

Recent Advances in Cointegration Analysis

by

Helmut Lütkepohl¹
Department of Economics
European University Institute
Via della Piazzuola 43
I-50133 Firenze
ITALY

Abstract

A small system of German economic variables consisting of the money stock M3, Gross National Product (GNP) and a bond rate is used to illustrate the power of cointegration analysis and the usefulness of some recently developed tools for this kind of analysis. Testing for the cointegrating rank and specifying a VECM, estimating the cointegrating relations and other parameters as well as model checking are discussed. The estimated model is analyzed with an impulse response analysis and a forecast error variance decomposition is performed. A quite reasonable long-run money demand relation is found.

Key Words: cointegration, German monetary system, vector error correction model

JEL classification: C32

¹I am grateful to Dmitri Boreiko for helping with the figures. An earlier version of the paper was presented at Venice University. I thank the seminar participants for their comments.

1 Introduction

The cointegration framework has been developed rapidly over the last years. Its fast progress is to a large extent due to its usefulness for applied work. Cointegration is a concept for modelling equilibrium or long-run relations of economic variables. Many economic issues have been reanalyzed using the cointegration toolkit with partly very interesting new findings and insights. In this study I will use a small German macro system consisting of the three variables log real M3 (m_t), log real GNP (gnp_t) and a long-term bond rate (R_t) to illustrate the power of cointegration analysis. The system was previously analyzed by Lütkepohl (2004b). It is modelled around a possible money demand relation. Thus, one would expect to find one long-run relation representing a money demand function. Hence, the cointegration framework is potentially useful for analyzing this system. The need for some of the recent developments will be demonstrated in the analysis.

In performing a cointegration analysis, the first step is to determine the order of integration of the individual variables. This part of the analysis will be discussed in Section 3. Then the number of cointegration relations has to be investigated. This issue is dealt with in Section 4. When the number of cointegration relations is known, their parameters can be estimated and restrictions may be placed on them as appropriate. This step is considered in Section 5. Although the cointegration relations often form the central part of interest, a complete model is necessary for assessing the general quality of the modelling exercise and for subsequent further investigations or forecasting. Therefore specifying and estimating the short-run part of the model for the DGP is discussed in Section 6. Model checking is treated in Section 7. Once a satisfactory complete model is available the dynamic interactions between the variables can be studied in more detail with the help of an impulse response analysis or a forecast error variance decomposition. These tools are presented in Section 8 and Section 9 concludes with a brief summary of some other interesting lines of research related to cointegration.

Throughout the issues are illustrated and the methods are guided by the small German monetary system sketched in the foregoing. The special data features call for special methods which have only recently been developed. Therefore the example system is useful for motivating the specific recent developments presented in this review. The data are discussed in more detail in the next section. The computations are performed with the software JMulTi (see Lütkepohl & Krätzig (2004) and the web page www.jmulti.de).

The following general notation will be used. The differencing and lag operators are denoted by Δ and L , respectively, that is, for a stochastic process y_t , $\Delta y_t = y_t - y_{t-1}$ and $Ly_t = y_{t-1}$. Convergence in distribution is signified by \xrightarrow{d} and \log denotes the natural

logarithm. The trace, determinant and rank of the matrix A are denoted by $\text{tr}(A)$, $\det(A)$ and $\text{rk}(A)$, respectively. The symbol vec is used for the column vectorization operator so that $\text{vec}(A)$ is the vector of columns of the matrix A . An $(n \times n)$ identity matrix is denoted by I_n . DGP, ML, LS, GLS, RR and LR are used to abbreviate data generation process, maximum likelihood, least squares, generalized least squares, reduced rank and likelihood ratio, respectively. VAR and VECM stand for vector autoregression and vector error correction model, respectively.

2 The Data

As mentioned in the introduction, an example model built around a money demand relation for Germany will be used for illustrative purposes throughout. The money demand relation is especially important for a monetary policy which targets the money stock growth. Such a monetary policy was conducted by the Bundesbank (German central bank) in Germany since the middle of the 1970s. Therefore investigating whether a stable money demand relation has existed for Germany for the period of monetary targeting by the Bundesbank is of interest.

According to economic theory real money demand should depend on the real transactions volume and a nominal interest rate. The latter variable represents opportunity costs of holding money. Because the quantity theory suggests a log-linear relationship, the three-dimensional system (m_t, gnp_t, R_t) is considered, where m_t is the log of real M3, gnp_t is the log of real GNP and R_t is the nominal long-term interest rate. The money stock measure M3 is used because the Bundesbank announced a target growth rate for that variable since 1988. In addition to currency holdings and sight deposits, M3 also includes savings deposits and time deposits for up to four years. Therefore it seems plausible to use a long-term interest rate as a measure for opportunity costs. Specifically, the so-called ‘Umlaufrendite’, an average bond rate will be used in the following. GNP represents the transactions volume. Clearly, in a system of this type there may be other important related variables. For instance, inflation or an exchange rate may be considered in addition to our three variables. A small three-dimensional system is preferable for illustrative purposes, however. An analysis of a related larger system was performed by Lütkepohl & Wolters (2003).

Figure 1 about here.

Our sample period is 1975Q1 – 1998Q4 because the Bundesbank started its monetary targeting policy in 1975 and at the beginning of 1999 the Euro was introduced so that the European Central Bank became responsible for the monetary policy. Quarterly, seasonally

unadjusted data is used. Both M3 and GNP are deflated by the GNP deflator.² The series m_t , gnp_t and R_t are plotted in Figure 1. The variables m_t and gnp_t have a seasonal pattern and a level shift in the third quarter of 1990 when the monetary unification of East and West Germany occurred. Before that date, the series only refer to West Germany and afterwards they refer to the unified Germany. The special data features and in particular the level shifts will be taken into account in the subsequent analysis.

3 Unit Root Analysis

We start by investigating the unit roots in the DGPs of the three individual series. In other words, their order of integration is determined.

3.1 The Augmented Dickey-Fuller Test

The point of departure is an $AR(p)$ with deterministic term ν_t , $y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \nu_t + u_t$. This process has a unit root and is hence integrated if $\alpha(1) = 1 - \alpha_1 - \dots - \alpha_p = 0$. The objective is therefore to test this null hypothesis against the alternative of stationarity of the process (i.e., $\alpha(1) > 0$). For this purpose the model is reparameterized by subtracting y_{t-1} on both sides and rearranging terms,

$$\Delta y_t = \pi y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \nu_t + u_t, \quad (1)$$

where $\pi = -\alpha(1)$ and $\gamma_j = -(\alpha_{j+1} + \dots + \alpha_p)$. The so-called *augmented Dickey-Fuller* (ADF) *test* for the pair of hypotheses $H_0 : \pi = 0$ versus $H_1 : \pi < 0$ is based on the t -statistic of the coefficient π from an OLS estimation of (1) (Fuller (1976), Dickey & Fuller (1979)). Its limiting distribution is nonstandard and depends on the deterministic terms in the model. Critical values have been simulated for different deterministic terms (see, e.g., Fuller (1976) and Davidson & MacKinnon (1993)). In these tests the number of lagged differences of y_t may be based on model selection criteria such as AIC, HQ or SC (see Lütkepohl (1991) for definitions) or a sequential testing procedure which eliminates insignificant coefficients (see, e.g., Ng & Perron (1995)).

If the time series under study may have more than one unit root, the series should first be differenced sufficiently often to make it stationary. Then an ADF test may be applied to the differenced series. If a unit root is rejected, an ADF test is applied to the series which is differenced one time less than in the previous test. This procedure is repeated until a unit root cannot be rejected. Suppose for instance that a series y_t is possibly $I(2)$. Then a unit

²Details of the data sources are given in the Appendix of Lütkepohl & Wolters (2003).

root test is applied to $\Delta^2 y_t$ first. If it rejects, a unit root test is applied to Δy_t . If the unit root cannot be rejected in Δy_t this result confirms that y_t is indeed best modelled as an I(2) series. If, however, a unit root is also rejected for Δy_t , the original series y_t is better not treated as I(2). This approach for determining the number of unit roots was proposed by Pantula (1989). It is therefore sometimes called the *Pantula principle*.

Table 1 about here.

For the German bond rate (R_t) ADF test results are given in Table 1. From Figure 1 one may conclude that the variable may be I(1). Therefore, the first differences are tested first. For both lag orders given in Table 1 the test clearly rejects the unit root. A deterministic term is not included in the test regression because a linear trend term is not regarded as plausible for the original series and a constant term vanishes upon differencing. The tests for a unit root in the original series do not reject the null hypothesis regardless of the lag order. Thus we conclude that the series should be treated as an I(1) variable in the subsequent analysis.

Both m_t and gnp_t have level shifts and therefore the deterministic term should be modified accordingly (see Perron (1989)). Suitable extensions of the ADF tests have been proposed recently and will be discussed next.

3.2 Unit Root Tests for Series with Structural Breaks

Perron (1989, 1990) considers models with deterministic terms $\mu_t = \mu_0 + \mu_0^s d_{tT_B} + \mu_1 t + \mu_1^s (t - T_B) d_{tT_B}$, where $d_{tT_B} = 0$ for $t \leq T_B$ and $d_{tT_B} = 1$ for $t > T_B$. Thus, if $\mu_0^s \neq 0$, there is a level shift after time T_B and a change in the trend slope occurs at the same time, if $\mu_1^s \neq 0$ (see also Amsler & Lee (1995) and Perron & Vogelsang (1992) for tests allowing for such deterministic terms). Saikkonen & Lütkepohl (2002) and Lanne, Lütkepohl & Saikkonen (2002) argue that a shift may not occur in a single period but may be spread out over a number of periods. Moreover, there may be a smooth transition to a new level. They consider shift functions of the general nonlinear form $f_t(\theta)' \gamma$ which are added to the deterministic term. Hence, if there is, e.g., a linear trend term and a shift, we have a model

$$y_t = \mu_0 + \mu_1 t + f_t(\theta)' \gamma + x_t, \quad (2)$$

where θ and γ are unknown parameters or parameter vectors and the errors x_t are assumed to be generated by an AR(p) process, $\alpha(L)x_t = u_t$ with $\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$.

Shift functions may, for example, be based on a simple shift dummy, d_{tT_B} or an exponential function such as $f_t(\theta) = 1 - \exp\{-\theta(t - T_B)\}$ for $t \geq T_B$ and zero elsewhere. The

simple shift dummy function does not involve any extra parameter θ and the parameter γ is a scalar. The exponential shift function allows for a nonlinear gradual shift to a new level, starting at time T_B . For this type of shift, both θ and γ are scalar parameters. The first one is confined to the positive real line ($\theta > 0$), whereas the second one may assume any value.

Saikkonen & Lütkepohl (2002) and Lanne et al. (2002) propose unit root tests for the model (2) which are based on estimating the deterministic term first by a generalized least squares procedure and subtracting it from the original series. Then an ADF type test is performed on the adjusted series $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t - f_t(\hat{\theta})' \hat{\gamma}$ based on a model which accounts for the estimation errors in the nuisance parameters and worked quite well in small sample simulations (Lanne et al. (2002)). As in the case of the ADF statistic, the asymptotic null distribution is nonstandard. Critical values are tabulated in Lanne et al. (2002). Again a different asymptotic distribution is obtained if the deterministic linear trend term is excluded a priori. Because the power of the test tends to improve when the linear trend is not present, it is advisable to use any prior information regarding the deterministic term. If the series of interest has seasonal fluctuations, it is also possible to include seasonal dummies in addition in the model (2). Another advantage of this approach is that it can be extended easily to the case where the break date is unknown (see Lanne, Lütkepohl & Saikkonen (2003)).

We have applied tests with a simple shift dummy and an exponential shift function to check the integration properties of the gnp_t and m_t series. It is known that the break has occurred in the third quarter of 1990 at the time of the German monetary unification. The break date T_B is therefore fixed accordingly. In Table 2 the test values for the two test statistics are given. They are all quite similar and do not provide evidence against a unit root in gnp_t and m_t .

Table 2 about here.

In Figure 2 the gnp series together with the estimated deterministic term and the adjusted series as well as the estimated exponential shift function are plotted. It turns out that in this case the exponential shift function looks almost like a shift dummy due to the large estimated value for θ . The sum of squared errors objective function which is minimized in estimating the deterministic parameters is also plotted as a function of θ in the lower right hand panel of Figure 2. Obviously, this function is decreasing in θ . Given that for large values of θ the exponential shift function is the same as a shift dummy for practical purposes, the shape of the shift function is not surprising. In the estimation procedure we have actually constrained the range of θ to the interval from zero to three because for $\theta = 3$ the exponential shift function almost represents an instantaneous shift to a new level. Therefore there is no need to consider larger values.

Figure 2 about here.

An analysis of the first differences of the two variables rejects unit roots in these series. Hence, there is some evidence that the variables are well modelled as I(1). The results for the first differences are not presented because the main purpose of this analysis is to illustrate the tests for series with level shifts. The first differences of the variables do not have a level shift anymore but just an outlying value for the third quarter of 1990 which is captured by using an impulse dummy variable in the tests.

