Analyzing I(2) Systems by Transformed Vector Autoregressions

Hans Christian Kongsted*

Institute of Economics, University of Copenhagen

Heino Bohn Nielsen

Institute of Economics, University of Copenhagen

February 15, 2002

Abstract

This paper derives the restrictions that apply to a transformed model obtained by a minimal I(2)-to-I(1) transformation. The relationship between the parameters of the I(2) vector autoregression and the transformed model is characterized, including the coefficients of polynomially cointegrating relationships. In a simulation experiment, unrestricted reduced rank regression in the transformed model, which is common in applied work, yields only a minor loss of efficiency compared to imposing the restrictions. A properly transformed vector autoregression thus provides a practical and effective means for inference on the parameters of the I(2) model.

*Studiestræde 6, DK-1455 Copenhagen K, Denmark. E-mail Hans.Christian.Kongsted at econ.ku.dk, phone +45 3532 3076, fax +45 3532 3064.

1 Introduction

This paper gives a general characterization of the properties of a vector process, $\tilde{X}_t$, obtained by partly differencing a process integrated of order two,[1] X_t. The transformation applied eliminates I(2) trends while retaining possible cointegrating relationships among the variables. A transformed model is the starting point of much empirical work that involves the rate of growth of nominal variables, *e.g.*, the rate of inflation, wage growth, or the rate of money growth, and real or relative magnitudes such as real wagees, real money, or the markup, see Banerjee, Cockerell and Russell (2001) or Doornik, Hendry and Nielsen (1998) for examples.

The main interest in the present paper lies in analyzing if the transformed model can provide a practical *and* effective means for inference on the parameters of the original I(2) model. A case is examined in which the original process, X_t, is generated by a vector autoregression (VAR) that satisfies the I(2) conditions of Johansen (1992). The transformation examined is minimal in the sense of the amount of *a priori* information on the parameters required to achieve a valid reduction from I(2) to I(1). Throughout, the validity of *a priori* parameter restrictions is taken as given. Clearly, in empirical applications their validity should be tested.[2] A properly transformed process is shown to satisfy a VAR on which inference on the full set of cointegrating parameters can be achieved by standard I(1) methods. This includes the so-called polynomially cointegrating parameters of the I(2) model. Inference on the latter has so far proven elusive in the I(2) model.

The parameters of the transformed VAR are subject to restrictions. One set of restrictions is shown to fit directly into the standard I(1) reduced rank regression analysis, see Johansen (1996). Another set of restrictions requires different estimation techniques and is commonly ignored in applied work. The likely consequences of the resulting efficiency loss are explored in

[1] Denoted I(2), see Johansen (1996).

[2] Kongsted (1998) suggests a simple test of the transformation based on the two-step algorithm of Johansen (1995). Kongsted (1999) examines the properties of $\tilde{X}_t$ under invalid *a priori* parameter restrictions.

a small-scale simulation experiment.

Some notation and definitions are used throughout: For a $p \times r$ matrix α of rank r, let $\alpha_\perp$ denote a basis of the $p \times (p-r)$ orthogonal complement and define $\bar{\alpha} = \alpha(\alpha'\alpha)^{-1}$. For β $(p \times r)$ and η $(p-r \times s)$, $s < p-r$, define $\beta_1 = \bar{\beta}_\perp \eta$ and $\beta_2 = \beta_\perp \eta_\perp$. The matrices β, β_1, and β_2 are thus mutually orthogonal. Also note the relationship $I = V(B'V)^{-1}B' + b(v'b)^{-1}v'$ where $V(p \times (r+s)) = v_\perp$, $B = b_\perp$, and $|v'b| \neq 0$, see Hansen and Johansen (1998).

2 Transforming from I(2) to I(1)

This section derives the properties of the transformed process. Because the precise I(2) conditions play a major role in deriving the implied restrictions on the transformed VAR, this section will briefly set up the VAR of the original data based on Johansen (1992) and Rahbek, Kongsted and Jørgensen (1999). Then, the transformation is defined and a transformed VAR is derived. Finally, the restrictions are examined.

