LAGRANGE-MULTIPLIER TESTS FOR WEAK EXOGENEITY: A SYNTHESIS

H. Peter Boswijk
Department of Econometrics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands

Jean-Pierre Urbain
Department of Quantitative Economics, University of Limburg, 6200 MD Maastricht, The Netherlands

Key Words and Phrases: error correction models; exogeneity; Lagrange-multiplier test; money demand
JEL Classification: C32, C22

ABSTRACT

This paper unifies two seemingly separate approaches to test weak exogeneity in dynamic regression models with Lagrange-multiplier statistics. The first class of tests focuses on the orthogonality between innovations and conditioning variables, and thus is related to the Durbin-Wu-Hausman specification tests. The second approach has been developed more recently in the context of cointegration and error correction models, and concentrates on the question whether the conditioning variables display error correcting behaviour. It is shown that the vital difference between the two approaches stems from the choice of the parameters of interest. A new test is derived, which encompasses both its predecessors. The test is applied to an error correction model of the demand for money in Switzerland.

1 Introduction

Since Engle et al.'s (1983) seminal paper, there has been an abundant literature on testing for weak exogeneity. The Lagrange-multiplier (LM) or efficient score testing principle proved to be particularly useful for this purpose, since it requires estimation under the null hypothesis only. Initially, the
focus of this literature has been on the orthogonality between regressors and disturbances, see, e.g., Engle (1982) and Holly (1985). This research is closely connected to the earlier literature on specification testing, including Durbin (1954), Sargent (1964), Wu (1973), Hausman (1978), and Nakamura and Nakamura (1981). We shall refer to these tests as orthogonality tests.

With the advent of cointegration theory, cf. Engle and Granger (1987), a new type of exogeneity test emerged. Johansen (1992) and Urbain (1992) showed that weak exogeneity of a set of conditioning variables for the cointegration parameters requires the absence of an error correction term in the marginal model for these variables. The Lagrange-multiplier test statistic for this restriction was derived by Boswijk (1992, 1995), who showed that a variable-addition F-test proposed by Johansen (1992) has an LM test interpretation. In addition, Johansen (1992) and Hendry and Mizon (1993) analysed likelihood ratio tests for weak exogeneity in cointegrated systems. These tests shall be called error-correction tests.

The present paper seeks to integrate and unify these apparently separate strands of literature. The reason that such different tests can be distinguished stems from the nature of Engle et al.'s (1983) definition. They define weak exogeneity in a fully specified model (i.e., a family of joint densities of the observables) and for a particular parameter (vector) of interest. Correspondingly, Urbain (1992) argues that the relevant restriction to be tested (orthogonality or no error correction), even in the same model, depends crucially on the choice of the parameters of interest. If these include the cointegration parameters as well as the parameters describing the short-run dynamics, then weak exogeneity entails both restrictions. In this paper we derive a new LM statistic which tests the joint hypothesis of orthogonality and no error correction. We shall show that this new statistic can be calculated simply as the sum of the two separate statistics, which is a reflection of the separability of the joint null hypothesis (cf. Smith, 1994), and the independence of the two statistics under this hypothesis.

The plan of the paper is as follows. In Section 2, we define weak exogeneity, and we analyse this concept in the classical simultaneous equations model under limited information, and in the cointegrated error correction model. In Section 3 we derive the new LM test statistic for the joint hypothesis of orthogonality and no error correction, and we show that it can be decomposed into its two predecessors. In Section 4 we apply the tests to a model of the demand for money in Switzerland, and the final section concludes.

### 2 Weak Exogeneity

Consider an n-vector time series \( \{ x_t, t = 1, \ldots, T \} \), with a set of starting values \( X_0 \); the analysis will be conditional on \( X_0 \). Define \( X_1 = (x_1, \ldots, x_T) \) and \( X_t = (X_0, X_t) \). A statistical model of \( \{ x_t \} \) may be expressed as a family of joint densities

\[
D_X(X_t | X_0; \theta) = \prod_{t=1}^{T} D(x_t | X_{t-1}; \theta), \quad \theta \in \Theta,
\]

where \( \theta \) is an \( \ell \)-dimensional parameter vector, with parameter space \( \Theta \subseteq \mathbb{R}^\ell \). Let \( x_t \) be partitioned as \( x_t = (y_t', z_t')' \), where the \( m \)-vector time series \( \{ y_t \} \) is to be explained, whereas the \( k \)-vector time series \( \{ z_t \} \) consists of explanatory variables (\( n = m + k \)). The joint density of \( x_t \), conditional upon the past history \( X_{t-1} \), can be factorized as
LAGRANGE-MULTIPLIER TESTS FOR WEAK EXOGENEITY

\[ D_x(x_t | X_{t-1}; \theta) = D_{y|x}(y_t | z_t; \theta_1)D_2(z_t | X_{t-1}; \theta_2), \]  

where \( \theta_1 \) and \( \theta_2 \) are the parameter vectors of the conditional density \( D_{y|x} \) of \( y_t \) given \( z_t \) and the past, and the marginal density \( D_2 \) of \( z_t \) given the past, with parameter spaces \( \Theta_1 \) and \( \Theta_2 \), respectively.

We shall be concerned with the question whether the conditioning variables \( z_t \) may be considered exogenous for the purpose of statistical inference on the parameter vector of interest \( \psi \in \Psi \). This parameter may equal the entire parameter vector \( \theta \), but often only a subset or, more generally, a function of \( \theta \) is considered of interest:

\[ f : \Theta \rightarrow \Psi; \psi = f(\theta). \]

The basic idea of weak exogeneity is the following. If \( \psi \) can be determined from the conditional density parameters \( \theta_1 \) only, then we need not analyse the marginal density for statistical inference on \( \psi \), provided that there are no restrictions linking \( \theta_1 \) and \( \theta_2 \). This idea is formalized in the next definition (Engle et al., 1983):

**Definition 1** In the model (1)-(2), the vector of conditioning variables \( z_t \) is said to be weakly exogenous for the parameters of interest \( \psi \) if and only if

(i) There exists a function \( h : \Theta_1 \rightarrow \Psi; \psi = h(\theta_1) \), i.e., \( \psi \) is a function of \( \theta_1 \) only;

(ii) \( (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \), i.e., the parameters \( \theta_1 \) and \( \theta_2 \) are variation free.

2.1 Weak Exogeneity in Simultaneous Equations

Let us first consider the conditions for weak exogeneity in the classical limited information set-up of a single structural equation, completed by reduced form equations for the explanatory variables. Suppose that the conditional density \( D_x(x_t | X_{t-1}; \theta) \) in (1) corresponds to a Gaussian vector autoregression (VAR) of order \( p \), i.e.,

\[ x_t | X_{t-1} \sim N(\mu_t, \Omega), \quad \mu_t = \sum_{i=1}^{p} \Pi_i x_{t-i}, \quad t = 1, \ldots, T, \]

where \( \{\Pi_i\}, \Omega \) are functions of \( \theta \). If we define the reduced form disturbance vector \( \varepsilon_t \equiv x_t - \mu_t \), then, by construction, \( \varepsilon_t \) is a homoskedastic Gaussian mean innovation process relative to \( X_{t-1} \) (cf. Hendry and Richard, 1982), with covariance matrix \( \Omega \). The normality assumption will be indispensable for many of the results to follow, since the conditions for weak exogeneity as well as the form of the LM statistic depend on the specific form of the joint density \( D_x \). However, as usual the asymptotic distribution of the LM statistic under the null hypothesis will hold under a wider class of distributional assumptions.