4 Cointegration Rank Tests

The next step of the analysis is to investigate the number of cointegration relations between the series. A great number of proposals have been made for this purpose. Many of them are reviewed and compared in Hubrich, Lütkepohl & Saikkonen (2001). Generally, there is a good case for using the Johansen (1995a) likelihood ratio (LR) approach based on Gaussian assumptions and its modifications because all other approaches were found to have severe shortcomings in some situations. Even if the actual DGP is non Gaussian, the resulting pseudo LR tests for the cointegrating rank may have better properties than many competitors. Only if specific data properties make this approach problematic, using other tests may be worth trying. However, even in the LR approach special data properties such as level shifts should be taken into account. Suitable modifications exist and will be outlined in the following after the standard setup has been presented.

4.1 The Model Setup

It is assumed that the DGP of a given K -dimensional vector of time series y_t can be decomposed in a deterministic part, μ_t , and a stochastic part, x_t ,

$$y_t = \mu_t + x_t. \quad (3)$$

The deterministic part will only be of secondary interest. It may contain, e.g., a constant, a polynomial trend, seasonals and other dummy variables. The stochastic part x_t is an I(1) process generated by a VECM of the form

$$\Delta x_t = \alpha\beta'x_{t-1} + \Gamma_1\Delta x_{t-1} + \cdots + \Gamma_{p-1}\Delta x_{t-p+1} + u_t, \quad (4)$$

where u_t is a K -dimensional unobservable zero mean white noise process with positive definite covariance matrix $E(u_t u_t') = \Sigma_u$. The parameter matrices α and β have dimensions $(K \times r)$ and rank r . They specify the long-run part of the model with β containing the *cointegrating*

relations and α representing the *loading coefficients*. The Γ_i ($i = 1, \dots, p-1$) are $(K \times K)$ short-run parameter matrices.

If there are deterministic terms in the DGP of the variables of interest, the x_t 's are unobserved whereas the y_t 's will be the observable variables. Left-multiplying y_t in (3) by the operator $\Delta I_K - \alpha\beta'L - \Gamma_1\Delta L - \dots - \Gamma_{p-1}\Delta L^{p-1}$, it is easy to see that y_t has the VECM representation

$$\begin{aligned}\Delta y_t &= \alpha(\beta'y_{t-1} + \delta^{co'}d_{t-1}^{co}) + \Gamma_1\Delta y_{t-1} + \dots + \Gamma_{p-1}\Delta y_{t-p+1} + Cd_t^s + u_t \\ &= \alpha\beta^{*'}y_{t-1}^* + \Gamma_1\Delta y_{t-1} + \dots + \Gamma_{p-1}\Delta y_{t-p+1} + Cd_t^s + u_t,\end{aligned}\tag{5}$$

where d_t^{co} is a vector of deterministic variables which can be absorbed into the cointegration relations. The corresponding coefficient matrix is denoted by δ^{co} . The vector d_t^s includes the remaining deterministic variables with coefficient matrix C . The matrix $\beta^{*'}$ $= [\beta' : \delta^{co}']$ is $(r \times K^*)$ and $y_{t-1}^* = [y_{t-1}', d_{t-1}^{co'}]'$ is $(K^* \times 1)$ with $K^* = K + \text{dimension}(d_t^{co})$. This is the form of the process on which much of the inference is based.

In practice, it is necessary to determine the lag order p and the cointegrating rank r . the former quantity may be chosen by model selection criteria or sequential testing procedures. If r is still unknown at that stage, the least restricted form of the model should be used. In other words, lag order selection may be based on (5) with cointegration rank K or, equivalently, on the corresponding levels VAR representation. In the following section it is assumed that the lag order p has been chosen in a previous step of the analysis and the determination of the cointegrating rank is discussed for a given lag order.

4.2 The Tests

Denoting the matrix $\alpha\beta^{*'}$ in the error correction term by Π , the following sequence of hypotheses is considered in the Johansen approach:

$$H_0(i) : \text{rk}(\Pi) = i \text{ versus } H_1(i) : \text{rk}(\Pi) > i, \quad i = 0, \dots, K-1.\tag{6}$$

The cointegrating rank specified in the first null hypothesis which cannot be rejected is then chosen as cointegrating rank r . If $H_0(0)$, the first null hypothesis in this sequence, cannot be rejected, a VAR process in first differences is considered. If all the null hypotheses can be rejected including $H_0(K-1)$, $\text{rk}(\Pi) = K$ and the process is $I(0)$. Given that the variables are supposed to be $I(1)$, the process cannot really be $I(0)$. In other words, under the present assumptions, there is strictly speaking no need to consider a test of $H_0(K-1)$. However, it may serve as a check of the unit root analysis.

Using (pseudo) LR tests is attractive because, for any given lag order p , they are easy to compute if the short-run parameters, $\Gamma_1, \dots, \Gamma_{p-1}$, are unrestricted, as shown by Johansen

(1995a) (see also Sec. 5 of the present paper). The LR statistics under their respective null hypotheses have nonstandard asymptotic distributions, however. They depend on the difference $K - r_0$ and on the deterministic terms included in the DGP but not on the short-run dynamics.

Although the asymptotic theory for quite general situations is available, a possible problem arises in practice because the small sample properties of the tests can be improved by specifying the deterministic terms as tightly as possible (see also Saikkonen & Lütkepohl (1999) for an asymptotic analysis of this problem). For example, if there is no deterministic linear trend term, it is desirable to perform the cointegration rank tests without such terms. On the other hand, leaving them out if they are part of the DGP can lead to major distortions in the tests. Johansen (1995a) also provides the asymptotic theory for testing hypotheses regarding the deterministic terms which can be helpful in this respect. Interestingly, under standard assumptions these tests have asymptotic χ^2 distributions with the number of degrees of freedom corresponding to the number of restrictions imposed under the null hypothesis.

A case which is not easily handled in this framework is a deterministic term with shift dummy variables of the form $d_{tT_B} = 0$ for $t \leq T_B$ and $d_{tT_B} = 1$ for $t > T_B$, as specified before. Shift dummy variables may be necessary to capture a level shift in the variables in time period T_B , as in the example series m_t and gnp_t . If such dummy variables belong to the deterministic term, the asymptotic null distribution of the LR test statistic for the cointegrating rank also depends on the shift period T_B . This is problematic if T_B is unknown.

In that case, a variant of an LR test for the cointegrating rank suggested in a series of papers by Saikkonen and Lütkepohl is convenient (e.g., Saikkonen & Lütkepohl (2000c)). The idea is to estimate the parameters of the model under the null hypothesis in a first step using a model such as (5) by Johansen's ML procedure with the shift dummy included in d_t^{co} (see Sec. 5). Then the estimated α , β and Γ_i ($i = 1, \dots, p - 1$) are used to construct an estimator of the covariance matrix of x_t and a feasible GLS estimation is performed to estimate the parameters of the deterministic part from a model such as (3) with shift dummy in the deterministic term. For example, if the deterministic term has the form $\mu_t = \mu_0 + \mu_1 t + \delta d_{tT_B}$, the parameter vectors μ_0 , μ_1 and δ are estimated by feasible GLS applied to $y_t = \mu_0 + \mu_1 t + \delta d_{tT_B} + x_t$. Using these estimates y_t may be adjusted for deterministic terms to obtain $\tilde{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t - \hat{\delta} d_{tT_B}$ and the Johansen LR test for the cointegrating rank is applied to \tilde{x}_t .

The advantage of this procedure is that the asymptotic null distribution of the resulting test statistic does not depend on the shift dummy or the shift date. Therefore the procedure can be used even if the shift date is unknown. In that case, the shift date can be estimated

first and the whole procedure may be based on an estimated T_B (see Lütkepohl, Saikkonen & Trenkler (2004)).

Although the short-run dynamics do not matter for the asymptotic theory, they have a substantial impact in small and moderate samples. Therefore the choice of the lag order p is quite important. Choosing p rather large to be on the safe side as far as missing out on important short-run dynamics is concerned, may lead to a drastic loss in power of the cointegrating rank tests. On the other hand, choosing the lag order too small may lead to dramatic size distortions even for well-behaved DGPs. In a small sample simulation study, Lütkepohl & Saikkonen (1999) found that using the AIC criterion for order selection may be a good compromise. It is also a good idea to use a few different lag orders and check the robustness of the results.

Because the dimension of the system also has an important impact on the test results (Gonzalo & Pitarakis (1999)), it is useful to apply cointegration tests to all possible subsystems as well and check whether the results are consistent with those for a higher-dimensional model. For example, in a system of three I(1) variables, if all pairs of variables are found to be cointegrated, the cointegrating rank of the three-dimensional system must be 2.

There are many interesting suggestions for modifying and improving the Johansen approach to cointegration testing. For example, to improve the performance of the Johansen cointegration tests in small samples, Johansen (2002) presents a Bartlett correction. Also there are a number of proposals based on different ideas. As mentioned previously, much of the earlier literature is reviewed in Hubrich et al. (2001). Generally, at present it appears that the Johansen approach should be the default and only if there are particular reasons other proposals are worth contemplating.

4.3 Cointegration Tests for the Example System

As suggested in the previous section, the rank of all pairs of series is investigated in addition to the rank of the three-dimensional system. Knowing the cointegrating ranks of the subsystems can also be helpful in finding a proper normalization of the cointegration matrix for the estimation stage (see Section 5). Because of the shift in the m_t and gnp_t series in the third quarter of 1990 a shift dummy variable will be allowed for and the cointegration tests proposed by Saikkonen and Lütkepohl (S&L tests) are used. Table 3 contains the results.

Table 3 about here.

The deterministic terms in addition to the shift dummies and the number of lagged differences in the model have to be specified before the tests can be carried out. Because m_t and gnp_t both have some seasonality and a trending behavior that may perhaps be captured

with a linear trend term, seasonal dummy variables and a linear trend term are included in the models used in the tests. To avoid a decision whether the trend is just in the variables and, hence, orthogonal to the cointegration relations or fully general, both types of tests are performed. Notice that if the trend is orthogonal to the cointegration relations, it is captured by an intercept term in the specification (5).

An easy way to choose the number of lagged differences to be included in the model is to apply model selection criteria. Using a maximum lag order of 10, the lag orders specified in Table 3 were suggested by AIC and HQ. The larger number of lagged differences is always the one chosen by AIC and the lower lag order is obtained with the HQ criterion. Considering different orders is useful in this context because choosing the order too small can lead to size distortions for the tests while selecting too large an order may result in power reductions.

In Table 3 the sample period is 1975Q1 – 1998Q4, including presample values needed in the estimation. There is strong evidence for a cointegration rank of zero for the (m_t, gnp_t) and (gnp_t, R_t) systems so that the two variables in each of these systems are not likely to be cointegrated. On the other hand, one cointegration relation is found for the (m_t, R_t) system under both alternative trend specifications. Thus, one would also expect to find at least one cointegration relation in the three-dimensional system of all variables. If no cointegration relation exists between m_t and gnp_t as well as between gnp_t and R_t as suggested by the bivariate analysis, then there cannot be a second cointegration relation between the three variables. If two linearly independent cointegration relations exist between the three variables, they can always be transformed so that they both involve just two of the variables, as we will see in Section 5. Consistent with the results for the bivariate models, there is some evidence of just one cointegration relation in the three-dimensional system.

5 Estimating the Cointegration Relations

5.1 Estimation Methods

5.1.1 Reduced Rank ML Estimation

For a given cointegrating rank and lag order, the VECM (5) can be estimated by RR regression as shown in Johansen (1991, 1995). Assuming that a sample of size T and p presample values are available, the estimator may be determined by denoting the residuals from regressing Δy_t and y_{t-1}^* on $\Delta Y'_{t-1} = [\Delta y'_{t-1}, \dots, \Delta y'_{t-p+1}, d_t^{s'}]$ by R_{0t} and R_{1t} , respectively, defining $S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R'_{jt}$, $i, j = 0, 1$, and solving the generalized eigenvalue problem

$$\det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}) = 0. \quad (7)$$

Let the ordered eigenvalues be $\lambda_1 \geq \dots \geq \lambda_{K^*}$ with corresponding matrix of eigenvectors $B = [b_1, \dots, b_{K^*}]$ satisfying $\lambda_i S_{11} b_i = S'_{01} S_{00}^{-1} S_{01} b_i$ and normalized such that $B' S_{11} B = I_{K^*}$. Estimators of β^* and α are then given by

$$\hat{\beta}^* = [b_1, \dots, b_r] \quad \text{and} \quad \hat{\alpha} = S_{01} \hat{\beta}^* (\hat{\beta}^{*'} S_{11} \hat{\beta}^*)^{-1}. \quad (8)$$

The corresponding estimator of $\Gamma = [\Gamma_1, \dots, \Gamma_{p-1}, C]$ is

$$\hat{\Gamma} = \left(\sum_{t=1}^T (\Delta y_t - \hat{\alpha} \hat{\beta}^{*'} y_{t-1}^*) \Delta Y'_{t-1} \right) \left(\sum_{t=1}^T \Delta Y_{t-1} \Delta Y'_{t-1} \right)^{-1}$$

Under Gaussian assumptions these estimators are ML estimators conditional on the presample values (Johansen (1995a)). The estimator of Γ is consistent and asymptotically normal under general assumptions and, if there are no deterministic terms,

$$\sqrt{T} \text{vec}(\hat{\Gamma} - \Gamma) \xrightarrow{d} N(0, \Sigma_{\hat{\Gamma}}).$$

Here the asymptotic distribution of $\hat{\Gamma}$ has a nonsingular covariance matrix $\Sigma_{\hat{\Gamma}}$ so that standard inference may be used for the short-run parameters Γ_j . The convergence rate for the deterministic terms may be different if polynomial trends are included.