2.1 The I(2) process of the original data

The starting point of the analysis is a p-dimensional vector time series X_t satisfying the kth order vector autoregression written in error-correction format,

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu_0 + \mu_1 t + \varepsilon_t, \tag{1}$$

for $t = 1, \ldots, T$, or equivalently, in a parameterization suitable for I(2) processes,

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 X_{t-i} + \mu_0 + \mu_1 t + \varepsilon_t. \tag{2}$$

Here the definitions $\Gamma = I - \sum_{i=1}^{k-1} \Gamma_i$ and $\Psi_i = -\sum_{j=i+1}^{k-2} \Gamma_j$ apply and ε_t denotes identically and independently distributed $N(0, \Omega)$ terms. The initial observations, $X_{-k+1}, \dots, X_0$, are taken to be fixed.

Assuming that the roots of the characteristic polynomial of (2) are either at one or outside the unit circle and maintaining a further rank condition, the parameters of (2) should satisfy the following reduced rank conditions for X_t to be I(2),

$$\Pi = \alpha\beta' \quad \text{and} \quad \alpha'_\perp \Gamma \beta_\perp = \xi\eta', \tag{3}$$

see Johansen (1992).

In many practical situations a deterministic specification is applied which restricts the parameters of the constant and linear drift terms in (2) by

$$\alpha'_\perp \mu_0 = -\xi\eta'_0 - \alpha'_\perp \Gamma \bar{\beta} \beta'_0 \tag{4}$$

and

$$\mu_1 = \alpha\beta'_0, \tag{5}$$

where η'_0 and β'_0 are vectors of dimensions s and r. Essentially, (4) and (5) imply that a linear trend is allowed in all directions whereas X_t has no quadratic trend, see Rahbek *et al.* (1999).

The matrix β_2 can be shown to be proportional to the loadings of the common I(2) trends. The full set of $r + s$ cointegrating combinations $(\beta, \beta_1)' X_t$ can be divided into s combinations, $\beta'_1 X_t$, which remain I(1), and $\beta' X_t$ which cointegrate to stationarity with the first-difference of the I(2) component of the data in the r polynomially cointegrating relations,

$$S_t = \beta' X_t - \delta\beta'_2 \Delta X_t. \tag{6}$$

The polynomially cointegrating parameter, $\delta = \bar{\alpha}' \Gamma \bar{\beta}_2$ $(r \times (p - r - s))$, is an object of primary interest. So far it has proven elusive in terms of inference in the I(2) analysis. Obtaining inference on δ is a main motivation for analyzing the transformed process to which we turn next.

2.2 A transformed VAR

The particular transformation[3] considered here is defined by a known matrix b of dimension $p \times (p-r-s)$ which is assumed to satisfy orthogonality in terms of the full set of cointegrating vectors,

$$b'(\beta, \beta_1) = 0. \tag{7}$$

A transformed vector process is then defined as

$$\tilde{X}_t = \begin{pmatrix} B'X_t \\ v'\Delta X_t \end{pmatrix} \equiv \begin{pmatrix} Z_t \\ U_t \end{pmatrix} \tag{8}$$

where $B = b_\perp$ is $p \times (r+s)$ and v satisfies $|v'b| \neq 0$.

Under the condition (7) the matrix of I(2) common trend loadings, β_2, is known (up to a normalization) and b is a valid basis for β_2. Similarly, the full set of cointegrating vectors has a representation as $(\beta, \beta_1) = B(\varphi, (B'B)^{-1}\varphi_\perp)$. The transformation requires b and thus the number of I(2) trends in the system, $p-r-s$, or equivalently, $r+s$, to be known. No further assumptions are needed on the precise relationship between B and each set of cointegrating vectors, β and β_1. In most practical situations, there is an *a priori* expectation of one common I(2) trend, typically a nominal trend, shared by the variables of the vector process. Theory is often less informative about the precise number of polynomially cointegrating relationships, r.

The first-differenced term defined by v ensures that the linear combinations of first differences needed in order to recover the polynomially cointegrating relationships are included in the transformed system.