The hypothesis of a set of \( g (\leq n) \) behavioural relationships may be formulated as

\[ B \mu_t + \sum_{i=1}^{p} C_i x_{t-i} = 0, \quad t = 1, \ldots, T, \]

where the \( g \times n \) matrices \( B \) (of rank \( g \)) and \( C_i \) are functions of a vector of structural parameters, denoted by \( \phi \). The structural disturbance \( u_t \) is defined by
\[ u_t \equiv Bx_t + \sum_{i=1}^{p} C_i x_{t-i} = B(x_t - \mu_t), \]

where the second equality only holds under the hypothesis (4); in that case, \( u_t \) is a homoskedastic Gaussian mean innovation process relative to \( X_{t-1} \) with covariance matrix \( \Sigma \). The relationship between the structural and reduced form parameters is given by

\[ B\Pi_i + C_i = 0, \quad i = 1, \ldots, p, \quad \Sigma = B\Omega B'. \] (5)

Global identification of the structural parameters requires that for any given parameter value of \( (\{\Pi_i\}, \Omega) \), (5) can be solved to yield unique values for \( (B, \{C_i\}, \Sigma) \).

Let \( x_t \) be partitioned as \( (y_t', z_t')' \) (of dimensions \( m \) and \( k \), respectively), and let the \( g \) structural equations be separated into \( g_1 \) equations of interest and \( g_2 (= g - g_1) \) remaining equations. Richard (1980) and Engle et al. (1983) derive sufficient conditions for weak exogeneity of \( z_t \) for the parameters of interest. Here we shall consider only the following special case.

**Assumption 1** In the simultaneous equations model characterized by the hypothesis (4),

(i) \( g = n \), so that the system is complete;

(ii) \( g_1 = m = 1 \), so that there is a single structural equation of interest, and a single modelled variable \( y_t \);

(iii) the matrix \( B \) is restricted to

\[ B = \begin{bmatrix} 1 & -\delta_0 \\ 0 & I_k \end{bmatrix}. \] (6)

Partition \( \Pi_i \), \( \Omega \), \( C_i \) and \( \Sigma \) conformably with \( x_t \):

\[ \Pi_i = \begin{bmatrix} \pi_{1i}' \\ \Pi_{2i} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}, \quad C_i = \begin{bmatrix} c_{1i}' \\ C_{2i} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}, \]

and let \( \varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2})' \) and \( u_t = (u_{1t}, u_{2t})' \). Because of (6), we have \( C_{2i} = -\Pi_{2i}, \quad u_{2t} = \varepsilon_{2t}, \quad \Sigma_{22} = \Omega_{22} \). Redefining \( \delta_i = -c_{1i} \) (note that these are \( n \)-vectors, whereas \( \delta_0 \) is an \((n-1)\)-vector), the first equation is given by

\[ y_t = \delta_0' z_t + \sum_{i=1}^{p} \delta_i' x_{t-i} + u_{1t}, \] (7)

whereas the remaining \( k \) equations are

\[ z_t = \sum_{i=1}^{p} \Pi_{2i} x_{t-i} + \varepsilon_{2t}. \] (8)

Thus Assumption 1 implies that the system consists of a single structural equation for \( y_t \), completed by \( k \) unrestricted reduced form equations for \( z_t \); this is the classical limited information framework. In the absence of restrictions on \( \sigma_{21} \), identification of the first equation requires at least \( k \) restrictions on \( \delta = (\delta_0', \delta_1', \ldots, \delta_k')' \). We formulate these as

**Assumption 2** The parameter space of \( \delta \) is defined by the linear restrictions

\[ R' \delta = 0, \] (9)
where \( R \) is a \((k + np) \times h\) matrix of full column rank, such that the rank condition

\[
\text{rank}(R'[-I_k : \Pi_{21} : \ldots : \Pi_{2p}]) = k
\]

is satisfied for almost all parameter matrices \( \{\Pi_{2i}\} \).

We focus on homogeneous restrictions for notational ease; non-homogeneous restrictions can be incorporated by redefining \( y_t \). Letting \( H \) denote a \((k + np) \times (k + np - h)\) matrix of full column rank satisfying \( R' H = 0 \), the restrictions can be expressed as

\[
\delta = H \phi, \quad \phi \in \mathbb{R}^{k + np - h},
\]

where \( \phi \) consists of the relevant part of the structural parameter vector \( \phi \). Assumption 2 obviously entails the order condition \( h > k \); if (10) is satisfied and \( h > k \), then the structural equation is over-identified.

In order to analyse weak exogeneity in this model, we require the conditional and marginal distributions corresponding to (3). These are given by (cf. Engle et al., 1983, Lemma 4.2):

\[
y_t | z_t, X_{t-1} \sim N \left( \pi_0' z_t + \sum_{i=1}^{p} \pi_{1,2,i}' x_{t-i}, \omega_{11,2} \right),
\]

\[
z_t | X_{t-1} \sim N \left( \sum_{i=1}^{p} \Pi_{2,i} x_{t-i}, \Omega_{22} \right),
\]

where \( \pi_0 = \Omega_{22}^{-1} \omega_{21}, \pi_{1,2,i} = \pi_{1,i} - \Pi_{2,i} \pi_0, \) and \( \omega_{11,2} = \omega_{11} - \omega_{21} \Omega_{22}^{-1} \omega_{21} \). Defining

\[
\varepsilon_{1,2,t} \equiv y_t - \mathbb{E}[y_t | z_t, X_{t-1}] = \varepsilon_{1,t} - \pi_{1,t}' \varepsilon_{2,t},
\]

we obtain the conditional model of \( y_t \) given \( z_t \):

\[
y_t = \pi_0' z_t + \sum_{i=1}^{p} \pi_{1,2,i}' x_{t-i} + \varepsilon_{1,2,t}.
\]

Notice the similarity between the conditional model and the structural equation (7); both equations have the same dependent and explanatory variables. However, the important difference is that, whereas (12) is derived from the conditional distribution \( D_{y_t}(y_t | z_t, X_{t-1}; \theta) \), the structural equation (7) represents the distribution \( D(y_t - \delta_0' z_t | X_{t-1}; \phi, \sigma_1) \). This implies that \( u_{1,t} \) is uncorrelated with \( X_{t-1} \), but not necessarily with \( z_t \); it is easily seen that \( \mathbb{E}[u_{1,t} | z_t] = \sigma_{21} \). On the other hand, the covariance between the \( \varepsilon_{1,2,t} \) and \( z_t \) is zero by construction.