For the estimators $\hat{\alpha}$ and $\hat{\beta}^*$, the product $\hat{\Pi}^* = \hat{\alpha} \hat{\beta}^{*'}$ is also a consistent and asymptotically normally distributed estimator,

$$\sqrt{T} \text{vec}(\hat{\Pi}^* - \Pi^*) \xrightarrow{d} N(0, \Sigma_{\hat{\Pi}^*}).$$

The $(KK^* \times KK^*)$ covariance matrix $\Sigma_{\hat{\Pi}^*}$ is singular if $r < K$, however. The matrices α and β are only identified individually with further restrictions. Identifying and overidentifying restrictions for these matrices have been the subject of some recent research (see, e.g., Johansen (1995a), Johansen & Juselius (1992, 1994), Boswijk (1995, 1996), Elliott (2000), Pesaran & Shin (2002), Boswijk & Doornik (2002)). The latter article allows for very general nonlinear restrictions. For our purposes imposing restrictions on α , β and other parameters may be useful either for reducing the parameter space and thereby improving estimation precision or to identify the cointegration relations to associate them with economic relations. Imposing just-identifying restrictions on α and/or β does not do any damage. Therefore, we are free to impose just-identifying restrictions on the cointegration parameters.

A triangular form has received some attention in the literature (see, e.g., Phillips (1991)). It assumes that the first part of β is an identity matrix, $\beta' = [I_r : \beta'_{(K-r)}]$ and, hence, $\beta^{*'} = [I_r : \beta^{*'}_{(K-r)}]$, where $\beta_{(K-r)}$ is $((K-r) \times r)$ and $\beta^{*}_{(K-r)}$ is a $((K-r) \times r)$ matrix. For $r = 1$, this restriction amounts to normalizing the coefficient of the first variable to be one. Given that $\text{rk}(\beta) = r$, there exists a nonsingular $(r \times r)$ submatrix of β' which motivates the

normalization. Notice that $\Pi^* = \alpha\beta^* = \alpha\Phi\Phi^{-1}\beta^*$ for any nonsingular $(r \times r)$ matrix Φ . Hence, choosing Φ such that it corresponds to the nonsingular $(r \times r)$ submatrix of β' results in a decomposition of Π^* where β and, hence, β^* contains an identity submatrix. By a suitable rearrangement of the variables it can be ensured that β' is of the form $[I_r : \beta'_{(K-r)}]$. It should be clear, however, that such a normalization requires a suitable order of the variables. Some care is necessary in choosing this order to make sure that only valid cointegration relations result. In practice, it is usually fairly easy to choose the order of the variables properly if the cointegrating ranks of all subsystems are known as well. In other words, in the initial analysis it will be useful to not only check the cointegrating rank of the system of interest but also of all smaller dimensional subsystems, as was done for the example system. There are also formal statistical tests for normalizing restrictions (e.g., Luukkonen, Ripatti & Saikkonen (1999)).

The normalization ensures identified parameters $\beta^*_{(K^*-r)}$ so that inference becomes possible. To simplify matters, it is now assumed that there are no deterministic terms in the model. The estimators for the parameters $\beta_{(K-r)}$ have an asymptotic distribution which is multivariate normal upon appropriate normalization. Partitioning R_{1t} as $R'_{1t} = [R^{(1)'}_{1t}, R^{(2)'}_{1t}]$ where $R^{(1)}_{1t}$ and $R^{(2)}_{1t}$ are $(r \times 1)$ and $((K-r) \times 1)$, respectively, it holds that

$$\begin{aligned} & \text{vec} \left\{ (\hat{\beta}'_{(K-r)} - \beta'_{(K-r)}) \left(\sum_{t=1}^T R^{(2)}_{1t} R^{(2)'}_{1t} \right)^{1/2} \right\} \\ &= \left[\left(\sum_{t=1}^T R^{(2)}_{1t} R^{(2)'}_{1t} \right)^{1/2} \otimes I_{K-r} \right] \text{vec}(\hat{\beta}'_{(K-r)} - \beta'_{(K-r)}) \xrightarrow{d} N(0, I_{K-r} \otimes (\alpha' \Sigma_u^{-1} \alpha)^{-1}) \end{aligned} \quad (9)$$

(e.g., Reinsel (1993, Chapter 6)). The asymptotic distribution of the untransformed estimator is mixed normal (see Johansen (1995a)). The present result is useful for deriving t -tests or Wald tests for restrictions on the parameters $\beta_{(K-r)}$.

Using that $T^{-2} \sum_{t=1}^T R^{(2)}_{1t} R^{(2)'}_{1t}$ converges weakly, it can be seen from this result that $T \text{vec}(\hat{\beta}'_{(K-r)} - \beta'_{(K-r)})$ has an asymptotic distribution. In other words, the estimator $\hat{\beta}'_{(K-r)}$ converges at a rate T rather than \sqrt{T} .

Imposing identifying restrictions on β , expressions for the asymptotic covariances of the other parameters are also readily available:

$$\sqrt{T} \text{vec}([\hat{\alpha}, \hat{\Gamma}_1, \dots, \hat{\Gamma}_{p-1}] - [\alpha, \Gamma_1, \dots, \Gamma_{p-1}]) \xrightarrow{d} N(0, \Omega^{-1} \otimes \Sigma_u),$$

where

$$\Omega = \text{plim} T^{-1} \sum_{t=1}^T \left(\begin{bmatrix} \beta' y_{t-1} \\ \Delta Y_{t-1} \end{bmatrix} [y'_{t-1} \beta, \Delta Y'_{t-1}] \right).$$

Asymptotically these parameters are distributed independently of $\hat{\beta}'_{(K-r)}$.

Deterministic terms can be included by just extending the relevant quantities in the foregoing formulas. For example deterministic terms not included in the cointegration relations

are taken into account by adding the components to ΔY_{t-1} and extending the parameter matrix Γ accordingly. Deterministic terms which are restricted to the cointegration relations are accounted for by using y_{t-1}^* and β^* instead of y_{t-1} and β in the error correction term. The convergence rates of the deterministic terms depend on the specific components included.

5.1.2 A Two-step Estimator

Ahn & Reinsel (1990), Reinsel (1993, Chapter 6) and Saikkonen (1992) proposed another estimator for the cointegration parameters. To focus on the latter parameters, we consider the concentrated model corresponding to the VECM (5),

$$R_{0t} = \alpha\beta^{*'}R_{1t} + \tilde{u}_t. \quad (10)$$

Using the normalization $\beta^{*'} = [I_r : \beta_{(K^*-r)}^{*'}]$, this model can be written in the form

$$R_{0t} - \alpha R_{1t}^{(1)} = \alpha\beta_{(K^*-r)}^{*'}R_{1t}^{(2)} + \tilde{u}_t, \quad (11)$$

where $R_{1t}^{(1)}$ and $R_{1t}^{(2)}$ again consist of the first r and last $K^* - r$ components of R_{1t} , respectively. Premultiplying (11) by $(\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}$ and defining $w_t = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}(R_{0t} - \alpha R_{1t}^{(1)})$, gives

$$w_t = \beta_{(K^*-r)}^{*'}R_{1t}^{(2)} + v_t, \quad (12)$$

where $v_t = (\alpha'\Sigma_u^{-1}\alpha)^{-1}\alpha'\Sigma_u^{-1}\tilde{u}_t$ is an r -dimensional error vector. The corresponding error term of the unconcentrated model is a white noise process with mean zero and covariance matrix $\Sigma_v = (\alpha'\Sigma_u^{-1}\alpha)^{-1}$.

From this model $\beta_{(K^*-r)}^{*'}$ can be estimated by a two step procedure. In the first step, the parameters in the model $R_{0t} = \Pi^*R_{1t} + \tilde{u}_t$ are estimated by unrestricted OLS. The first r columns of Π^* are equal to α and hence these columns from the estimated matrix are used as an estimator $\tilde{\alpha}$. This estimator and the usual residual covariance estimator are used to obtain a feasible version of w_t , say $\tilde{w}_t = (\tilde{\alpha}'\tilde{\Sigma}_u^{-1}\tilde{\alpha})^{-1}\tilde{\alpha}'\tilde{\Sigma}_u^{-1}(R_{0t} - \tilde{\alpha}R_{1t}^{(1)})$. This quantity is substituted for w_t in (12) in the second step and $\beta_{(K^*-r)}^{*'}$ is estimated from that model by OLS. The resulting two step estimator has the same asymptotic distribution as the ML estimator (see Ahn & Reinsel (1990) and Reinsel (1993, Chapter 6)).

5.1.3 Other Estimators

So far we have started from a parametrically specified model setup. If interest centers on the cointegration parameters only, it is always possible to find a transformation of the variables such that the system of transformed variables can be written in so-called triangular form,

$$y_{1t} = \beta'_{(K-r)}y_{2t} + z_{1t} \quad \text{and} \quad \Delta y_{2t} = z_{2t},$$

where $z_t = [z'_{1t}, z'_{2t}]'$ is a general stationary linear process. Phillips (1991) considers inference for the cointegration parameters in this case and shows that the covariance structure of z_t has to be taken into account for optimal inference. Very general nonparametric estimators are sufficient, however, to obtain asymptotic optimality. Hence, it is not necessary to assume a specific parametric structure for the short-run dynamics.

There are also other systems methods for estimating the cointegrating parameters. For example, Stock & Watson (1988) consider an estimator based on principal components and Bossaerts (1988) uses canonical correlations. These estimators were shown to be inferior to Johansen's ML method in a comparison by Gonzalo (1994) and are therefore not further considered here.

5.1.4 Restrictions for the Cointegration Relations

In case just identifying restrictions for the cointegration relations are available, estimation may proceed by RR regression and then the identified estimator of β may be obtained by a suitable transformation of $\hat{\beta}$. For example, if β is just a single vector, a normalization of the first component may be obtained by dividing the vector $\hat{\beta}$ by its first component, as discussed previously.

Sometimes over-identifying restrictions are available for the cointegration matrix. They can be handled easily if they can be written in the form $\beta^* = H\varphi$, where H is some known, fixed ($K^* \times s$) matrix and φ is ($s \times r$) with $s \geq r$. In this case R_{1t} is simply replaced by $H'R_{1t}$ in the quantities entering the generalized eigenvalue problem (7), that is, we have to solve

$$\det(\lambda H'S_{11}H - H'S'_{01}S_{00}^{-1}S_{01}H) = 0 \quad (13)$$

for λ to get $\lambda_1^H \geq \dots \geq \lambda_s^H$. The eigenvectors corresponding to $\lambda_1^H, \dots, \lambda_r^H$ are the estimators of the columns of φ . Denoting the resulting estimator by $\hat{\varphi}$ gives a restricted estimator $\hat{\beta}^* = H\hat{\varphi}$ for β^* and corresponding estimators of α and Γ as previously.

More generally, restrictions may be available in the form $\beta^* = [H_1\varphi_1, \dots, H_r\varphi_r]$, where H_j is ($K \times s_j$) and φ_j is ($s_j \times 1$) ($j = 1, \dots, r$). In that case, restricted ML estimation is still not difficult but requires an iterative optimization whereas the two-step estimator is available in closed form, as will be shown now.

In general, if the restrictions can be represented in the form

$$\text{vec}(\beta_{(K^*-r)}^*) = \mathcal{H}\eta + h,$$

where \mathcal{H} is a fixed matrix, h a fixed vector and η a vector of free parameters, the second

step of the two-step estimator given in (13) may be adapted using the vectorized form

$$\begin{aligned} w_t &= (R_{1t}^{(2)'} \otimes I_r) \text{vec}(\beta_{(K^*-r)}^*) + v_t \\ &= (R_{1t}^{(2)'} \otimes I_r)(\mathcal{H}\eta + h) + v_t \end{aligned}$$

so that

$$\tilde{w}_t - (R_{1t}^{(2)'} \otimes I_r)h = (R_{1t}^{(2)'} \otimes I_r)\mathcal{H}\eta + v_t$$

can be used in the second step. The feasible GLS estimator of η , say $\tilde{\eta}$, has an asymptotic normal distribution upon appropriate normalization so that t -ratios can be obtained and interpreted in the usual manner.