The properties of the transformed process, $\tilde{X}_t$, can now be characterized. Following Johansen (1996, Section 4.3) the conditions $\Pi = \alpha\beta'$ and $\mu_1 = \alpha\beta_0'$

[3] The transformation is denoted a nominal-to-real transformation in Kongsted (1999) since it typically involves going from a system in nominal variables to a real system. In the present paper the transformation is assumed to achieve the reduction from I(2) to I(1) whereas Kongsted (1999) examines the general case where that is not necessarily true.

are imposed and (2) is premultiplied by $(B,v)'$ to obtain

$$\begin{pmatrix} B'\Delta^2 X_t \\ v'\Delta^2 X_t \end{pmatrix} = \begin{pmatrix} B'\alpha\beta' X_{t-1} \\ v'\alpha\beta' X_{t-1} \end{pmatrix} - \begin{pmatrix} B'\Gamma\Delta X_{t-1} \\ v'\Gamma\Delta X_{t-1} \end{pmatrix} + \sum_{i=1}^{k-2} \begin{pmatrix} B'\Psi_i\Delta^2 X_{t-i} \\ v'\Psi_i\Delta^2 X_{t-i} \end{pmatrix}$$
$$+ \begin{pmatrix} B'\alpha\beta_0' t \\ v'\alpha\beta_0' t \end{pmatrix} + \begin{pmatrix} B'(\mu_0+\varepsilon_t) \\ v'(\mu_0+\varepsilon_t) \end{pmatrix}. \qquad (9)$$

Using the definitions of Z_t and U_t, substituting $\Delta X_t = c\Delta Z_t + aU_t$ in (9) with $c = V(B'V)^{-1}$ and $a = b(v'b)^{-1}$, and finally collecting terms yields a VAR(k) in error correction format for the transformed variables, $\tilde{X}_t$,

$$\Delta\tilde{X}_t = \tilde{\Pi}\tilde{X}_{t-1} + \sum_{i=1}^{k-1} \tilde{\Gamma}_i \Delta\tilde{X}_{t-i} + \tilde{\mu}_1 t + \tilde{\mu}_0 + \tilde{\varepsilon}_t. \qquad (10)$$

This is the standard representation for cointegration analysis, see Johansen (1996). It provides the starting point of much empirical analysis.

2.3 Restrictions implied by the transformation

The parameters of (10) are subject to restrictions deriving from the transformation. Comparing (9) and (10) the parameters of the transformed VAR are given by:

$$\tilde{\Pi} = \begin{pmatrix} B'\alpha\varphi' & -B'\Gamma a \\ v'\alpha\varphi' & -v'\Gamma a \end{pmatrix},$$

$$\tilde{\Gamma}_1 = \begin{pmatrix} I + B'(\Psi_1 - \Gamma)c & B'\Psi_1 a \\ v'(\Psi_1 - \Gamma)c & v'\Psi_1 a \end{pmatrix},$$

$$\tilde{\Gamma}_i = \begin{pmatrix} I + B'(\Psi_i - \Psi_{i-1})c & B'\Psi_i a \\ v'(\Psi_i - \Psi_{i-1})c & -v'\Psi_i a \end{pmatrix}, \quad i = 2, \ldots, k-2,$$

$$\tilde{\Gamma}_{k-1} = \begin{pmatrix} -B'\Psi_{k-2}c & 0 \\ -v'\Psi_{k-2}c & 0 \end{pmatrix},$$

$$\tilde{\mu}_0 = \begin{pmatrix} B'\mu_0 \\ v'\mu_0 \end{pmatrix}, \quad \tilde{\mu}_1 = \begin{pmatrix} B'\alpha\beta_0' \\ v'\alpha\beta_0' \end{pmatrix}, \quad \tilde{\varepsilon}_t = \begin{pmatrix} B'\varepsilon_t \\ v'\varepsilon_t \end{pmatrix}.$$