A second difference is that the parameters \( (\pi_0, \{\pi_{1,2,i}\}, \omega_{11,2}) \) of the conditional model are unrestricted, and variation independent of the marginal model parameters \( \{\Pi_{2i}\}, \Omega_{22} \). Thus, if the parameters of interest are a function of \( (\pi_0, \{\pi_{1,2,i}\}, \omega_{11,2}) \) only, then \( z_t \) is weakly exogenous for those parameters by construction. In contrast, the structural parameters \( \delta_0 \) and \( \{\delta_i\} \) are subject to the restrictions (9), and not variation independent of \( \Pi_{2i} \), because of their definition in (5).

A specific case where the two equations (and the corresponding distributions) coincide is when instead of (9), the identifying restrictions are \( \sigma_{21} = 0 \), or equivalently when the parameters \( \delta_0 \) are chosen to equal \( \pi_0 \). Then the structural model is said to be block-recursive, and \( z_t \) is weakly exoge-
nous for \( \delta \). The following result, which is a special case of Theorem 4.1 of Engle et al. (1983), shows that the same covariance restriction is required when further restrictions of the form (9) are imposed.

**Theorem 1** In the simultaneous equations model defined by Assumptions 1 and 2, \( z_t \) is weakly exogenous for the parameters of interest \( \psi \) under the following set of conditions:

(i) \( \sigma_{21} \in \{0\} \);
(ii) \((\phi_1, \sigma_{11})\) and \((\{\Pi_{21}\}, \Omega_{22})\) are variation free;
(iii) \( \psi \) is a function of \((\phi_1, \sigma_{11})\).

The three conditions of the theorem are all statements about parameter spaces. In particular, the first condition restricts the parameter space of \( \sigma_{21} \) to the single point \( \{0\} \). If the model satisfies conditions (ii) and (iii), with \( \sigma_{21} \) unrestricted, then weak exogeneity holds only in the sub-model defined by \( \sigma_{21} = 0 \), and hence a test for the reduction of the general model to this sub-model is a test for weak exogeneity. In a similar fashion, one could have circumstances where (ii) is violated in the general model but satisfied in a sub-model (the converse is also possible), but this will not be considered explicitly here. Finally, whether the parameters of interest are defined solely from \((\phi_1, \sigma_{11})\) depends upon the issue to be addressed.

It should be emphasized that condition (i) is testable only if the parameter vector \( \delta \) is just- or over-identified according to the rank condition (10), i.e., without any restriction on \( \sigma_{21} \). As we have discussed above, in a recursive system \( \sigma_{21} = 0 \) is the only identifying restriction, so that weak exogeneity is imposed and hence not testable. If (10) is satisfied, then we may construct a test for \( \sigma_{21} = 0 \), and hence for weak exogeneity, as will be discussed in the next section. Notice that this condition is the same as the conventional predeterminedness or orthogonality restriction.

A parameterization that will be particularly convenient is the following. From the fact that \( u_t \) is a Gaussian mean innovation process relative to \( X_{t-1} \) with constant covariance matrix \( \Sigma \), we have

\[
\mathbb{E}[u_{1t}|z_t, X_{t-1}] = \sigma_{21}' \Sigma_{22}^{-1} u_{2t}, \quad \mathbb{V}[u_{1t}|z_t, X_{t-1}] = \sigma_{11.2} = \sigma_{11} - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21}.
\]

Thus we obtain

\[
y_t = \delta_0' z_t + \sum_{i=1}^{p} \delta_i' x_{t-i} + \tau' u_{2t} + u_{1,2,t}
\]

\[
= \delta_0' z_t + \sum_{i=1}^{p} \delta_i' x_{t-i} + \tau' \left( z_t - \sum_{i=1}^{p} \Pi_{2,i} x_{t-i} \right) + u_{1,2,t},
\]

where \( \tau = \Sigma_{22}^{-1} \sigma_{21} \) and \( u_{1,2,t} = u_{1t} - \tau' u_{2t} \), a Gaussian mean innovation process relative to \((z_t, X_{t-1})\) with constant variance \( \sigma_{11.2} \). From the definition of \( \tau \) it follows that weak exogeneity holds if \( \tau = 0 \). This is only a testable restriction if \( \tau \) is identified in (13), which in turn requires that the restrictions (9) identify \( \delta \). Otherwise the vector of regressors in (13) would become collinear, either perfectly or asymptotically, since \( u_{2t} \) consists of linear combinations of \((z_t', x_{t-1}', \ldots, x_{t-p}')\).
2.2 Weak Exogeneity in Error Correction Models

Let us again start from the Gaussian VAR($p$) model for an $n$-vector time series $\{x_t, t = 1, \ldots, T\}$:

$$x_t = \sum_{i=1}^{p} \Pi_i x_{t-i} + \epsilon_t, \quad t = 1, \ldots, T,$$

where $\epsilon_t$ is a homoskedastic Gaussian mean innovation process relative to $X_{t-1}$ with covariance matrix $\Omega$. Let $L$ denote the lag operator, so that $L^i x_t = x_{t-i}, \ i \in \mathbb{N}$, and define the matrix lag polynomial $\Pi(L) = I_n - \sum_{i=1}^{p} \Pi_i L^i$, so that the model can be expressed as $\Pi(L) x_t = \epsilon_t$.

Assumption 3 (Cointegration) In the VAR model (14),

(i) $\Pi(1) = -\alpha \beta'$, where $\alpha$ and $\beta$ are $n \times r$ matrices of rank $r$, $0 < r < n$;

(ii) $|\Pi(\zeta)|$ has $n - r$ roots equal to 1 and all other roots outside the unit circle.

This assumption implies (see Johansen, 1991, and Boswijk, 1992) that the process $x_t$ is cointegrated of order $(1,1)$, cf. Engle and Granger (1987), i.e., $x_t$ is integrated of order 1 ($x_t \sim I(1)$), but with $r$ stationary linear combinations $\beta' x_t$. The columns of $\beta$ span the space of cointegrating vectors, and the elements of $\alpha$ are the corresponding adjustment coefficients or factor loadings. A convenient decomposition of the matrix lag polynomial $\Pi(L)$ is

$$\Pi(L) = \Pi(1)L + \left(I_n - \sum_{i=1}^{p} \Gamma_i L^i\right)(1 - L),$$

where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$. Defining the first-difference operator $\Delta = (1 - L)$, this leads to the vector error correction model (VECM):

$$\Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t.$$  

(15)

Partition $x_t$ as $(y_t, z_t)'$ so that $y_t$ is a scalar time series and $z_t$ is a $k$-vector time series ($k = n - 1$). Partitioning $\epsilon_t, \alpha, \Gamma_i$ and $\Omega$ again conformably with $x_t$, the conditional and marginal error correction models are given by

$$\Delta y_t = \alpha_{1,2} \beta' z_{t-1} + \gamma_{0,1} \Delta z_t + \sum_{i=1}^{p-1} \gamma_{1,2,i} \Delta x_{t-i} + \epsilon_{1,2,t},$$

(16)

$$\Delta z_t = \alpha_{2,2} \beta' z_{t-1} + \sum_{i=1}^{p-1} \Gamma_{2,i} \Delta x_{t-i} + \epsilon_{2,t},$$

(17)

where $\gamma_0 = \Omega_{2,1}^{-1} \omega_{2,1} \epsilon_{1,2,t} + \gamma_{1,2,t}$ (with variance $\omega_{1,1} = \omega_{1,1} - \omega_{2,1} \Omega_{2,1}^{-1} \omega_{2,1}$), $\alpha_{1,2} = \alpha_{1,2} - \gamma_0 \alpha_{2,1}$ and $\gamma_{1,2,t} = \gamma_{1,2,t} - \Gamma_{2,1} \gamma_0$. Notice that although $\beta'$ is an $r \times n$ matrix, $\alpha_{1,2} \beta'$ is an $1 \times n$ vector. Thus only one cointegrating vector is identifiable from the single-equation conditional model; if more than one cointegrating vectors are to be analysed, then the dimension of the conditional model should be increased correspondingly. However, here we confine ourselves to the following case:
Assumption 4 In addition to Assumption 3, \( r = 1 \) and \( \beta \) can be normalized as \( \beta = (1, -\kappa')' \).