5.2 Estimating the Example Cointegration Relation

Using the results of Section 4.3, we consider a VECM for the example series with cointegrating rank one, four lagged differences and and seasonal dummy variables. Moreover, the shift dummy is included in differenced form only because it turned out to be unnecessary in the cointegration relation. In other words, an impulse dummy variable $I90Q3_t = \Delta S90Q3_t$ is included instead of the shift dummy. A linear trend term was also included initially but was found to be insignificant. The money variable m_t is the first variable in our model because we want its coefficient to be normalized to one in the cointegration relation. The resulting ML estimator of the cointegration relation with standard errors in parentheses is

$$ec_t^{ML} = m_t - 1.093 gnp_t + 6.514 R_t$$

(0.090) (1.267)

or

$$m_t = 1.093 gnp_t - 6.514 R_t + ec_t^{ML}.$$

(0.090) (1.267)

This equation is easily interpreted as a money demand relation, where increases in the transactions volume increase money demand and increases in the opportunity costs (R_t) reduce the demand for money. The coefficient 1.093 of gnp_t is the estimated output elasticity because m_t and gnp_t appear in logarithms. For a constant velocity of money a 1% increase in the transactions volume is expected to induce a 1% increase in money demand. In other words, the output elasticity is expected to be one in a simple theoretical model. Therefore it is appealing that the gnp_t coefficient is close to 1. In fact, taking into account its standard deviation of 0.090, it is not significantly different from 1 at common significance levels. Using the two-step estimator for estimating the cointegration relation with a unit income elasticity gives

$$m_t = gnp_t - 3.876 R_t + ec_t^{2S}. \tag{14}$$

(0.742)

Notice that in this relation the coefficient of R_t is a semi elasticity because the interest rate is not in logarithms.

Taking into account the results of the cointegrating rank tests in Section 4.3, it may be puzzling that we found a cointegration relation between m_t and R_t that does not involve gnp_t in testing the bivariate system. This result suggests that the single cointegration relation found in the three-dimensional analysis may be one between m_t and R_t only which does not fit together with our money demand function (14). Because gnp_t enters significantly in the cointegration relation there is indeed a slight inconsistency between the bivariate and the three-dimensional analysis. Maintaining all three variables in the cointegration relation may still be reasonable because eliminating gnp_t from the cointegration relation imposes a restriction on the model which is rejected by the full three-dimensional information set.

6 Estimation of Short-run Parameters and Model Reduction

A VECM may also be estimated with restrictions on the loading coefficients (α), the short-run (Γ) and other parameter matrices. Restrictions for α are typically zero constraints, meaning that some cointegrating relations are excluded from some of the equations of the system. Usually it is possible to estimate β^* in a first stage. For example, ignoring the restrictions for the short-run parameters, the RR regression ML procedure or the two-step procedure may be used.

The first stage estimator $\hat{\beta}^*$, say, may be treated as fixed in a second stage estimation of the restricted VECM, because the estimators of the cointegrating parameters converge at a better rate than the estimators of the short-run parameters. In other words, a systems estimation procedure may be applied to

$$\Delta y_t = \alpha \hat{\beta}^{*'} y_{t-1}^* + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + C d_t^s + \hat{u}_t. \quad (15)$$

If only exclusion restrictions are imposed on the parameter matrices in this form, standard econometric systems estimation procedures such as feasible GLS or SURE (e.g., Judge, Griffiths, Hill, Lütkepohl & Lee (1985)) or similar methods may be applied which result in estimators of the short-run parameters with the usual asymptotic properties. A substantial number of articles deals with estimating models containing integrated variables. Examples are Phillips (1987, 1991), Phillips & Durlauf (1986), Phillips & Hansen (1990) and Phillips & Loretan (1991). A textbook treatment is given in Davidson (2000).

Some care is necessary with respect to the treatment of deterministic variables. For the parameters of those terms which are properly restricted to the cointegration relations the

properties can be recovered from a result similar to that given in (9). Thus, for example, t -ratios can be interpreted in the usual way. The properties of the estimators corresponding to d_t^s are not treated in detail here because in a subsequent analysis of the model, the parameters of the deterministic terms are often of minor interest (see, however, Sims, Stock & Watson (1990)).

The standard t -ratios and F -tests retain their usual asymptotic properties if they are applied to the short-run parameters in a VECM. Hence, individual zero coefficients can be imposed based on the t -ratios of the parameter estimators and one may sequentially eliminate those regressors with the smallest absolute values of t -ratios until all t -ratios (in absolute value) are greater than some threshold value γ . Alternatively, restrictions for individual parameters or groups of parameters in VECMs may be based on model selection criteria. Brüggemann & Lütkepohl (2001) discuss the relation between sequential testing procedures and using model selection criteria in this context.

Using the cointegration relation in (14) I have performed a model reduction starting from a model with four lagged differences of the variables. The model reduction procedure was based on a sequential selection of variables and the AIC. The following estimated model was obtained:

$$\begin{aligned}
\begin{bmatrix} \Delta m_t \\ \Delta gnp_t \\ \Delta R_t \end{bmatrix} &= \begin{bmatrix} -0.04 \\ (-3.1) \\ 0 \\ -0.01 \\ (-1.6) \end{bmatrix} (m_{t-1} - gnp_{t-1} + 3.876R_{t-1}) \\
+ \begin{bmatrix} 0.15 & -0.18 & -0.58 \\ (2.5) & (-2.9) & (-3.4) \\ 0.22 & -0.36 & 0 \\ (2.8) & (-4.2) & \\ 0 & 0 & 0.18 \\ & & (1.8) \end{bmatrix} \begin{bmatrix} \Delta m_{t-1} \\ \Delta gnp_{t-1} \\ \Delta R_{t-1} \end{bmatrix} &+ \begin{bmatrix} 0 & 0 & -0.30 \\ & & (-1.6) \\ 0.25 & -0.22 & 0.37 \\ (3.1) & (-2.4) & (1.5) \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta m_{t-2} \\ \Delta gnp_{t-2} \\ \Delta R_{t-2} \end{bmatrix} \\
+ \begin{bmatrix} 0 & -0.09 & 0 \\ & (-1.8) & \\ 0 & 0 & 0 \\ 0 & 0 & 0.18 \\ & & (1.8) \end{bmatrix} \begin{bmatrix} \Delta m_{t-3} \\ \Delta gnp_{t-3} \\ \Delta R_{t-3} \end{bmatrix} &+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.28 & 0 \\ & (4.0) & \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta m_{t-4} \\ \Delta gnp_{t-4} \\ \Delta R_{t-4} \end{bmatrix}
\end{aligned}$$

$$+ \begin{bmatrix} 0.15 & 0.07 & -0.03 & -0.02 & -0.02 \\ (17.5) & (4.9) & (-5.4) & (-3.5) & (-4.4) \\ 0.11 & 0.04 & -0.07 & -0.03 & -0.03 \\ (8.9) & (7.7) & (-9.1) & (-4.2) & (-3.5) \\ 0 & 0.01 & 0 & 0 & 0 \\ & (1.5) & & & \end{bmatrix} \begin{bmatrix} I90Q3_t \\ c \\ s_{1,t} \\ s_{2,t} \\ s_{3,t} \end{bmatrix} + \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \\ \hat{u}_{3,t} \end{bmatrix}. \quad (16)$$

Here estimation of the final model was done by feasible GLS applied to the full system and the t -values are given in parentheses. For the residuals the following covariance and correlation matrices were estimated:

$$\tilde{\Sigma}_u = \begin{bmatrix} 6.85 & -0.01 & 0.40 \\ \cdot & 13.3 & 1.12 \\ \cdot & \cdot & 2.59 \end{bmatrix} \times 10^{-5} \quad \text{and} \quad \widetilde{\text{Corr}} = \begin{bmatrix} 1 & -0.00 & 0.10 \\ \cdot & 1 & 0.19 \\ \cdot & \cdot & 1 \end{bmatrix}.$$

The off-diagonal elements of $\widetilde{\text{Corr}}$ are all quite small, given the effective sample size of $T = 91$ observations. Clearly, they are all smaller than $2/\sqrt{T} = 0.21$. Hence, they may be classified as not significantly different from zero. This result is good to remember at a later stage when an impulse response analysis is performed (see Section 8).

7 Model Checking

7.1 Some Tools

Various checks of the adequacy of a given model are available for VECMs. One group of checks considers the estimated residuals and another one investigates the time invariance of the model parameters. Residual based tests for autocorrelation, nonnormality, conditional heteroskedasticity etc. are available for stationary VAR models (e.g., Lütkepohl (1991), Doornik & Hendry (1997)). Many of the tests have been extended to VECMs with cointegrated variables as well. The modifications relative to the stationary VAR case are usually straightforward. Therefore these tests will not be discussed here. The situation is somewhat different with respect to checks of parameter constancy. In addition to more classical tests, specific tools for this purpose have been developed which are especially suitable for VECMs. Some of them will be presented in the following.

7.1.1 Chow Tests for Structural Stability

Chow tests check the null hypothesis of time invariant parameters throughout the sample period against the possibility of a change in the parameter values in period T_B . The model under consideration is estimated from the full sample of T observations and from the first T_1

and the last T_2 observations, where $T_1 < T_B$ and $T_2 \leq T - T_B$. The test is constructed using the LR principle based on Gaussian assumptions. In other words, the likelihood maximum from the constant parameter model is compared to the one with different parameter values before and after period T_B , leaving out the observations between T_1 and $T - T_2 + 1$. Denoting the conditional log-density of the t -th observation vector by $l_t = \log f(y_t | y_{t-1}, \dots, y_1)$, the Chow test statistic can be written as

$$\lambda_{Chow} = 2 \left[\sup \left(\sum_{t=1}^{T_1} l_t \right) + \sup \left(\sum_{t=T-T_2+1}^T l_t \right) - \left(\sum_{t=1}^{T_1} l_t^* + \sum_{t=T-T_2+1}^T l_t^* \right) \right],$$

where l_t^* is the log-likelihood increment for observation t evaluated at the parameter values which maximize the likelihood over the full sample. If the model is time invariant, the statistic has an asymptotic χ^2 -distribution. The degrees of freedom are given by the number of restrictions imposed by assuming a constant coefficient model for the full sample period, that is, it is the difference between the sum of the number of free coefficients estimated in the first and last subperiods and the number of free coefficients in the full sample model (see Hansen (2003)). The parameter constancy hypothesis is rejected if the value of the test statistic is large.

From the point of view of asymptotic theory there is no need to leave out any observations between the two subsamples. So $T_1 = T_B - 1$ and $T_2 = T - T_B$ is a possible choice. In practice, if the parameter change has not occurred instantaneously at the beginning of period T_B , but is spread out over a few periods or its exact timing is unknown, leaving out some observations may improve the small sample power of the test.

Various generalizations of these tests are possible. For example, one could test for more than one break or one could check constancy of a subset of parameters keeping the remaining ones fixed. Moreover, there may be deterministic terms in the cointegration relations or the number of cointegration relations may change in different subperiods. These generalizations are also treated by Hansen (2003). A Chow forecast test version for multivariate time series models was considered by Doornik & Hendry (1997). It tests the null hypothesis that the forecasts from a model fitted to the first T_B observations are in line with the actually observed data. Doornik & Hendry (1997) also proposed small sample corrections of the tests which may be used in conjunction with critical values from F distributions.

Candelon & Lütkepohl (2001) pointed out that especially for multivariate time series models the asymptotic χ^2 distribution may be an extremely poor guide for small sample inference. Even adjustments based on F approximations can lead to drastically distorted test sizes. Therefore they proposed to use bootstrap versions of the Chow tests in order to improve their small sample properties.

Chow tests are sometimes performed repeatedly for a range of potential break points T_B .

If the test decision is based on the maximum of the test statistics, the test is effectively based on the test statistic $\sup_{T_B \in \mathbb{T}} \lambda_{Chow}$, where $\mathbb{T} \subset \{1, \dots, T\}$ is the set of periods for which the test statistic is determined. The asymptotic distribution of this test statistic is not χ^2 . Distributions of test statistics of this kind are discussed by Andrews (1993), Andrews & Ploberger (1994) and Hansen (1997).

7.1.2 Recursive Eigenvalues

For parameter constancy analysis, Hansen & Johansen (1999) proposed recursive statistics based on the eigenvalues that were encountered in the RR ML estimation procedure. Let $\lambda_i^{(\tau)}$ be the i -th largest eigenvalue based on sample moments from the first τ observations only. They present approximate 95% confidence intervals for the nonzero true eigenvalues corresponding to $\lambda_1^{(\tau)}, \dots, \lambda_r^{(\tau)}$ under the assumption of time invariance of the DGP. The plots of the intervals for consecutive sample sizes $\tau = T_{\min}, \dots, T$, can reveal structural breaks in the DGP.