In practical terms, the restrictions on the transformed VAR fall in two categories. First, there are restrictions related to the fact that the coefficient matrix of the levels term, $\tilde{\Pi}$, is of reduced rank, r. Kongsted (1999) shows that when (7) is satisfied X_t is an I(1) process and rank($\tilde{\Pi}$) $= r$. To see the relationship with δ it is illuminating to briefly sketch the argument here as well. Johansen (1992) shows that for the I(2) process it holds that

$$\Gamma = \Gamma\bar{\beta}\beta' + (\alpha\bar{\alpha}'\Gamma\bar{\beta}_1 + \alpha_1)\beta_1' + \alpha\bar{\alpha}'\Gamma\bar{\beta}_2\beta_2'. \tag{11}$$

Using b as a valid basis of β_2 under (7) it is found that $\Gamma a = \alpha\tilde{\delta}$ where $\tilde{\delta} = \delta b'b(v'b)^{-1}$. Substituting in the expression for $\tilde{\Pi}$, it is seen to be the product of $p \times r$ matrices

$$\tilde{\beta} = \begin{pmatrix} \varphi \\ \tilde{\delta}' \end{pmatrix}, \quad \tilde{\alpha} = \begin{pmatrix} B'\alpha \\ b'\alpha \end{pmatrix},$$

which define the cointegrating relationships among the transformed variables, Z_t and U_t, and the loadings into the transformed process. The former is of special interest since it reflects the polynomially cointegrating relationships of the original I(2) process.

The parameter $\tilde{\delta}$ can be estimated from the transformed process as an I(1) cointegrating parameter and inference on the polynomially cointegrating relationship (6), including δ, can thus be obtained by from the transformed VAR. Also related to the reduced rank of $\tilde{\Pi}$ is the coefficient of the linear trend which in (9) can be seen to be restricted corresponding to (5), that is,

$$\tilde{\mu}_1 = \tilde{\alpha}\beta_0'. \tag{12}$$

The reduced rank of $\tilde{\Pi}$ and (12) are straightforward to implement in the reduced rank regression algorithm for I(1) models with a restricted linear term, see Johansen (1996).

A second set of restrictions on the parameters of the transformed VAR does not fall naturally into the reduced rank regression framework. First, the coefficients of the last lag, ΔU_{t-k+1}, are zero which restricts the last columns of $\tilde{\Gamma}_{k-1}$. Secondly, if the I(2) process has restricted deterministic terms, conditions such as (4) carry over to the transformed model. Specifically, (4) serves to exclude the possibility of any quadratic trends in X_t, see Rahbek *et al.* (1999). The first-differenced component of $\tilde{X}_t$, $U_t = v'\Delta X_t$, must therefore be without linear trend. Both restrictions require other estimation algorithms than standard reduced rank regression and are commonly ignored in applied work. The simulation experiment below will assess the importance of the resulting efficiency loss.

3 Simulation experiment

Unrestricted estimation of (10) allows for a redundant lag in $v'\Delta X_t$ and for a redundant linear trend in that variable.[4] In this section we want to evaluate the likely consequences of using an unrestricted VAR for analysis of the transformed variables rather than the restricted one.

Simulation setup As the data generating process (DGP) for the simulations we set up a simple model for a trivariate vector process, Y_t, which could encompass e.g. a money demand analysis with $Y_t = (m_t, y_t, p_t)'$, m_t being nominal money (in logs), y_t nominal income, and p_t the price index. It is a simplified version of the DGP in Kongsted (1999) used for analysing the properties of tests of the nominal-to-real transformation. Here, the single I(2) trend feeds proportionally into all three variables and a nominal-to-real transformation is a valid reduction from I(2) to I(1).