Notice that cointegrating vectors are only unique up to non-singular linear transformations, so that if \( r = 1 \), some normalization is required for identification. Assumption 4 implies \( \beta' \varrho_t = y_t - \kappa' \varrho_t \), so that we exclude the case where the first component of \( \beta \) equals zero, in which case \( y_t \) would not enter the cointegrating relationship.

Assume that the parameters of interest are given by the adjustment parameter and the (unique components of the) cointegrating vector in the conditional error correction model, i.e., \( \psi = (\alpha_{1,2}, \kappa) \). It is easily shown that the conditional model parameters \( (\alpha_{1,2}, \gamma_0, \{ \gamma_{1,2}, \}, \omega_{1,2}) \) are variation independent of the marginal model parameters \( (\alpha_2, \{ \Gamma_2 \}, \Omega_{22}) \). However, the vector of cointegration parameters enters both the conditional and marginal model, leading to a violation of the variation independence requirement, unless \( \alpha_2 \) is restricted to zero. This leads to the following theorem, cf. Boswijk (1992), Johansen (1992) and Urbain (1992):

**Theorem 2** In the VECM (15), and under Assumptions 3 and 4, \( \varrho_t \) is weakly exogenous for \( (\alpha_{1,2}, \kappa) \) if \( \alpha_2 \in \{0\} \).

Extensions of Theorem 2 to cases with more than one cointegrating vector and multiple-equation conditional models have been obtained by Johansen (1992), Boswijk (1992), and Hendry and Mizon (1993). Observe that, in contrast to the previous section, predeterminedness of \( \varrho_t \) is a maintained assumption. The consequences of a violation of weak exogeneity are analysed by Johansen (1992) and Hendry (1995). Although the maximum likelihood (ML) estimator of \( \kappa \) is still consistent if \( \alpha_2 = 0 \) is incorrectly imposed, it is inefficient, and hypothesis test statistics will no longer have standard (normal or \( \chi^2 \)) asymptotic null distributions.

The condition \( \alpha_2 = 0 \) is sufficient for weak exogeneity of \( \varrho_t \) for all parameters of the conditional model. However, if we consider an error correction model in structural form, then additional conditions are required. Such a system is obtained by pre-multiplication of the VAR model by a particular \( g \times n \) matrix \( B \), yielding

\[
B \Delta \varrho_t + \sum_{i=1}^{p-1} C_i \Delta \varrho_{t-i} + \alpha \beta' \varrho_{t-1} = \varrho_t,
\]

where \( \varrho_t \) is again a homoskedastic Gaussian mean innovation process relative to \( X_{t-1} \) with covariance matrix \( \Sigma \). The parameter matrices \( B, C_i, \alpha \) and \( \beta \) are functions of the vector of structural parameters \( \phi \), and are subject to identifying restrictions, so that they are uniquely related to the parameters of the VECM via

\[
B \Gamma_i + C_i = 0, \quad i = 1, \ldots, p - 1, \quad B \alpha + \alpha = 0, \quad \Sigma = B \Omega B'.
\]

If we confine ourselves again to the case defined by Assumption 1, so that only the first equation of (18) is of interest, we obtain the single-equation error correction model in structural form. Letting \( \alpha = -\lambda (\alpha_2')' \), this equation is given by

\[
\Delta y_t = \lambda (y_{t-1} - \kappa' \varrho_{t-1}) + \delta' \Delta \varrho_t + \sum_{i=1}^{p-1} \delta_i (\Delta \varrho_{t-i}) + \varrho_{1,t}.
\]
where \( \delta = (\delta_0, \delta_1, \ldots, \delta_{p-1}) \) is again subject to identifying restrictions \( R' \delta = 0 \), or equivalently \( 6 = H \phi_1 \), satisfying the rank condition \( \text{rank}(R'[-I_k : \Gamma_{21} : \cdots : \Gamma_{2,p-1}]) = k \). The system is completed by the marginal error correction model (17). The next theorem from Urbain (1992) gives sufficient conditions for weak exogeneity in this model.

**Theorem 3** In the error correction model in structural form, defined by Assumptions 1–4, \( z_t \) is weakly exogenous for the parameters of interest \( \psi \) under the following set of conditions:

(i) \( \sigma_{21} \in \{0\} \);

(ii) \( \alpha_2 \in \{0\} \);

(iii) \( (\lambda, \kappa, \phi_1, \sigma_{11}) \) and \( \{\Omega_{21}\}, \Omega_{22} \) are variation free;

(iv) \( \psi \) is a function of \( (\lambda, \kappa, \phi_1, \sigma_{11}) \).

Observe that this theorem is in effect a synthesis of Theorems 1 and 2. Similarly, we shall see in the next section that the Lagrange-multiplier statistic for weak exogeneity in this model is a blend of the traditional orthogonality \( (\sigma_{21} = 0) \) test statistic and a statistic for the hypothesis of no error correction in the marginal system \( (\alpha_2 = 0) \).

Condition (iii) will always be fulfilled if the parameters of the marginal system are unrestricted. Dolado (1992) analyses the possibility that variation independence and hence weak exogeneity is violated, even if \( \sigma_{21} = 0 \) and \( \alpha_2 = 0 \). Consider the following bivariate model:

\[
\begin{align*}
\Delta y_t &= \lambda (y_{t-1} - \kappa z_{t-1}) + \delta_0 \Delta z_t + u_{1t}, \\
\Delta z_t &= \gamma_{21} \Delta y_{t-1} + \gamma_{22} \Delta z_{t-1} + \varepsilon_{2t}.
\end{align*}
\]

If we impose the restriction \( \gamma_{22} = -\kappa \gamma_{21} \), so that \( \Delta z_t \) only depends on \( \Delta (y_{t-1} - \kappa z_{t-1}) \), then \( z_t \) is no longer weakly exogenous for \( \kappa \). However, Dolado argues that the (asymptotic) consequences of this violation are much less severe than if \( \alpha_2 \neq 0 \). Therefore, he defines a weaker form of exogeneity, called long-run weak exogeneity. An interpretation of these results is provided by Hendry (1995).

