.02	21.43
4	14.83	17.38	19.69	22.24	24.63
5	17.09	19.87	22.48	24.77	27.11
6	19.25	22.14	24.88	27.05	30.22
7	21.49	24.62	27.27	29.84	32.87
8	23.67	26.96	29.65	32.27	35.27
9	25.77	29.11	32.13	34.90	38.16
10	27.88	31.39	34.52	37.40	40.79

Table B.3
Critical values for the test statistic ξ_{μ}^{*}

k	20%	10%	5%	2.5%	1%
1	8.41	10.44	12.25	14.18	16.19
2	10.90	13.15	15.30	17.16	19.39
3	13.34	15.84	18.08	19.77	22.30
4	15.73	18.22	20.50	22.89	25.28
5	17.93	20.82	23.35	25.35	28.24
6	20.19	23.17	25.66	28.01	30.96
7	22.38	25.47	28.07	30.81	33.43
8	24.51	27.78	30.64	33.11	36.27
9	26.74	29.96	33.02	35.77	38.83
10	28.82	32.31	35.31	38.20	41.72

Table B.4
Critical values for the test statistic ξ_{τ}

k	20%	10%	5%	2.5%	1%
1	10.16	12.32	14.28	16.30	18.53
2	12.49	14.91	17.20	19.17	21.62
3	14.84	17.57	19.81	21.96	24.65
4	17.08	19.96	22.47	24.79	27.52
5	19.37	22.33	24.88	27.48	30.24
6	21.48	24.64	27.39	29.86	32.60
7	23.74	26.88	29.91	32.17	35.54
8	25.92	29.23	32.21	34.93	37.99
9	27.93	31.45	34.59	37.31	40.94
10	30.06	33.62	36.95	39.70	43.49

Table B.5
Critical values for the test statistic ξ_{τ}^{*}

k	20%	10%	5%	2.5%	1%
1	11.08	13.22	15.24	17.02	19.30
2	13.44	15.85	18.03	20.05	22.50
3	15.75	18.45	20.66	22.79	25.46
4	17.95	20.76	23.33	25.54	28.51
5	20.23	23.20	25.79	28.44	31.13
6	22.42	25.58	28.12	30.87	33.27
7	24.63	27.76	30.72	33.14	36.35
8	26.77	30.16	33.05	35.64	38.88
9	28.86	32.31	35.35	38.20	42.06
10	30.99	34.63	37.66	40.70	44.58

References

Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith, 1986, Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence, Oxford Bulletin of Economics and Statistics 48, 253–277.

Bårdsen, G., 1989, Estimation of long-run coefficients in error correction models, Oxford Bulletin of Economics and Statistics 51, 345–350.

Billingsley, P., 1968, Convergence of probability measures (Wiley, New York, NY).

Boswijk, H.P., 1989, Estimation and testing for cointegration with trended variables: A comparison of a static and a dynamic regression procedure, Report no. AE 12/89 (University of Amsterdam, Amsterdam).

Boswijk, H.P., 1992, Cointegration, identification and exogeneity: Inference in structural error correction models (Thesis Publishers, Amsterdam).

Boswijk, H.P. and P.H. Franses, 1992, Dynamic specification and cointegration, Oxford Bulletin of Economics and Statistics 54, 369–381.

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least-squares estimates of unstable autoregressive processes, Annals of Statistics 16, 367–401.

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427–431.

Dickey, D.A. and W.A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057–1072.

Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error correction: Representation, estimation, and testing, Econometrica 55, 251–276.

Granger, C.W.J., 1981, Some properties of time series data and their use in econometric model specification, Journal of Econometrics 16, 121–130.

Hendry, D.F. and N.R. Ericsson, 1991, Modeling the demand for narrow money in the United Kingdom and the United States, European Economic Review 35, 833–881.

Hendry, D.F. and G.E. Mizon, 1978, Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England, Economic Journal 88, 549–563.

Hendry, D.F. and G.E. Mizon, 1993, Evaluating dynamic econometric models by encompassing the VAR, in: P.C.B. Phillips, ed., Models, methods and applications of econometrics: Essays in honor of Rex Bergstrom (Basil Blackwell, Oxford).

Johansen, S., 1988, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1991a, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59, 1551–1580.

Johansen, S., 1991b, The role of the constant term in cointegration analysis of $I(1)$ variables, Discussion paper (University of Copenhagen, Copenhagen).

Johansen, S., 1992a, Cointegration in partial systems and the efficiency of single-equation analysis, Journal of Econometrics 52, 389–402.

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

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

Kiviet, J.F. and G.D.A. Phillips, 1992, Exact similar tests for unit roots and cointegration, Oxford Bulletin of Economics and Statistics 54, 349–367.

Kremers, J.J.M., N.R. Ericsson, and J.J. Dolado, 1992, The power of cointegration tests, Oxford Bulletin of Economics and Statistics 54, 325–348.

Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, Econometric Theory 5, 95–131.

Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, Econometrica 58, 165–193.

Spanos, A., 1990, The simultaneous-equations model revisited: Statistical adequacy and identification, Journal of Econometrics 44, 87–105.