Hansen & Johansen (1999) also proposed formal tests for parameter constancy. The following notation will be used in stating them:

$$\xi_i^{(\tau)} = \log \left(\frac{\lambda_i^{(\tau)}}{1 - \lambda_i^{(\tau)}} \right)$$

and

$$\mathcal{T}(\xi_i^{(\tau)}) = \frac{\tau}{T} |(\xi_i^{(\tau)} - \xi_i^{(T)}) / \hat{\sigma}_{ii}|,$$

where $\hat{\sigma}_{ii}$ is a suitable estimator of the standard deviation of $(\xi_i^{(\tau)} - \xi_i^{(T)})$. The statistic $\mathcal{T}(\xi_i^{(\tau)})$ compares the i -th eigenvalue obtained from the full sample to the one estimated from the first τ observations only and Hansen & Johansen (1999) have shown that the maximum over all τ ,

$$\sup_{T_{\min} \leq \tau \leq T} \mathcal{T}(\xi_i^{(\tau)}),$$

has a limiting distribution which depends on a Brownian bridge. Critical values were tabulated by Ploberger, Krämer & Kontrus (1989). If the difference between the eigenvalues based on the subsamples and the full sample gets too large so that $\mathcal{T}(\xi_i^{(\tau)})$ exceeds the relevant critical value, the parameter constancy is rejected.

An alternative test considers the sum of the r largest recursive eigenvalues. It is based on the statistic

$$\mathcal{T} \left(\sum_{i=1}^r \xi_i^{(\tau)} \right) = \frac{\tau}{T} \left| \left[\sum_{i=1}^r (\xi_i^{(\tau)} - \xi_i^{(T)}) \right] / \hat{\sigma}_{1-r} \right|.$$

Here $\hat{\sigma}_{1-r}$ is an estimator of the standard deviation of the quantity $\sum_{i=1}^r (\xi_i^{(\tau)} - \xi_i^{(T)})$. The limiting distribution of

$$\sup_{T_{\min} \leq \tau \leq T} \mathcal{T} \left(\sum_{i=1}^r \xi_i^{(\tau)} \right)$$

is also given by Hansen & Johansen (1999).

7.2 Checking the Example Model

Estimating a VECM as in (16) with cointegrating rank one but otherwise unrestrictedly by the RR ML procedure and checking the residuals with autocorrelation and nonnormality tests, it turned out that the model is a quite satisfactory representation of the DGP. Detailed results are not shown to save space. Also a stability analysis based on the recursive eigenvalues and the $\mathcal{T}(\xi_1^{(\tau)})$ statistic for 1986Q1 – 1998Q4 did not give rise to concern. The value of the test statistic did not exceed the critical value for a 5% level test.

The sample-split Chow tests in Figure 3 show a somewhat different picture, however. The p -values are computed by a bootstrap on the assumption that a test is made for a single break point only. The cointegration relation is fixed throughout the sample. Moreover, the test assumes a time invariant residual covariance matrix. Notice that the test statistic is only computed for the center part of the sample because sufficiently many degrees of freedom have to be available for estimation in the two subsamples. Clearly, quite small p -values are estimated for part of the sample. Thus, one may conclude that there is a stability problem for the model parameters. A closer investigation reveals, however, that there is a possible ARCH problem in the residuals of the interest rate equation. ARCH effects in the residuals of financial data series such as interest rates are fairly common in practice. They are not necessarily a signal of inadequate modelling of the conditional mean of the DGP. Because interest centers on the latter part in the present analysis, the possibly remaining ARCH in the residuals of the interest rate equation is ignored. ARCH in the residuals signals volatility clusters that can lead to significant values of Chow tests because these tests compare the residual variability in different subperiods to decide on parameter instability. Higher volatility is indeed found in the first part of the sample and may be responsible for the significant sample-split Chow tests.

Figure 3 about here.

The usual diagnostic tests for autocorrelation in the residuals of the restricted model (16) did not give rise to concern about the adequacy of the subset model either. Given the results of the stability test based on the recursive eigenvalues, the model is used as a basis for

further analysis in the following section. Possible tools for such an analysis are considered next.

8 Impulse Response Analysis

8.1 Background

For an $I(0)$ process y_t , the effects of shocks in the variables are easily seen in its Wold moving average (MA) representation,

$$y_t = u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \dots \quad (17)$$

The coefficient matrices of this representation may be obtained by recursive formulas from the coefficient matrices A_j of the levels VAR representation, $y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \nu_t + u_t$, where ν_t contains all deterministic terms (see Lütkepohl (1991, Chapter 2)). The elements of the Φ_s 's may be interpreted as the responses to impulses hitting the system. In particular, the ij th element of Φ_s represents the expected marginal response of $y_{i,t+s}$ to a unit change in y_{jt} holding constant all past values of the process. Because u_{it} is the forecast error in y_{it} given $\{y_{t-1}, y_{t-2}, \dots\}$, the elements of Φ_s represent the impulse responses of the components of y_t with respect to the u_t innovations. Because these quantities are just the 1-step ahead forecast errors the corresponding impulse responses are sometimes referred to as *forecast error impulse responses* (Lütkepohl (1991)). In the presently considered $I(0)$ case, $\Phi_s \rightarrow 0$ as $s \rightarrow \infty$. Consequently, the effect of an impulse vanishes over time and is hence transitory.

These impulse responses have been criticized on the grounds that the underlying shocks may not occur in isolation if the components of u_t are instantaneously correlated. Therefore, orthogonal innovations are preferred in an impulse response analysis. Using a Choleski decomposition of the covariance matrix $E(u_t u_t') = \Sigma_u$ is one way to obtain uncorrelated innovations. Let \mathbf{B} be a lower-triangular matrix with the property that $\Sigma_u = \mathbf{B}\mathbf{B}'$. Then orthogonalized shocks are given by $\varepsilon_t = \mathbf{B}^{-1}u_t$. Substituting in (17) and defining $\Psi_i = \Phi_i \mathbf{B}$ ($i = 0, 1, 2, \dots$) gives

$$y_t = \Psi_0 \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots \quad (18)$$

Notice that $\Psi_0 = \mathbf{B}$ is lower triangular so that the first shock may have an instantaneous effect on all the variables, whereas the second shock can only have an instantaneous effect on y_{2t} to y_{Kt} but not on y_{1t} . This way a recursive Wold causal chain is obtained. The ε shocks are sometimes called *orthogonalized impulse responses* because they are instantaneously uncorrelated (orthogonal).

A drawback of these shocks is that many matrices \mathbf{B} exist which satisfy $\mathbf{B}\mathbf{B}' = \Sigma_u$. The Choleski decomposition is to some extent arbitrary if there are no good reasons for a

particular recursive structure. Clearly, if a lower triangular Choleski decomposition is used to obtain \mathbf{B} , the actual innovations will depend on the ordering of the variables in the vector y_t so that different shocks and responses may result if the vector y_t is rearranged. In response to this problem, Sims (1981) recommended to consider different triangular orthogonalizations and check the robustness of the results if no particular ordering is suggested by economic theory. Taking into account subject matter theory in identifying the relevant impulses is the idea underlying structural VAR modelling. I do not discuss that issue here in detail here but refer the reader to Breitung, Brüggemann & Lütkepohl (2004) for a recent introduction.

For nonstationary cointegrated processes the Wold representation does not exist. Still the Φ_s impulse response matrices can be computed as for stationary processes from the levels version of a VECM (Lütkepohl (1991, Chapter 11), Lütkepohl & Reimers (1992)). Generally the Φ_s will not converge to zero as $s \rightarrow \infty$ in this case. Consequently, some shocks may have permanent effects. Distinguishing between shocks with permanent and transitory effects can also help in finding identifying restrictions for the innovations and impulse responses of a VECM. For an introduction to structural VECMs see also Breitung et al. (2004).

8.2 Statistical Inference for Impulse Responses

8.2.1 Asymptotic Theory Considerations

Suppose an estimator $\hat{\theta}$, say, of the model parameters θ is available. Then an impulse response coefficient $\phi = \phi(\theta)$, say, can be estimated as $\hat{\phi} = \phi(\hat{\theta})$. If $\hat{\theta}$ has an asymptotic normal distribution, $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_{\hat{\theta}})$, then $\hat{\phi}$ is also asymptotically normally distributed. Denoting by $\partial\phi/\partial\theta$ the vector of first order partial derivatives of ϕ with respect to the elements of θ and using the delta method gives

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma_{\hat{\phi}}^2), \quad (19)$$

where $\sigma_{\hat{\phi}}^2 = \frac{\partial\phi}{\partial\theta'} \Sigma_{\hat{\theta}} \frac{\partial\phi}{\partial\theta}$. This result holds if $\sigma_{\hat{\phi}}^2$ is nonzero which is guaranteed if $\Sigma_{\hat{\theta}}$ is nonsingular and $\partial\phi/\partial\theta \neq 0$. The covariance matrix $\Sigma_{\hat{\theta}}$ may be singular if the system contains I(1) variables. The partial derivatives will also usually be zero in some points of the parameter space because the ϕ generally consist of sums of products of the VAR coefficients. Then the partial derivatives will also be sums of products of such coefficients which may be zero. The partial derivatives are nonzero if all elements of θ are nonzero. Therefore, fitting subset models where all those coefficients are restricted to zero which are actually zero, helps to make the asymptotics for impulse responses work (see Benkwitz, Lütkepohl & Wolters (2001)).

8.2.2 Bootstrapping Impulse Responses

In practice, confidence intervals (CIs) for impulse responses are often constructed by bootstrap methods because they have some advantages over asymptotic CIs. In particular, they were found to be more reliable in small samples than those based on asymptotic theory (e.g., Kilian (1998)). Moreover, precise expressions for the asymptotic variances of the impulse response coefficients are not needed if a bootstrap is used. The asymptotic variances are rather complicated (e.g., Lütkepohl (1991, Chapter 3)) and it may therefore be an advantage if they can be avoided.

Typically, a residual based bootstrap is used in this context. Let ϕ , $\hat{\phi}$ and $\hat{\phi}^*$ denote some general impulse response coefficient, its estimator implied by the estimators of the model coefficients and the corresponding bootstrap estimator, respectively. The *standard percentile interval* is perhaps the most common method in setting up CIs for impulse responses in practice. It is given by $\left[s_{\gamma/2}^*, s_{(1-\gamma/2)}^* \right]$, where $s_{\gamma/2}^*$ and $s_{(1-\gamma/2)}^*$ are the $\gamma/2$ - and $(1 - \gamma/2)$ -quantiles, respectively, of the empirical distribution of the $\hat{\phi}^*$ (see, e.g., Efron & Tibshirani (1993)). Benkwitz et al. (2001) also consider *Hall's percentile interval* (Hall (1992)) which is derived using that the distribution of $\sqrt{T}(\hat{\phi} - \phi)$ is approximately equal to that of $\sqrt{T}(\hat{\phi}^* - \hat{\phi})$ in large samples. The resulting CI is $\left[\hat{\phi} - t_{(1-\gamma/2)}^*, \hat{\phi} - t_{\gamma/2}^* \right]$. Here $t_{\gamma/2}^*$ and $t_{(1-\gamma/2)}^*$ are the $\gamma/2$ - and $(1 - \gamma/2)$ -quantiles, respectively, of the empirical distribution of $(\hat{\phi}^* - \hat{\phi})$.

Unfortunately, the bootstrap generally does not overcome the problems due to a singularity in the asymptotic distribution which results from a zero variance in (19). In these cases bootstrap CIs may not have the desired coverage probability as discussed by Benkwitz, Lütkepohl & Neumann (2000). To overcome these problems one may (i) consider bootstrap procedures that adapt to the kind of singularity in the asymptotic distribution, (ii) fit subset models or (iii) assume an infinite VAR order. The first one of these approaches has drawbacks in empirical applications (see Benkwitz et al. (2000)). Either they are not very practical for processes of realistic dimension and autoregressive order or they do not perform well in samples of typical size.

Fitting subset models may also be problematic because this only solves the singularity problem if indeed all zero coefficients are found (Benkwitz et al. (2001)). Usually there is uncertainty regarding the actual zero restrictions if statistical methods are used for subset modelling, however. The third possible solution to the singularity problem is to assume a VAR or VECM with infinite lag order and letting the model order increase when more sample information becomes available. In this approach the model order is assumed to go to infinity with the sample size at a suitable rate. Relevant asymptotic theory was developed by Lütkepohl (1988, 1996), Lütkepohl & Poskitt (1991, 1996), Lütkepohl & Saikkonen (1997) and Saikkonen & Lütkepohl (1996, 2000) based on work by Lewis & Reinsel (1985) and

Saikkonen (1992). The disadvantage of this approach is that the greater generality of the models implies an inefficiency relative to the model with finite fixed order, provided the latter is a proper representation of the actual DGP. For practical purposes, subset modelling may be the best solution.