### 3 The Lagrange-Multiplier Test

In this section we derive the Lagrange-multiplier (LM) statistic for weak exogeneity for the case analysed in Theorem 3. As in the previous section, we shall focus on the simplest case, i.e., a single structural equation and a single cointegrating relationship. Extensions to multivariate systems can be (and partly have been) derived analogously, and will be briefly discussed in the final section.

Consider the conditional and marginal model, parameterized similarly as in (13):

\[
\begin{align*}
\Delta y_t &= \phi_2 w_{2t} + \lambda (y_{t-1} - \kappa' z_{t-1}) + \tau' u_{2t} + u_{1;2,t}, \\
\Delta z_t &= \Gamma_2 w_{2t} + \alpha z_{t-1} + \omega z_{t-1} + \varepsilon_{2t},
\end{align*}
\]

where \( w_{1t} = H'((\Delta z_t', \Delta z_{t-1}', \ldots, \Delta z_{t-p+1}')\) is the vector of included differenced regressors in the structural equation, with (unrestricted) parameter vector \( \phi_1 \) (cf. (11)); where \( w_{2t} = (\Delta z_{t-1}', \ldots, \Delta z_{t-p+1}')\).
$\Delta \mathbf{x}_{t-p+1}'$ is the corresponding regressor vector in the marginal model, with coefficient matrix $\Gamma_2 = (\Gamma_{21}, \ldots, \Gamma_{2,p-1})$; and where $\tau = \Sigma^{-1}_{22}\sigma_{21}$ (recall that $u_{2t} = \varepsilon_{2t}$). The null hypotheses are

$$H_0^\tau : \tau = 0; \quad (\text{orthogonality})$$

$$H_0^{\alpha_2} : \alpha_2 = 0; \quad (\text{no error correction})$$

$$H_0 = H_0^\tau \cap H_0^{\alpha_2} : \tau = 0, \alpha_2 = 0.$$ Letting $H_1^\tau : \tau \neq 0$ and $H_1^{\alpha_2} : \alpha_2 \neq 0$ denote the partial alternative hypotheses, we are interested in testing the joint null hypothesis $H_0$ against the joint alternative $H_1 : H_1^\tau \cup H_1^{\alpha_2}$. However, we shall also discuss the partial testing problems of $H_0$ against $H_1^\tau$ or against $H_1^{\alpha_2}$.

Let variables without a subscript $t$ denote the corresponding data vectors or matrices (the latter in upper case), each consisting of $T$ rows. Then the model in matrix notation becomes

$$\Delta \mathbf{y} = W_1 \phi_1 + \mathbf{y}_1 - Z_{-1} \kappa \lambda + U_2 \tau + u_{1,2}$$

$$\Delta \mathbf{Z} = W_2 \phi_1' + \mathbf{y}_1 - Z_{-1} \kappa \lambda + [\Delta \mathbf{Z} - W_2 \phi_1' - \mathbf{y}_1 - Z_{-1} \kappa \lambda] \alpha_2' \tau + u_{1,2},$$

The vector of error correction terms $(\mathbf{y}_1 - Z_{-1} \kappa)$ will be denoted by $u_{-1}$, and further $\Delta \mathbf{z} = \text{vec}(\Delta \mathbf{Z})$ and $\gamma_2 = \text{vec}(\Gamma_{21})$. Because $u_{1,2,t} \sim \text{IN}(0, \sigma_{11,2})$, independently of $u_{2t} \sim \text{IN}(0, \Sigma_{22})$, the log-likelihood is given by

$$\ln L(\theta) = c - \frac{T}{2} \ln \sigma_{11,2} - \frac{1}{2\sigma_{11,2}} u_{1,2}' u_{1,2} - \frac{T}{2} \ln |\Sigma_{22}| - \frac{1}{2} u_{2,2}' (\Sigma_{22}^{-1} \otimes I_T) u_{2,2}, \quad (20)$$

where $c$ is a constant and $u_{2,2} = \text{vec}(U_2)$. Here $u_{1,2}$ and $u_{2,2}$ should be understood as functions of the data and parameters.

For the Lagrange-multiplier statistic, we require the score vector $q$, i.e., the vector of first derivatives of the log-likelihood, and the Hessian matrix $Q$. Because the information matrix satisfies the usual block-diagonality with respect to the variance parameters $(\sigma_{11,2}, \Sigma_{22})$ and the other parameters, the derivatives with respect to the former will not be given here. Let $\theta_1 = (\phi_1', \lambda, \tau', \kappa')'$ and $\theta_2 = (\gamma_2', \alpha_2')'$, and let $q_{ij}(\theta) \equiv \partial \ln L(\theta)/\partial \theta_i$ and $Q_{ij}(\theta) \equiv \partial^2 \ln L(\theta)/\partial \theta_i \partial \theta_j$, $i, j = 1, 2$; we shall often suppress the argument $\theta$ below. Matrix differential calculus (see Magnus and Neudecker, 1988) yields the following result for the score vector $q = (q_1, q_2)'$:

$$q_1 = \begin{pmatrix}
\partial \ln L/\partial \phi_1 \\
\partial \ln L/\partial \lambda \\
\partial \ln L/\partial \tau \\
\partial \ln L/\partial \kappa
\end{pmatrix} = \begin{pmatrix}
\sigma_{11,2}^{-1} W_1' u_{1,2} \\
\sigma_{11,2}^{-1} v_{1,1}' u_{1,2} \\
\sigma_{11,2}^{-1} U_2' u_{1,2} \\
(\lambda - \tau' \alpha_2) \sigma_{11,2}^{-1} Z_{-1}' u_{1,2} - (\alpha_2' \Sigma_{22}^{-1} \otimes Z_{-1}' ) u_{2,2}
\end{pmatrix},$$

$$q_2 = \begin{pmatrix}
\partial \ln L/\partial \gamma_2 \\
\partial \ln L/\partial \alpha_2
\end{pmatrix} = \begin{pmatrix}
\sigma_{11,2}^{-1} (\tau \otimes W_2' ) u_{1,2} + (\Sigma_{22}^{-1} \otimes W_2' ) u_{2,2} \\
\sigma_{11,2}^{-1} (\tau \otimes v_{-1}' ) u_{1,2} + (\Sigma_{22}^{-1} \otimes v_{-1}' ) u_{2,2}
\end{pmatrix}.$$
LAGRANGE-MULTIPLIER TESTS FOR WEAK EXOGENEITY

\[ Q_{11} = -\sigma_{1112}^{-1} \begin{bmatrix} W'_1 W_1 & W'_1 U_2 & W'_1 Z_{-1} \nu \\ v'_1 v_{-1} & v'_1 U_2 & v'_1 Z_{-1} \nu + u'_{12} Z_{-1} \\ U'_2 U_2 & U'_2 Z_{-1} \nu - \alpha_2 u'_{12} Z_{-1} & Z'_{-1} Z_{-1} [\nu^2 + \sigma_{112} (\alpha_2^2 \Sigma_{22}^{-1} \alpha_2)] \end{bmatrix} \]

where \( \nu = -(\lambda - \tau' \alpha_2) \), and the lower-diagonal elements follow from symmetry. Next,

\[ Q_{12} = \sigma_{1112}^{-1} \begin{bmatrix} \tau' \otimes W'_1 W_2 & W'_1 v_{-1} \tau' \\ \tau' \otimes U'_1 W_2 & U'_1 v_{-1} \tau' - u'_{12} v_{-1} I_6 \\ \xi' \otimes Z'_{-1} W_2 & Z'_{-1} v_{-1} \xi' + \tau u'_{12} Z_{-1} - \sigma_{112} \Sigma_{22}^{-1} U'_2 Z_{-1} \end{bmatrix} \]

where \( \xi' = \nu \tau' + \sigma_{112} \alpha_2^2 \Sigma_{22}^{-1} \). Finally,

\[ Q_{22} = -(\Sigma_{22}^{-1} + \sigma_{112} \tau \tau') \otimes \begin{bmatrix} W'_2 W_2 & W'_2 v_{-1} \\ v'_1 W_2 & v'_1 v_{-1} \end{bmatrix} \]