To be specific, the stochastic part of the process is defined by the three generating equations

$$\begin{aligned} m_t - y_t &= \kappa \Delta p_{t-1} + \epsilon_{1t} - \epsilon_{2t} && (13) \\ \Delta y_t &= \Delta p_{t-1} + \epsilon_{2t} && (14) \\ \Delta^2 p_t &= \epsilon_{3t}, && (15) \end{aligned}$$

[4]Furthermore, one observation is lost by applying the transformation. In this section, however, we ignore this effect by assuming to have one unused pre-sample observation.

and we assume $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})' \sim i.i.d.N\,(0, I)$. Equation (13) mimics a polynomially cointegrating velocity equation, where the parameter $\kappa = -6$ reflects a typical estimate of the interest rate elasticity of real money demand, e.g. Doornik, Hendry and Nielsen (1998). Equation (15) defines the prices to be the I(2) trend of the system and (14) shows that this I(2) trend is embodied in nominal income. Written in the error correction form (2) the model is given by

$$\begin{pmatrix} \Delta^2 m_t \\ \Delta^2 y_t \\ \Delta^2 p_t \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} m_{t-1} \\ y_{t-1} \\ p_{t-1} \end{pmatrix} - \begin{pmatrix} 1 & 0 & -(1+\kappa) \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta m_{t-1} \\ \Delta y_{t-1} \\ \Delta p_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}.$$

The levels matrix Π has reduced rank $r = 1$ and we can chose $\alpha = (-1, 0, 0)'$ and $\beta = (1, -1, 0)'$. It is easily verified that $\alpha'_\perp \Gamma \beta_\perp$ has reduced rank $s = 1$, that there is $p - r - s = 1$ I(2) trend in the process, and that Y_t satisfies the I(2) assumptions. Further we can choose $\beta_1 = (1, 1, -2)'$ and $\beta_2 = (1, 1, 1)'$ implying that the I(2) trend loads proportionally into the three nominal variables.

We generate samples of $T+100$ observations, $t = -101, -100, ..., 0, 1, 2, ..., T$ from this setup by replacing ϵ_t with pseudo-random independent drawings in Ox 3.0 (Doornik, 2001). We consider sample lengths of $T = \{25, 50, 75, 100, 150, 200\}$ effective observations and discard 100 presample observations to eliminate the importance of the initial values $Y_{-101}, Y_{-100} = 0$.

To be able to analyze the restrictions implied on the deterministic part of the process, we augment the DGP to allow for a linear trend in all directions, see Rahbek *et al.* (1999). The observed variables are defined from the factor representation

$$X_t = Y_t + \theta_0 + \theta_1 t. \tag{16}$$

The test statistics and parameter estimates for the I(1) model with a restricted linear trend and the I(2) model with the deterministic specification of Rahbek *et al.* (1999) are invariant to the true coefficient to the trend term, θ_1, see Nielsen and Rahbek (2000). The important thing is not the true coefficient but the fact that we allow for it in the estimation. Without loss of generality we can therefore set the coefficient to $\theta_0 = \theta_1 = 0$.

The transformation is defined by

$$B = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

which can seen to satisfy (7) with $b = (1, 1, 1)'$ and $|v'b| \neq 0$. The transformed data vector is

$$\widetilde{X}_t = \begin{pmatrix} B'X_t \\ v'\Delta X_t \end{pmatrix} = \begin{pmatrix} m_t - p_t \\ y_t - p_t \\ \Delta p_t \end{pmatrix}.$$

For the transformed data we have an error correction model written via the factor representation

$$\Delta \widetilde{Y}_t = \widetilde{\alpha}\widetilde{\beta}'\widetilde{Y}_{t-1} + \widetilde{\Gamma}_1 \Delta \widetilde{Y}_{t-1} + \widetilde{\epsilon}_t \qquad (17)$$

$$\widetilde{X}_t = \widetilde{Y}_t + \widetilde{\theta}_0 + \widetilde{\theta}_1 t \qquad (18)$$

and it holds that $\widetilde{\alpha} = (-1, 0, 0)'$ and $\widetilde{\beta} = (1, -1, -\kappa)'$. The restrictions implied by the transformation are that the last column of $\widetilde{\Gamma}_1$ is zero, i.e.

$$H_0 : \widetilde{\Gamma}_1 = (\gamma_{3\times 2}, 0_{3\times 1}),$$

where γ contains the free parameters, and that the last element in $\widetilde{\theta}_1$ is zero, i.e.

$$H_1 : \widetilde{\theta}_1 = \left(\vartheta'_{1\times 2}, 0\right)',$$

where ϑ contains the free parameters.