8.3 Impulse Response Analysis of the Example System

For illustrative purposes an impulse response analysis is performed based on the subset VECM (16). Thereby we hope to account for the problems related to constructing bootstrap confidence intervals. Because the estimated instantaneous residual correlations were found to be small, it may be reasonable to consider the forecast error impulse responses. They are shown in Figure 4 with standard percentile and Hall's percentile confidence intervals, based on 2000 bootstrap replications. According to the bootstrap literature the number of bootstrap replications has to be quite large in order to obtain reliable results. Therefore one may wonder if using 2000 replications is adequate. We have also computed CIs with 1000 replications and found that they are not very different from those based on 2000 replications. Hence, 2000 replications should be sufficient for the present example.

Figure 4 about here.

The two different methods for constructing CIs result in very similar intervals for the present example system (see Figure 4). The impulse responses are all quite plausible. For example, an interest rate impulse leads to a reduction in money demand and in output, whereas a money shock raises output and, in the long-run, tends to decrease the nominal interest rate. Not surprisingly, shocks in all three variables have long-term impacts because the variables are all I(1).

8.4 Forecast Error Variance Decomposition

Forecast error variance decompositions are related to impulse responses and may also be used for interpreting dynamic models. The h -step forecast error for the y_t variables in terms of structural innovations $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Kt})' = \mathbf{B}^{-1}u_t$ can be shown to be

$$\Psi_0\varepsilon_{T+h} + \Psi_1\varepsilon_{T+h-1} + \dots + \Psi_{h-1}\varepsilon_{T+1},$$

so that the k th element of the forecast error vector is

$$\sum_{n=0}^{h-1} (\psi_{k1,n}\varepsilon_{1,T+h-n} + \dots + \psi_{kK,n}\varepsilon_{K,T+h-n}),$$

where $\psi_{ij,n}$ denotes the ij th element of Ψ_n (see Lütkepohl (1991)). Because, by construction, the ε_{kt} are contemporaneously and serially uncorrelated and have unit variances, the corresponding forecast error variance is

$$\sigma_k^2(h) = \sum_{n=0}^{h-1} (\psi_{k1,n}^2 + \cdots + \psi_{kK,n}^2) = \sum_{j=1}^K (\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2).$$

The quantity $(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2)$ is interpreted as the contribution of variable j to the h -step forecast error variance of variable k . This interpretation is justified if the ε_{it} can be viewed as shocks in variable i . The percentage contribution of variable j to the h -step forecast error variance of variable k is obtained by dividing the above terms by $\sigma_k^2(h)$,

$$\omega_{kj}(h) = (\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2) / \sigma_k^2(h).$$

The corresponding estimated quantities are often reported for various forecast horizons.

Figure 5 about here.

In Figure 5, a forecast error variance decomposition of the German macro system based on the subset VECM (16) is shown. It uses orthogonalized innovations obtained via a Choleski decomposition of the covariance matrix. In Figure 5 it appears that the interest rate dominates its own development as well as that of m_t at least in the long-run, whereas the gnp_t variable is largely determined by its own innovations. This interpretation relies on the point estimates, however, because the forecast error variance components are computed from estimated quantities. They are therefore uncertain. Also, the ordering of the variables may have an impact on the results. Although the instantaneous residual correlation is small in our subset VECM, it may have some impact on the outcome of a forecast error variance decomposition. This possibility was checked by reversing the ordering of the variables. It turned out that for the present system the ordering of the variables has a very small effect only.

9 Conclusions and Extensions

Cointegration analysis has become a standard tool in econometrics during the last two decades after its introduction by Granger (1981) and Engle & Granger (1987). In this paper some recent developments are reviewed. A small German macro system around a money demand relation is used to illustrate the methodology. The example data have special features for which new methods have been developed recently. In particular, they have level shifts that have to be taken into account in unit root and cointegration testing and in modelling

the DGP. Some recent tools for handling such data properties have been discussed. Also methods for parameter estimation and model checking have been presented and applied. A satisfactory model for the example data set is found. This model is then used to study the dynamic interactions between the variables within an impulse response analysis and by means of a forecast error variance decomposition. Some recent developments in using these tools are also discussed.

In this review I have not tried to present all the interesting and exciting developments of cointegration analysis over the last two decades. The present review focusses explicitly on developments related to a specific example data set and on methodology to which I have contributed. An interesting development that has not been considered is, for instance, the analysis of systems with higher integration orders. Considerable progress has been made on the theory for analyzing models of this type (see, e.g., Johansen (1995b, (1997)), Kitamura (1995), Haldrup (1998), Paruolo (2000, 2002), Paruolo & Rahbek (1999), Boswijk (2000)). Moreover, models for variables with seasonal unit roots have been analyzed (Johansen & Schaumburg (1999), Ghysels & Osborn (2001, Chapter 3)). Generally these models are more complicated than the I(1) case and the theory is not as complete as that of I(1) models. Not surprisingly, there are also fewer applications.

Another generalization of the models considered so far is obtained by allowing the order of integration to be a real number rather than restricting it to an integer value. For real numbers d , one may define the differencing operator Δ^d by the following power series expansion:

$$\Delta^d = \sum_{j=0}^{\infty} L^j \Gamma(j-d)/\Gamma(j+1)\Gamma(-d),$$

where $\Gamma(\cdot)$ denotes the Gamma function. With this definition, processes y_t may be considered for which $\Delta^d y_t$ is stationary for real values of $d > -1$. The concept of cointegration has been extended to this type of *fractionally integrated* processes (e.g., Velasco (2003), Marmol, Escribano & Aparicio (2002), Robinson & Yajima (2002), Breitung & Hassler (2002)). They allow more flexibility with respect to the persistence of shocks to the system.

It is also possible to extend the linear VECMs by considering nonlinear error correction terms. For example, the term $\alpha\beta'y_{t-1}$ may be replaced by a nonlinear function $f(\beta'y_{t-1})$ or more generally by $g(y_{t-1})$. Such extensions may be of interest because the implications of linear models are not always fully satisfactory. For example, in a linear model a positive deviation from the long-run equilibrium relation has the same effect as a negative deviation of the same magnitude except that it has the opposite sign. Such a reaction is not always realistic in economic systems, where, for instance, the reaction may depend on the state of the business cycle. Models with nonlinear error correction terms have been proposed and considered, for example, by Balke & Fomby (1997), Escribano & Mira (2002), Saikkonen

(2001a, 2001b) and Lo & Zivot (2001).

Other forms of nonlinearities may also be considered. For example, nonlinearities may be present in the short-run dynamics in addition or alternatively to the error correction term. For example, Krolzig (1997) extends Markov regime switching models which were originally introduced to econometrics by Hamilton (1989), to systems of cointegrated variables.

Other extensions of the basic model include VECMs with finite order vector MA terms (Lütkepohl & Claessen (1997), Lütkepohl (2002)) and models which condition on some of the variables (Harbo, Johansen, Nielsen & Rahbek (1998), Pesaran, Shin & Smith (2000)).

To do a cointegration analysis it is usually not necessary anymore to develop the software because some packages exist which can be used comfortably. Examples are EViews (EViews (2000)), PcFiml (Doornik & Hendry (1997)), Microfit (Pesaran & Pesaran (1997)), CATS (Hansen & Juselius (1994)), JMulTi (Lütkepohl & Krätzig (2004)). The latter program was also used for the computations related to the example discussed in the present paper.

References

- Ahn, S. K. & Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* **85**: 813–823.
- Amsler, C. & Lee, J. (1995). An lm test for a unit root in the presence of a structural change, *Econometric Theory* **11**: 359–368.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–856.
- Andrews, D. W. K. & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**: 1383–1414.
- Balke, N. S. & Fomby, T. B. (1997). Threshold cointegration, *International Economic Review* **38**: 627–645.
- Benkwitz, A., Lütkepohl, H. & Neumann, M. (2000). Problems related to bootstrapping impulse responses of autoregressive processes, *Econometric Reviews* **19**: 69–103.
- Benkwitz, A., Lütkepohl, H. & Wolters, J. (2001). Comparison of bootstrap confidence intervals for impulse responses of German monetary systems, *Macroeconomic Dynamics* **5**: 81–100.
- Bossaerts, P. (1988). Common nonstationary components of asset prices, *Journal of Economic Dynamics and Control* **12**: 347–364.
- Boswijk, H. P. (1995). Efficient inference on cointegration parameters in structural error correction models, *Journal of Econometrics* **69**: 133–158.
- Boswijk, H. P. (1996). Testing identifiability of cointegrating vectors, *Journal of Business & Economic Statistics* **14**: 153–160.
- Boswijk, H. P. (2000). Mixed normality and ancillarity in I(2) systems, *Econometric Theory* **16**: 878–904.

- Boswijk, H. P. & Doornik, J. A. (2002). Identifying, estimating and testing restricted cointegrated systems: An overview, *Paper presented at the Henri Theil Memorial Conference*, Universiteit van Amsterdam.
- Breitung, J., Brüggemann, R. & Lütkepohl, H. (2004). Structural vector autoregressive modelling and impulse responses, in H. Lütkepohl & M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press.
- Breitung, J. & Hassler, U. (2002). Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics* **110**: 167–185.
- Brüggemann, R. & Lütkepohl, H. (2001). Lag selection in subset VAR models with an application to a U.S. monetary system, in R. Friedmann, L. Knüppel & H. Lütkepohl (eds), *Econometric Studies: A Festschrift in Honour of Joachim Frohn*, LIT Verlag, Münster, pp. 107–128.
- Candelon, B. & Lütkepohl, H. (2001). On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models, *Economics Letters* **73**: 155–160.
- Davidson, J. (2000). *Econometric Theory*, Blackwell, Oxford.
- Davidson, R. & MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, London.
- Dickey, D. A. & Fuller, W. A. (1979). Estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**: 427–431.
- Doornik, J. A. & Hendry, D. F. (1997). *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, International Thomson Business Press, London.
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- Elliott, G. (2000). Estimating restricted cointegrating vectors, *Journal of Business & Economic Statistics* **18**: 91–99.
- Engle, R. F. & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing, *Econometrica* **55**: 251–276.
- Escribano, A. & Mira, S. (2002). Nonlinear error correction models, *Journal of Time Series Analysis* **23**: 509–522.
- EViews (2000). *EViews 4.0 User's Guide*, Quantitative Micro Software, Irvine, CA.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*, John Wiley & Sons, New York.
- Ghysels, E. & Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press, Cambridge.
- Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics* **60**: 203–233.
- Gonzalo, J. & Pitarakis, J.-Y. (1999). Dimensionality effect in cointegration analysis, in R. Engle & H. White (eds), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 212–229.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**: 121–130.

- Haldrup, N. (1998). An econometric analysis of I(2) variables, *Journal of Economic Surveys* **12**: 595–650.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**: 357–384.
- Hansen, B. E. (1997). Approximate asymptotic p values for structural-change tests, *Journal of Business & Economic Statistics* **15**: 60–67.
- Hansen, H. & Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR-models, *Econometrics Journal* **2**: 306–333.
- Hansen, H. & Juselius, K. (1994). *CATS for RATS4: Manual to Cointegration Analysis to Time Series*, Estima.
- Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* **114**: 261–295.
- Harbo, I., Johansen, S., Nielsen, B. & Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems, *Journal of Business & Economic Statistics* **16**: 388–399.
- Hubrich, K., Lütkepohl, H. & Saikkonen, P. (2001). A review of systems cointegration tests, *Econometric Reviews* **20**: 247–318.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**: 1551–1581.
- Johansen, S. (1995a). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Johansen, S. (1995b). A statistical analysis of cointegration for I(2) variables, *Econometric Theory* **11**: 25–59.
- Johansen, S. (1997). Likelihood analysis of the I(2) model, *Scandinavian Journal of Statistics* **24**: 433–462.
- Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica* **70**: 1929–1961.
- Johansen, S. & Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK, *Journal of Econometrics* **53**: 211–244.
- Johansen, S. & Juselius, K. (1994). Identification of the long-run and the short-run structure. An application to the ISLM model, *Journal of Econometrics* **63**: 7–36.
- Johansen, S. & Schaumburg, E. (1999). Likelihood analysis of seasonal cointegration, *Journal of Econometrics* **88**: 301–339.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. & Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, 2nd edn, John Wiley and Sons, New York.
- Kilian, L. (1998). Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* **80**: 218–230.