Let \( \hat{\theta} \) denote the restricted \( (\tau = 0, \alpha_2 = 0) \) ML estimator. The Lagrange-multiplier statistic is given by \( LM = q(\hat{\theta})' V[\hat{\theta}] q(\hat{\theta}) \), where \( V[\hat{\theta}] \) is the inverse of either Fisher's expected information matrix, evaluated at \( \hat{\theta} \), or of the observed information matrix \( -Q(\hat{\theta}) \), or of any asymptotically equivalent matrix. Here we consider

\[ LM = -q(\hat{\theta})' Q_{\hat{\theta}}(\hat{\theta})^{-1} q(\hat{\theta}), \] (21)

where

\[ Q_{\hat{\theta}}(\hat{\theta}) = \begin{bmatrix} Q_{11}(\hat{\theta}) & 0 \\ 0 & Q_{22}(\hat{\theta}) \end{bmatrix} \]

It can be shown that \( Q_{\hat{\theta}} \) is asymptotically equivalent to \( Q \). Moreover, the expected information satisfies the same block-diagonality as \( Q_{\hat{\theta}} \). The motivation for dropping the asymptotically negligible term \( Q_{12} \) is that it yields a useful decomposition of the LM statistic:

\[ LM = -q_1(\hat{\theta})' Q_{11}(\hat{\theta})^{-1} q_1(\hat{\theta}) - q_2(\hat{\theta})' Q_{22}(\hat{\theta})^{-1} q_2(\hat{\theta}) = LM_1 + LM_2. \] (22)

In order to obtain an expression for \( LM_1 \) and \( LM_2 \), let us first analyse \( \hat{\theta} \). Under the restrictions \( \tau = 0 \) and \( \alpha_2 = 0 \), the ML estimators of \( \phi_1, \lambda \) and \( \kappa \) are obtained from the score equations

\[ \hat{\sigma}_{112}^{-1} \begin{bmatrix} W'_1 \\ (y_{-1} - Z_{-1} \hat{\kappa})' \\ -\hat{\lambda} Z_{-1}^{'} \end{bmatrix} [\Delta y - W_1 \hat{\phi}_1 - (y_{-1} - Z_{-1} \hat{\kappa})] = 0. \]

Defining \( \beta = (\beta_1, \beta_2)' = (\lambda, -\lambda \kappa)' \), it is easily shown that \( \hat{\lambda} = \hat{\beta}_1 \) and \( \hat{\kappa} = -\hat{\beta}_2 / \hat{\beta}_1 \), where \( \hat{\phi}_1 \) and \( \hat{\beta} \) are the ordinary least-squares estimators in

\[ \Delta y = W_1 \phi_1 + X_{-1} \beta + u_{1,2}, \] (23)
with \( X_{-1} = [y_{-1}, Z_{-1}] \). Moreover, \( \hat{\sigma}_{11.2} = T^{-1} \hat{u}_{1.2}' \hat{u}_{1.2} \), where \( \hat{u}_{1.2} = \Delta y - W_1 \hat{\phi}_1 - X_{-1} \hat{\beta} \).

For the marginal model, the restricted estimate of \( \gamma_2 \) is obtained by applying least-squares to

\[
\Delta z = [I_k \otimes W_2] \gamma_2 + u_2,
\]

yielding the residuals \( \hat{u}_2 = \text{vec}(\hat{U}_2) \), and the variance estimator \( \hat{\Sigma}_{12} = T^{-1} \hat{U}_2' \hat{U}_2 \).

Inserting these estimates (as well as \( \tau = 0 \) and \( \alpha_2 = 0 \)) in \( q_1 \) and \( Q_{11} \) yields

\[
q_1(\hat{\theta}) = \begin{pmatrix}
0 \\
0 \\
\hat{\sigma}_{11.2}^{-1} \hat{U}_2' \hat{u}_{1.2}
\end{pmatrix},
\]

\[
Q_{11}(\hat{\theta}) = -\hat{\sigma}_{11.2}^{-1} \begin{pmatrix}
W_1' W_1 & W_1' \hat{\theta}_{-1} & W_1' \hat{U}_2 & -W_1' Z_{-1} \hat{\lambda} \\
\hat{\theta}_{-1} W_1 & \hat{\theta}_{-1} \hat{\theta}_{-1} & \hat{\theta}_{-1} \hat{U}_2 & -\hat{\theta}_{-1} Z_{-1} \hat{\lambda} \\
\hat{U}_2' W_1 & \hat{U}_2' \hat{\theta}_{-1} & \hat{U}_2' \hat{U}_2 & -\hat{U}_2' Z_{-1} \hat{\lambda} \\
-\hat{\lambda} Z_{-1} W_1 & -\hat{\lambda} Z_{-1} \hat{\theta}_{-1} & -\hat{\lambda} Z_{-1} \hat{U}_2 & Z_{-1} \hat{\lambda}^2
\end{pmatrix},
\]

where \( \hat{\theta}_{-1} = (y_{-1} - Z_{-1} \hat{\kappa}) \), the estimated vector of error correction terms. For an arbitrary \( T \times k \) matrix \( A \) of full column rank, let \( M(A) = I_T - A(A'A)^{-1} A' \). Note that \( M(W_1, \hat{\theta}_{-1}, -\hat{\lambda} Z_{-1}) = M(W_1, X_{-1}) \), because \( X_{-1} \) and \( [\hat{\theta}_{-1}, -\hat{\lambda} Z_{-1}] \) span the same column space. Since \( \hat{u}_{1.2} = M(W_1, X_{-1}) \Delta y \), we find

\[
LM_1 = \frac{\Delta y'M(W_1, X_{-1}) \hat{U}_2' \hat{U}_2 M(W_1, X_{-1}) \hat{U}_2' M(W_1, X_{-1}) \Delta y}{\hat{\sigma}_{11.2}^{-1}}
\]

\[
= \hat{\tau}' \left( \hat{V}[\hat{\tau}] \right)^{-1} \hat{\tau},
\]

where \( \hat{\tau} \) is the least-squares estimate in the augmented regression

\[
\Delta y = W_1 \hat{\phi}_1 + X_{-1} \hat{\beta} + \hat{U}_2 \tau + \hat{u}_{1.2},
\]

and \( \hat{V}[\hat{\tau}] = \hat{\sigma}_{11.2}^{-1} \hat{U}_2' M(W_1, X_{-1}) \hat{U}_2 \hat{\sigma}_{11.2}^{-1} \) is an estimate of its variance (based on the restricted error variance estimate). Therefore, \( LM_1 \) is a variable-addition test statistic for marginal model residuals \( \hat{U}_{-1} \) in the structural equation (23). This is exactly the orthogonality test for \( \sigma_{12} = 0 \) derived by, e.g., Holly (1985), who shows that, up to a factor of proportionality, this statistic is numerically identical to the Durbin-Wu-Hausman statistic.