In each of 1000 replications we estimate the unrestricted VAR(2) model corresponding to (17) and (18), which is the standard cointegrated VAR model with a restricted trend term and an unrestricted constant, $H^*(r)$, see Johansen (1996). Furthermore we estimate the model subject to each of the restrictions implied by the two true hypotheses H_0 and H_1. And we compare the properties of the three estimators to assess the importance of the efficiency gain implied by imposing the restrictions in H_0 and H_1.

Firstly, we consider the size and power properties of the Trace test for rank determination.[5] Secondly, we assess the precision of the estimated $\widetilde{\beta}$ in each of the three models with the true one. The comparison is based on the distance measure of Larsson and Villani (2001). In the present case with $r = 1$ we transform it to an angle (measured in degrees) between the estimated and the true $\widetilde{\beta}$ vector.

Unrestricted estimation First we consider the unrestricted model for the transformed variables $\widetilde{X}_t$, which is a cointegrated VAR(2) with a restricted trend term. Maximum likelihood estimation of this model reduces to solving the eigenvalue problem

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0,$$

where $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}$ are sample moment matrices, and R_{0t} and R_{1t} are least squares residuals of regressing $\Delta\widetilde{X}_t$ and $\left(\widetilde{X}'_{t-1}, t\right)'$ respectively on $V_t = \left(\Delta\widetilde{X}'_{t-1}, 1\right)'$, see Johansen (1996, chapter 6). This yields $p+1$ ordered eigenvalues $1 > \widehat{\lambda}_1 > \widehat{\lambda}_2 > ... > \widehat{\lambda}_p > \widehat{\lambda}_{p+1} = 0$ and the estimate of the cointegrating relations, $\left(\widehat{\widetilde{\beta}}', \widehat{\beta}_0\right)'$, is given by the eigenvectors corresponding to the r largest eigenvalues. Furthermore the likelihood ratio test for a reduced rank of r compared to the full rank alternative $H(p)$, can be written as a

[5] If the dataset is initially analyzed in an I(2) model and the validity of the transformation is tested, the rank will be determined in the I(2) step and no new rank determination is necessary in the I(1) model. Here we consider the rank determination anyway, since in many applications the transformation is not tested and the starting point is the unrestricted I(1) model for the transformed data.

function of the eigenvalues as the Trace test statistics

$$Q_r = -2 \log Q \left(H^* \left(r \right) \mid H^* \left(p \right) \right) = -T \sum_{i=r+1}^{p} \log \left(1 - \widehat{\lambda}_i \right).$$

The simulation results for the baseline unrestricted reduced rank regression are summarized in table 1A. First part of the table illustrates the average angle between the true and estimated parameter, $\widetilde{\beta}$. The precision is high and with 75 observations, say, the average angle is well below 1 degree. Only in the case of 25 observations the average angle is marked with a value of approximately 5 degrees. Also reported are the rejection frequency of the Trace test statistic on a nominal 5 per cent level. It appears that the rank determination procedure is effective even for relatively small samples. The test of $r \leq 1$ is somewhat oversized, but the power of rejecting $r = 0$ is extremely high.

The extraordinarily good performance of the unrestricted reduced rank regression reflects the properties of the GDP, in particular the very strong error correction implied by $\widetilde{\alpha} = (-1, 0, 0)'$. This means that the samples are very informative on the long-run relation. In practical applications results similar to the ones in table 1A would probably require more observations. The properties of the baseline case is not our main focus here and below we focus on the deviations from the baseline case.

Restriction on the lagged first differences Next we want to assess the importance of the redundant lag included for Δp_t. In each iteration we therefore estimate the model under the restrictions implied by the hypothesis H_0. To impose the restriction we modify the reduced rank regression described above. In particular we modify the vector of unrestricted variables to obtain

$$V_t^* = \begin{pmatrix} \left(I_2, 0_{2 \times 1} \right) \Delta \widetilde{X}_{t-1} \\ 1 \end{pmatrix},$$

which excludes Δp_{t-1}. Note that each equation of the VAR still contains the same set of variables and the effect can still be partialled out using least squares.