- Kitamura, Y. (1995). Estimation of cointegrated systems with I(2) processes, *Econometric Theory* **11**: 1–24.
- Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin.
- Lanne, M., Lütkepohl, H. & Saikkonen, P. (2002). Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis* **23**: 667–685.
- Lanne, M., Lütkepohl, H. & Saikkonen, P. (2003). Test procedures for unit roots in time series with level shifts at unknown time, *Oxford Bulletin of Economics and Statistics* **65**: 91–115.
- Lewis, R. & Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**: 393–411.
- Lo, M. C. & Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price, *Macroeconomic Dynamics* **5**: 533–576.
- Lütkepohl, H. (1988). Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process, *Econometric Theory* **4**: 77–85.
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.
- Lütkepohl, H. (1996). Testing for nonzero impulse responses in vector autoregressive processes, *Journal of Statistical Planning and Inference* **50**: 1–20.
- Lütkepohl, H. (2002). Forecasting cointegrated VARMA processes, in M. P. Clements & D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 179–205.
- Lütkepohl, H. (2004a). Univariate time series analysis, in H. Lütkepohl & M. Kräätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.
- Lütkepohl, H. (2004b). Vector autoregressive and vector error correction models, in H. Lütkepohl & M. Kräätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.
- Lütkepohl, H. & Claessen, H. (1997). Analysis of cointegrated VARMA processes, *Journal of Econometrics* **80**: 223–239.
- Lütkepohl, H. & Kräätzig, M. (eds) (2004). *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.
- Lütkepohl, H. & Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models, *Econometric Theory* **7**: 487–496.
- Lütkepohl, H. & Poskitt, D. S. (1996). Testing for causation using infinite order vector autoregressive processes, *Econometric Theory* **12**: 61–87.
- Lütkepohl, H. & Reimers, H.-E. (1992). Impulse response analysis of cointegrated systems, *Journal of Economic Dynamics and Control* **16**: 53–78.
- Lütkepohl, H. & Saikkonen, P. (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes, *Journal of Econometrics* **81**: 127–157.
- Lütkepohl, H. & Saikkonen, P. (1999). Order selection in testing for the cointegrating rank of a VAR process, in R. F. Engle & H. White (eds), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 168–199.

- Lütkepohl, H. & Saikkonen, P. (2000). Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics* **95**: 177–198.
- Lütkepohl, H., Saikkonen, P. & Trenkler, C. (2004). Testing for the cointegrating rank of a VAR process with level shift at unknown time, *Econometrica* **72**: 647–662.
- Lütkepohl, H. & Wolters, J. (2003). The transmission of German monetary policy in the pre-euro period, *Macroeconomic Dynamics* **7**: 711–733.
- Luukkonen, R., Ripatti, A. & Saikkonen, P. (1999). Testing for a valid normalization of cointegrating vectors in vector autoregressive processes, *Journal of Business & Economic Statistics* **17**: 195–204.
- Marmol, F., Escribano, A. & Aparicio, F. M. (2002). Instrumental variable interpretation of cointegration with inference results for fractional cointegration, *Econometric Theory* **18**: 646–672.
- Ng, S. & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* **90**: 268–281.
- Pantula, S. G. (1989). Testing for unit roots in time series data, *Econometric Theory* **5**: 256–271.
- Paruolo, P. (2000). Asymptotic efficiency of the two stage estimator in I(2) systems, *Econometric Theory* **16**: 524–550.
- Paruolo, P. (2002). Asymptotic inference on the moving average impact matrix in cointegrated I(2) VAR systems, *Econometric Theory* **18**: 673–690.
- Paruolo, P. & Rahbek, A. (1999). Weak exogeneity in I(2) VAR systems, *Journal of Econometrics* **93**: 281–308.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* **57**: 1361–1401.
- Perron, P. (1990). Testing for a unit root in a time series with a changing mean, *Journal of Business & Economic Statistics* **8**: 153–162.
- Perron, P. & Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business & Economic Statistics* **10**: 301–320.
- Pesaran, M. H. & Pesaran, B. (1997). *Working with Microfit 4.0: Interactive Econometric Analysis*, Oxford University Press, Oxford.
- Pesaran, M. H. & Shin, Y. (2002). Long-run structural modelling, *Econometric Reviews* **21**: 49–87.
- Pesaran, M. H., Shin, Y. & Smith, R. J. (2000). Structural analysis of vector error correction models with I(1) variables, *Journal of Econometrics* **97**: 293–343.
- Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**: 277–301.
- Phillips, P. C. B. (1991). Optimal inference in cointegrated systems, *Econometrica* **59**: 283–306.
- Phillips, P. C. B. & Durlauf, S. N. (1986). Multiple time series regression with integrated processes, *Review of Economic Studies* **53**: 473–495.
- Phillips, P. C. B. & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* **57**: 99–125.

- Phillips, P. C. B. & Loretan, M. (1991). Estimating long-run economic equilibria, *Review of Economic Studies* **58**: 407–436.
- Ploberger, W., Krämer, W. & Kontrus, K. (1989). A new test for structural stability in the linear regression model, *Journal of Econometrics* **40**: 307–318.
- Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.
- Robinson, P. M. & Yajima, Y. (2002). Determination of cointegrating rank in fractional systems, *Journal of Econometrics* **106**: 217–241.
- Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.
- Saikkonen, P. (2001a). Consistent estimation in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations, *Econometric Theory* **17**: 296–326.
- Saikkonen, P. (2001b). Statistical inference in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations, *Econometric Theory* **17**: 327–356.
- Saikkonen, P. & Lütkepohl, H. (1996). Infinite order cointegrated vector autoregressive processes: Estimation and inference, *Econometric Theory* **12**: 814–844.
- Saikkonen, P. & Lütkepohl, H. (1999). Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory* **15**: 50–78.
- Saikkonen, P. & Lütkepohl, H. (2000a). Asymptotic inference on nonlinear functions of the coefficients of infinite order cointegrated VAR processes, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim & A. Würtz (eds), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge University Press, Cambridge, pp. 165–201.
- Saikkonen, P. & Lütkepohl, H. (2000b). Testing for the cointegrating rank of a VAR process with an intercept, *Econometric Theory* **16**: 373–406.
- Saikkonen, P. & Lütkepohl, H. (2000c). Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics* **18**: 451–464.
- Saikkonen, P. & Lütkepohl, H. (2002). Testing for a unit root in a time series with a level shift at unknown time, *Econometric Theory* **18**: 313–348.
- Sims, C. A. (1981). An autoregressive index model for the U.S. 1948-1975, in J. Kmenta & J. B. Ramsey (eds), *Large-Scale Macro-Econometric Models*, North-Holland, Amsterdam, pp. 283–327.
- Sims, C. A., Stock, J. H. & Watson, M. W. (1990). Inference in linear time series models with some unit roots, *Econometrica* **58**: 113–144.
- Stock, J. H. & Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association* **83**: 1097–1107.
- Velasco, C. (2003). Gaussian semi-parametric estimation of fractional cointegration, *Journal of Time Series Analysis* **24**: 345–378.

| variable | deterministic term | no. of lagged differences | test statistic | 5% critical value |
|--------------|--------------------|---------------------------|----------------|-------------------|
| ΔR_t | none | 0 | -8.75 | -1.94 |
| | | 2 | -4.75 | |
| R_t | constant | 1 | -1.48 | -2.86 |
| | | 3 | -1.93 | |

Note: Critical values from Davidson & MacKinnon (1993, Table 20.1).

Table 2: Unit Root Tests in the Presence of Structural Shifts for gnp_t and m_t Using Four Lagged Differences, a Constant, Seasonal Dummies and a Linear Trend

| Variable | Shift function | test statistic | 5% critical value |
|----------|------------------|----------------|-------------------|
| gnp_t | $d_{tT_B}\gamma$ | -1.41 | -3.03 |
| | exponential | -1.36 | |
| m_t | $d_{tT_B}\gamma$ | -1.70 | -3.03 |
| | exponential | -1.69 | |

Note: Critical value from Lanne et al. (2002).

Table 3: S&L Cointegration Tests for German Macro System, Sample Period: 1975Q1 – 1998Q4

| variables | deterministic terms | no. of lagged differences | $H_0 : \text{rk}(\Pi) = r_0$ | test statistic | 5% critical value |
|-------------------|--------------------------|---------------------------|------------------------------|----------------|-------------------|
| m_t, gnp_t | $c, tr, sd, shift$ | 0 | $r_0 = 0$ | 6.86 | 15.92 |
| | | | $r_0 = 1$ | 0.37 | 6.83 |
| | $c, orth\ tr, sd, shift$ | 4 | $r_0 = 0$ | 4.91 | 15.92 |
| | | | $r_0 = 1$ | 1.75 | 6.83 |
| | | 0 | $r_0 = 0$ | 9.13 | 9.79 |
| | | | | | |
| m_t, R_t | $c, tr, sd, shift$ | 0 | $r_0 = 0$ | 26.71 | 15.92 |
| | | | $r_0 = 1$ | 0.00 | 6.83 |
| | $c, orth\ tr, sd, shift$ | 0 | $r_0 = 0$ | 22.98 | 9.79 |
| | | | | | |
| | | 6 | $r_0 = 0$ | 8.26 | 15.92 |
| | | | $r_0 = 1$ | 0.20 | 6.83 |
| gnp_t, R_t | $c, tr, sd, shift$ | 0 | $r_0 = 0$ | 8.26 | 15.92 |
| | | | $r_0 = 1$ | 0.20 | 6.83 |
| | $c, orth\ tr, sd, shift$ | 6 | $r_0 = 0$ | 8.42 | 15.92 |
| | | | $r_0 = 1$ | 0.56 | 6.83 |
| | | 0 | $r_0 = 0$ | 4.04 | 9.79 |
| | | | 6 | $r_0 = 0$ | 9.36 |
| m_t, gnp_t, R_t | $c, tr, sd, shift$ | 0 | $r_0 = 0$ | 38.36 | 28.47 |
| | | | $r_0 = 1$ | 9.07 | 15.92 |
| | | $r_0 = 2$ | 0.00 | 6.83 | |
| | | 4 | $r_0 = 0$ | 19.58 | 28.47 |
| | | | $r_0 = 1$ | 4.93 | 15.92 |
| | | $r_0 = 2$ | 4.53 | 6.83 | |
| | $c, orth\ tr, sd, shift$ | 0 | $r_0 = 0$ | 33.62 | 20.66 |
| | | | $r_0 = 1$ | 9.47 | 9.79 |
| | | 4 | $r_0 = 0$ | 20.14 | 20.66 |
| | | | $r_0 = 1$ | 4.53 | 9.79 |

Notes: c - constant, tr - linear trend, $orth\ tr$ - linear trend orthogonal to the cointegration relations, sd - seasonal dummies, $shift$ - shift dummy variable $S90Q3$; critical values from Lütkepohl & Saikkonen (2000, Table 1) for models with unrestricted trend and from Saikkonen & Lütkepohl (2000b, Table 1) for models with trend orthogonal to the cointegration relations. This table is adapted from Lütkepohl (2004b).

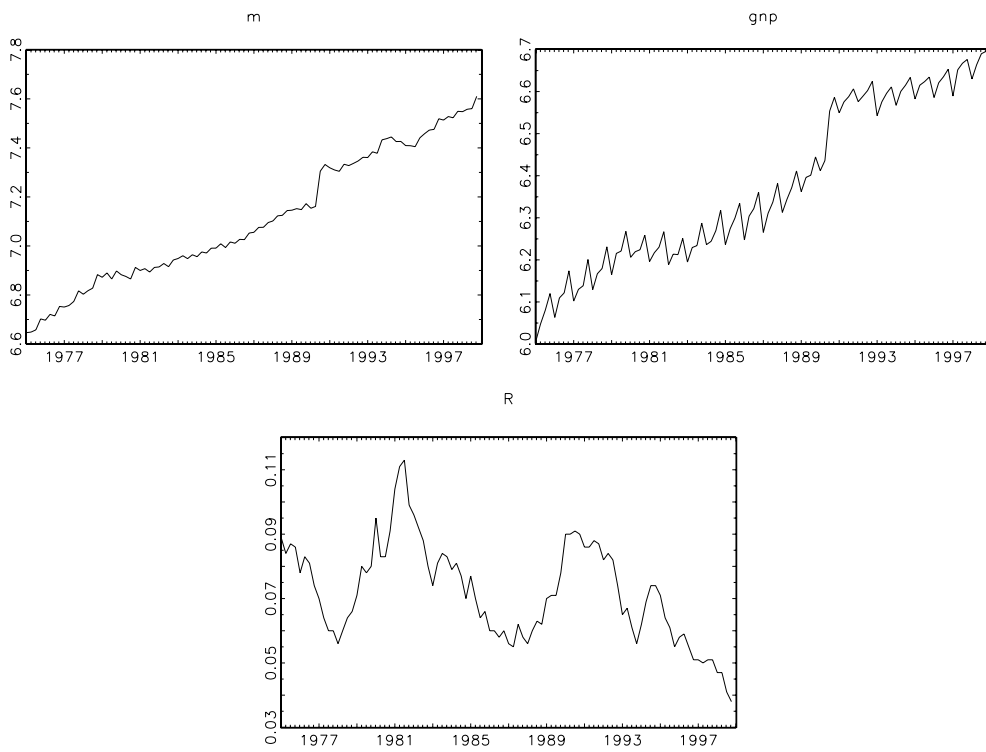


Figure 1: Seasonally unadjusted, quarterly German log real M3 (m), log real GNP (gnp) and average bond rate (R), 1975Q1 – 1998Q4.