If in the regression (26) the dependent variable \( \Delta y \) is replaced by \( \hat{u}_{1.2} \), then \( LM_1 = TR^2 \), with \( R^2 \) the coefficient of determination. Replacing \( \hat{\sigma}_{11.2} \) by \( \hat{\sigma}_{11.2} \), the degrees-of-freedom corrected residual variance in (26), and dividing by \( k \) yields the statistic \( LM_{F_1} \), i.e., the conventional \( F \)-statistic for \( \tau = 0 \) in (26), calculated by standard regression packages. Using the general results on regressions with integrated processes, see, e.g., Park and Phillips (1989), it can be shown that under \( H_0 \), \( LM_1 \) has an asymptotic \( \chi^2(k) \) distribution, so that \( LM_{F_1} \) may be compared with critical values from the \( F(k, T - \ell_i) \) distribution (where \( \ell_i \) is the dimension of \( \theta_i, i = 1, 2 \)).

Insertion of the restricted MLE in \( q_2 \) and \( Q_{22} \) yields

\[
LM_2 = \hat{u}_{2} \left( \hat{\Sigma}_{22}^{-1} \otimes \hat{\theta}_{-1} \right) \left[ \hat{\Sigma}_{22}^{-1} \otimes \hat{\theta}_{-1} \right]^{-1} \left( \hat{\Sigma}_{22}^{-1} \otimes \hat{\theta}_{-1} \right) \hat{u}_{2}
\]
where $\Delta z$ is the multivariate least-squares estimate in the augmented regression and $V(\tilde{\alpha}_2) = \begin{bmatrix} \hat{\Sigma}_1^{-1} & \hat{\Sigma}_2^{-1} \end{bmatrix}^{-1}$ is its covariance matrix estimate (again based on the restricted error covariance matrix estimator). If we replace $\hat{\Sigma}_2$ in (27) by $\hat{\Sigma}_2$, the covariance matrix estimator from (28), we obtain a simple Wald-type variable addition statistic for $\alpha_2 = 0$ in (28). A generalization of this test is considered by Boswijk (1992, 1995) and Johansen (1992). Because $\hat{k}$ is super-consistent, this statistic is asymptotically equivalent to a variable addition statistic with $K$ known, i.e., with $\hat{v}_{-1}$ replaced by $v_{-1}$. Because both $W_2$ and $v_{-1}$ are stationary, this implies that $LM_2$ has an asymptotic $\chi^2(k)$ distribution under the null hypothesis. An $F$-version is again obtained if $\hat{V}$ involves a degrees of freedom correction, and the resulting statistic is divided by $k$. This $LM_2$ statistic may be compared with critical values from the $F(k, T-C_2)$ distribution.

From the block-diagonality of the information matrix, it follows that the two estimated scores $q_1(\hat{\theta})$ and $q_2(\hat{\theta})$ are asymptotically independent, so that the two $LM$ statistics have asymptotically independent $\chi^2(k)$ distributions under the null hypothesis. This in turn implies that $LM$ has an asymptotic $\chi^2(2k)$ null distribution. There is no obvious $F$-version of this joint test.

A similar decomposition and independence result was obtained by Smith (1994), who analysed classical tests for weak exogeneity in a general class of simultaneous equations models. In his set-up, weak exogeneity requires orthogonality ($\Sigma_1 = 0$) and no feedback from $y$ to $z$ ($B_2 = 0$). Smith showed that these two hypotheses are separable, which entails that under the null hypothesis (as well as local alternatives), the information matrix is block-diagonal. Separability implies that a classical test of the joint null hypothesis against the union of alternatives can be decomposed into the sum of statistics for the separate hypotheses. In our model, the block-diagonality between the conditional model parameters $\theta_1$ and the marginal model parameters $\theta_2$ shows that our joint hypothesis $H_0$ is indeed separable into its constituents $H_0^1$ and $H_0^2$. This explains the decomposition (22) and the associated independence of $LM_1$ and $LM_2$.

In summary, weak exogeneity of the conditioning variables for the parameters of a structural error correction model can be tested as follows. First, estimate the structural model (23) and marginal model (24) by (multivariate) least-squares. Construct the residuals $u_{2t}$ from the marginal model and the lagged disequilibrium errors $\hat{v}_{t-1} = (y_{t-1} - \hat{k}z_{t-1})$ from the structural model. Next, calculate $LM_1$, the variable-addition test statistic for the significance of $u_{2t}$ in the structural equation, and the statistic $LM_2$ for the significance of $\hat{v}_{t-1}$ in the marginal model. If exogeneity for the long-run parameters is maintained, and only exogeneity for the short-run parameters is to be tested, then $LM_1$ is to be compared with critical values from the $\chi^2(k)$ distribution. If one wishes to test $\alpha_2 = 0$ while maintaining the orthogonality assumption $\sigma_{21} = 0$, $LM_2$ should be compared with $\chi^2(k)$ quantiles. Finally, the joint hypothesis is tested by comparing $LM = LM_1 + LM_2$ with critical values from the $\chi^2(2k)$ distribution.
4 The Demand for Money in Switzerland

In this section we apply the exogeneity test derived above to a model of the demand for money in Switzerland inspired by the analysis of Fischer and Peytrignet (1991). They consider the effect of changes in the Swiss central bank policy on the (weak and super-) exogeneity status of prices, income and interest rates in a conditional money demand equation. The data used are quarterly observations (1973.1-1989.4) on $m$, the natural logarithm of $M_1 + B$ (defined as $M_1$ plus savings deposits); $p$, the natural logarithm of the CPI; $y$, the natural logarithm of real GDP; $R'$, the long-term (Bundesobligation) rate; and $R^*$, the three-month Euro (Swiss franc) rate. The variables $m$, $p$ and $y$ have been seasonally adjusted.

Figure 1 displays the data. In the bottom left graph, the close connection between velocity $(p + y - m)$ and the long-term interest rate is evident (the location and scale of $R'$ and velocity have been matched). The bottom right graph shows how large movements in $R^*$ are often reflected in the (yearly-based) inflation rate $infl = 4\Delta p$.

The full vector of variables is defined as $\mathbf{x}_t = ([m - p], p, y, R'_t, R^*_t)'$. We start with a multivariate cointegration analysis in a fifth-order VAR model of $\mathbf{x}_t$, using Johansen’s (1991) likelihood ratio tests. An unrestricted constant term is included (allowing for a drift in the integrated series), as well as a dummy variable $D_t$ which takes on the value 1 in the first quarter of 1988, and -1 in the same quarter a year later; see Fischer and Peytrignet (1991) for an explanation of an outlier in the
LAGRANGE-MULTIPLIER TESTS FOR WEAK EXOGENEITY

first quarter of 1988 (the outlier a year later may be caused by seasonal adjustment). We assume that \( x_t \) is at most integrated of order 1. Price series such as \( p_t \) sometimes appear to be I(2), but univariate analysis suggests that this can be rejected here. All computations are performed using *PcGive Professional Version 8.0*, see Doornik and Hendry (1994).

Both single-equation and multivariate diagnostic tests do not indicate any misspecification of the original VAR(5) specification at the 5% level; a lower-order VAR does suffer from serial correlation. Next, we estimate the cointegrating rank using van Giersbergen's (1994) bootstrap procedure to obtain tail probabilities of Johansen's likelihood ratio (LR) trace statistic. The LR statistic for the hypothesis \( r = 0 \) equals 127.7, which has a p-value of 0.013, so that this hypothesis can be rejected (at the 5% level). On the other hand, the LR statistic for \( r \leq 1 \) has a p-value of 0.253, so we cannot reject this hypothesis at the 5% level, and we end up with \( \hat{r} = 1 \).

The next step in the analysis is the specification of the structural error correction equation. An initial single-equation ECM with four lagged differences of all variables, a constant and the dummy variable \( D_t \), appears to be well-specified. A subsequent reduction process leads to the following, more parsimonious error correction model (standard errors between parentheses):

\[
\Delta(m-p)_t = -1.13 + 0.02 D_t + 0.28 \Delta(m-p-y)_{t-1} - 0.54 \Delta p_t \\
- 0.17 \left\{ (m-p-y)_{t-1} + 0.16 p_{t-1} - 0.38 y_{t-1} - 7.25 R^2_t - 2.50 R^3_t \right\},
\]

(29)

Here \( p_1 \) denote p-values of diagnostic test statistics against first- and fourth-order serial correlation, fourth-order ARCH, non-normality, heteroskedasticity and functional form misspecification (RESET). The F-test for the hypothesis that (29) parsimoniously encompasses the initial single-equation ECM has a p-value of 0.084. Note that the interest rates appear in the error correction mechanism, but without a lag; this is due to a restriction on the short-run dynamics. The model is remarkably close to the one obtained for the UK by Hendry and Ericsson (1991), which was also found to be quite successful for Germany by Westphal (1991). The price and income homogeneity restriction imposed in those models cannot be rejected here either (with a p-value of 0.08).

For the marginal system, we specify an unrestricted system with four lagged differences of all variables, a constant and \( D_t \). The residual vector from this multivariate regression constitutes \( \hat{u}_{2t} \). The variable-addition F-test for \( \hat{u}_{2t} \) in (29) equals \( LMF_1 = 0.32 \), which has a p-value of 0.86. In addition, none of the individual t-statistics for the components of \( \hat{u}_{2t} \) are significant. Thus, the orthogonality hypothesis cannot be rejected. Next, the variable-addition F-test for \( \hat{v}_{t-1} \) in the marginal system equals \( LMF_2 = 2.17 \), with a p-value of 0.09. Thus, rejection of the hypothesis of no error correction is not as clear-cut as of the orthogonality hypothesis. In fact, if we would have used the \( \chi^2 \) form of the test, then we would have rejected the null hypothesis, since \( LM_2 = 14.79 \) has a p-value of 0.005. Since \( LM_1 = 1.62 \), the overall exogeneity LM-statistic becomes \( LM = LM_1 + LM_2 = 16.41 \), which has a p-value of 0.037. Hence, we are led to reject the null hypothesis of weak ex-
ogeneity at the 5% significance level. Given the lack of a degrees-of-freedom correction for $LM$, one might doubt the accuracy of its $p$-value. However, the relatively low $p$-value of $LMF_2$ already indicated that the assumption of no error correction may be troublesome.

In a model similar to (29), but with both interest rates in natural logarithms, Fischer and Peytrignet (1991) could not reject weak exogeneity of the conditioning variables. However, their result was based solely on the orthogonality test, so that implicitly, their parameters of interest were a function of the short-run parameters only. The present empirical analysis therefore illustrates how conclusions regarding the exogeneity status of a set of conditioning variables may change with the choice of the parameters of interest.

5 Discussion

If the LM tests for weak exogeneity analysed in this paper do not reject the null hypothesis, then consistent and asymptotically efficient inference on the cointegration parameters as well as on the short-run parameters is possible in the single-equation structural error correction model. If, on the other hand, weak exogeneity is rejected, then different system-based methods can be employed, depending on which condition is violated.

First, if there is error correction in the marginal system (17), but the orthogonality condition is maintained, then an efficient estimator of the cointegrating vector may be obtained from Johansen’s (1991) procedure. Conditional upon this estimate, the conditional error correction model may then be estimated using least-squares.

Secondly, if $\alpha_2 = 0$ but the orthogonality condition is violated, then either limited-information maximum likelihood (LIML) or instrumental variable-type approximations thereof should be employed to yield consistent estimators of the short-run structural parameters.

Thirdly, if both conditions are violated, then again two possibilities emerge. On the one hand, if the long-run and short-run parameters are to be estimated jointly, then some form of (non-linear) full-information maximum likelihood (FIML) should be used, taking into account the non-linear cross-equation restrictions that arise from having the same error correction term in different equations. A feasible approximation of FIML is again to estimate the cointegrating vector in an unrestricted VAR model, and replace the cointegrating vector in the structural model by this estimate. The remaining parameters may then be estimated by standard linear simultaneous equations procedures. Applications of such a two-step procedure are given in Hendry and Mizon (1993) and Hendry and Doornik (1994).

The test proposed in this paper can easily be extended to allow for more than one structural equation of interest, and for multiple cointegrating vectors; see Johansen (1992) and Hendry and Mizon (1993) for likelihood ratio tests, and Boswijk (1992, 1995) for LM tests. In such cases the number of endogenous variables $m$ should at least equal the number of cointegrating vectors $r$, since otherwise the weak exogeneity restriction $\alpha_2 = 0$ is in conflict with $\text{rank}(\alpha \beta') = r$. A variable-addition test for the estimated error correction terms in the marginal system again serves as a test for weak exogeneity for the cointegration parameters. Provided that sufficient identifying restrictions are imposed on the structural model, orthogonality can be tested by adding the residuals of the marginal system to the structural model and testing their significance.
LAGRANGE-MULTIPLIER TESTS FOR WEAK EXOGENEITY

ACKNOWLEDGEMENTS

We would like to thank Franz Palm, Peter Schotman, and two anonymous referees, for helpful comments; Andreas Fischer, for providing the data; and Jurgen Doornik and David Hendry, for the use of a beta-test version of GiveWin Version 9.0 for constructing the graphs. The first author gratefully acknowledges financial support from the Royal Netherlands Academy of Arts and Sciences and the Tinbergen Institute.

REFERENCES