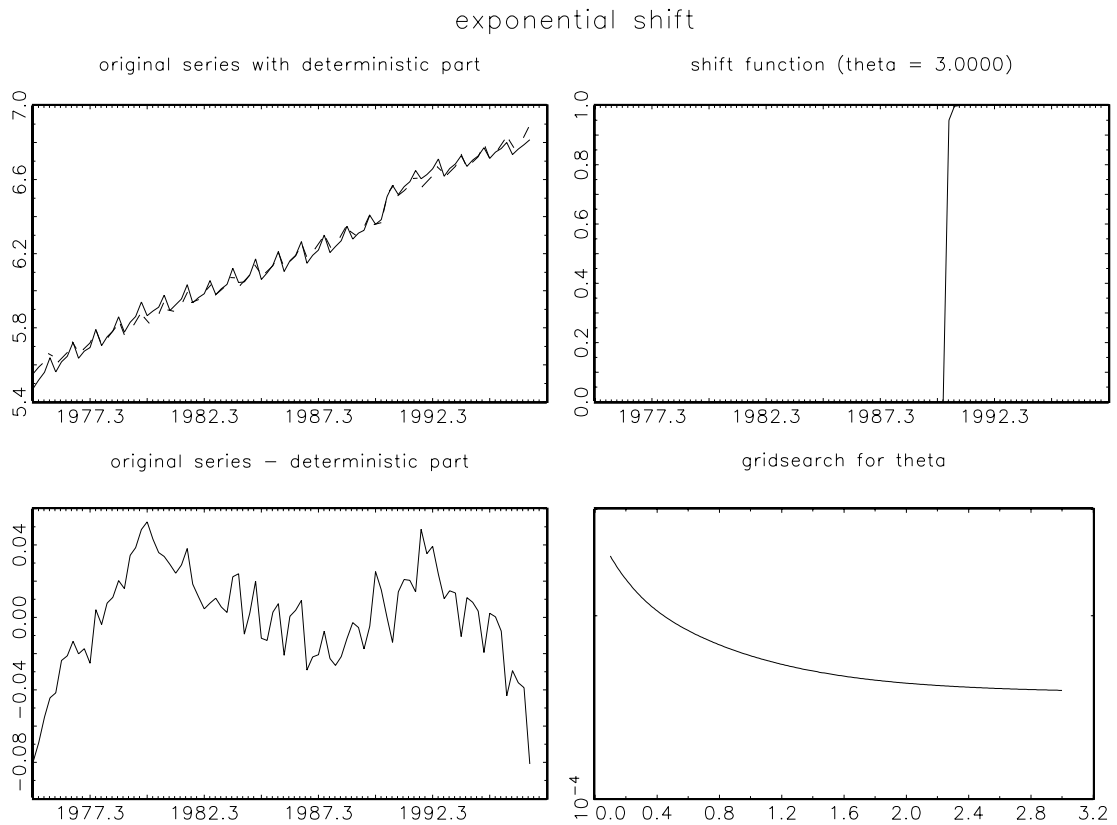


Figure 2: Deterministic terms and adjusted series used in unit root tests for log GNP series, based on a model with four lagged differences, sample period 1976Q2 – 1996Q4.
 (Note: The figure is extracted from Figure 2.18 of Lütkepohl (2004a).)

sample-split Chow test: bootstrapped p -values
(2000 repl.)

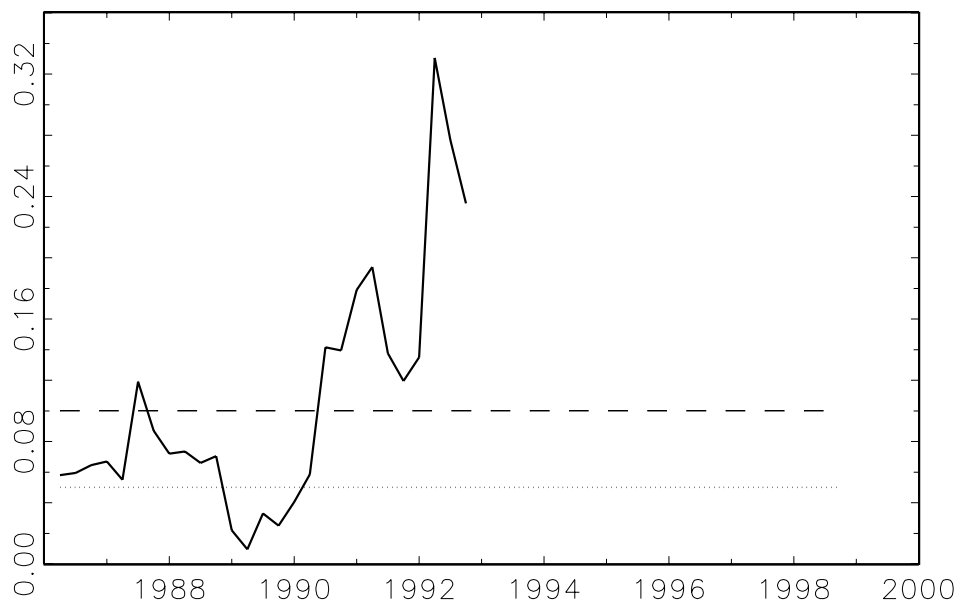


Figure 3: Chow test p -values for unrestricted VECM with cointegrating rank one, four lagged differences, constants, impulse dummy and seasonal dummies for German money demand system; sample period: 1975Q1 – 1998Q4 (including presample values).

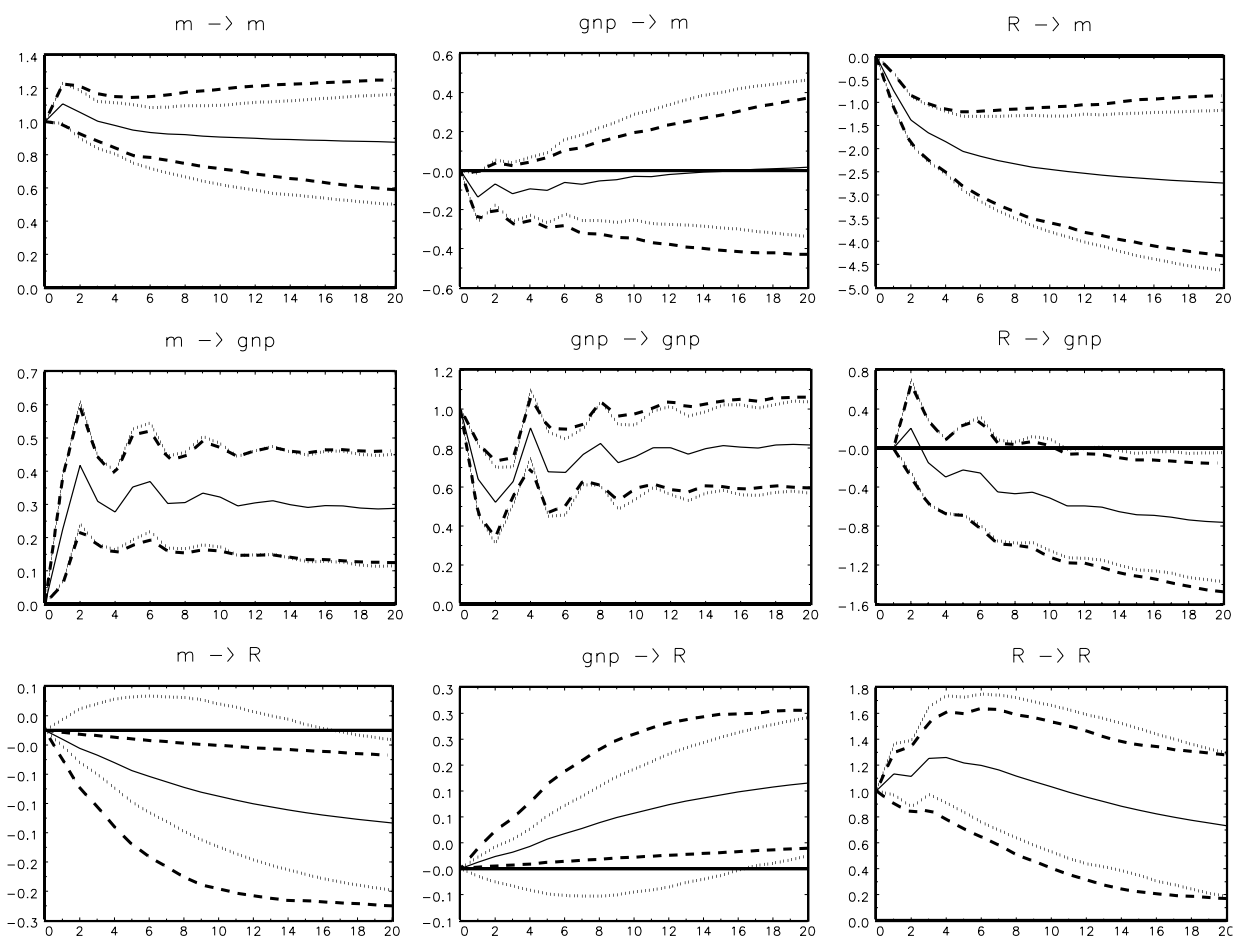


Figure 4: Forecast error impulse responses of German macro system based on subset VECM (16) with 95% standard (- - -) and Hall's percentile confidence intervals (.....) (2000 bootstrap replications).

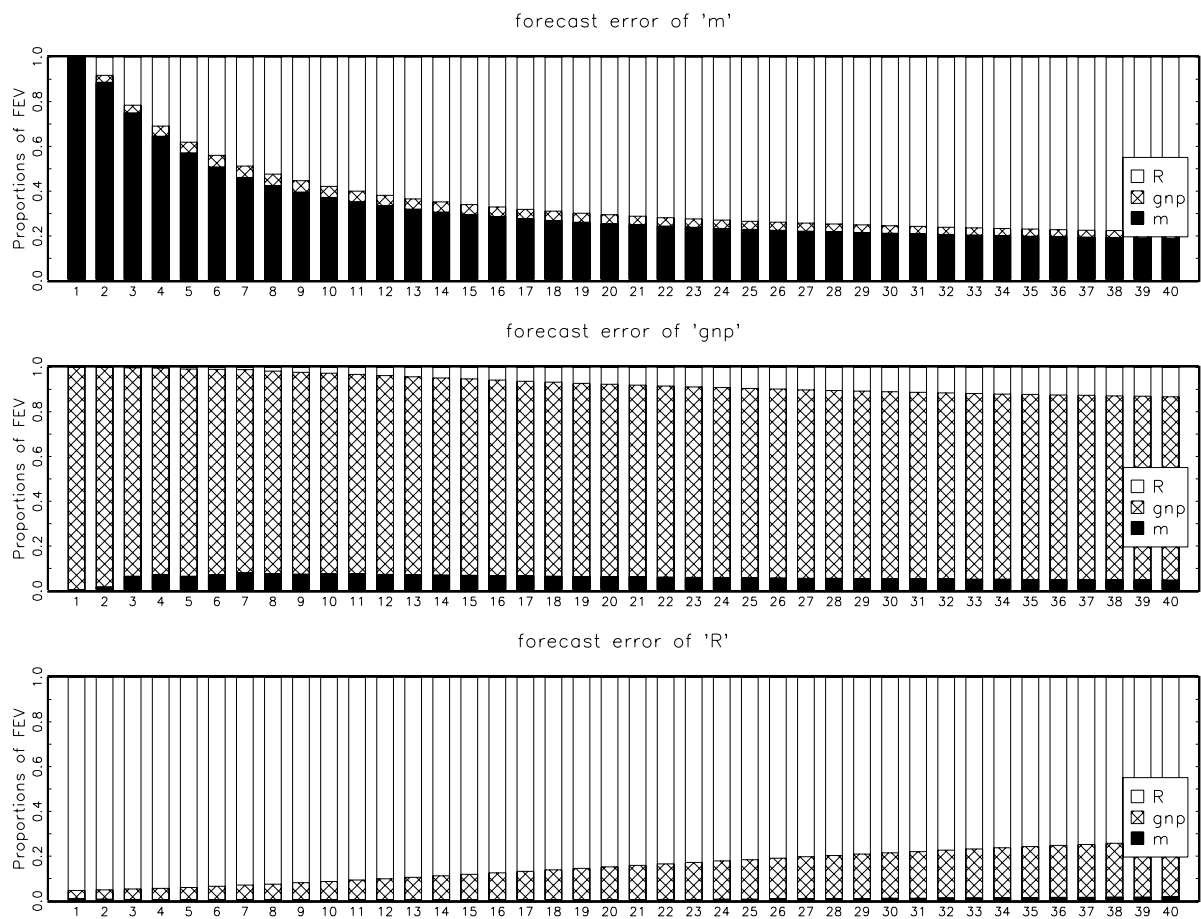


Figure 5: Forecast error variance decomposition of German macro system based on subset VECM (16).