The results are given in table 1B. For a limited number of observations, the included obsolete lag in the unrestricted model seems to be important and the efficiency gain under H_0 eliminates half of the average distance. For larger sample sizes, however, the difference is negligible. Second part of the table reports the rejection frequency of the likelihood ratio test for the restriction implied by the hypothesis H_0. This test is simply comparing the likelihoods of the restricted and unrestricted model and is asymptotically distributed $\chi^2(3)$ under the null. The test for the hypothesis is significantly oversized in small samples but converges relatively fast to a level close to the nominal 5 per cent level. Last part reports the Trace test rejection frequencies. Imposing the restriction increases the empirical power of the Trace test in small samples.

Restriction on the trend term Finally we want to assess the efficiency loss of not having imposed the restriction implied by H_1. In each iteration we therefore want to estimate the model, which under H_1 is given by

$$\begin{aligned} \Delta\widetilde{Y}_t &= \widetilde{\alpha}\widetilde{\beta}'\widetilde{Y}_{t-1} + \widetilde{\Gamma}_1\Delta\widetilde{Y}_{t-1} + \widetilde{\epsilon}_t & (19) \\ \widetilde{X}_t &= \theta D_t, \qquad \textit{where} \quad \theta D_t = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} + \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ 0 \end{pmatrix} t. & (20) \end{aligned}$$

To estimate this model we apply the switching algorithm of Nielsen (2002) that allows for maximum likelihood estimation of the cointegrated I(1) model with additive deterministic components and therefore is capable of estimating a factor representation like (19) and (20). The idea is that conditional on an estimate, $\widehat{\theta}$, of the parameters, θ, to the additive components, D_t, the parameters of (19) can be estimated using a usual reduced rank regression of the corrected data $\widetilde{X}_t - \widehat{\theta}D_t$. With these estimates we can construct the estimated characteristic polynomial, $\widehat{A}(L)$, and the estimated residual, $e_t = \widehat{A}(L)\widetilde{X}_t$, which under the model is given by

$$\widehat{e}_t = \widehat{A}(L)\,\theta D_t + \epsilon_t. \tag{21}$$

For each of the n columns, D_{it} $i = 1, 2, ..., n$, in D_t, we define $\widehat{H}_{it} \equiv \widehat{A}(L) D_{it}$ and collect into $\widehat{H}_t = \left(\widehat{H}_{1t}, \widehat{H}_{2t}, ..., \widehat{H}_{nt}\right)$. We can then refomulate (21) into

$$\widehat{e}_t = \widehat{H}_t vec(\theta) + \epsilon_t, \tag{22}$$

where $vec(\theta)$ stacks the columns of θ. The likelihood function conditional on the parameters in $\widehat{A}(L)$ is maximized over θ by a GLS estimation in (22), i.e.

$$vec\left(\widehat{\theta}_j\right) = \left(\sum_{i=1}^{T}\left(\widehat{H}_t'\widehat{\Omega}^{-1}\widehat{H}_t\right)\right)^{-1}\left(\sum_{i=1}^{T}\left(\widehat{H}_t'\widehat{\Omega}^{-1}\widehat{e}_t\right)\right), \tag{23}$$

see Tsay, Peña and Pankratz (2000) and Saikkonen and Lütkepohl (2000) for a similar GLS step used in a two step estimator. Here we follow Nielsen (2002) and iterate between the two conditional steps until convergence. Due to the simple GLS estimation in step two, it is simple to impose the required restriction, see also Nielsen (2002).

Note, however, that the constants in (20) are not identified in the nonstationary directions and the procedure cannot be implemented directly on (19) and (20). We therefore modify the procedure slightly, considering a VAR model with a restricted constant and adding the linear trend only via the factor representation (20). To be specific we estimate the model

$$\begin{aligned}
\Delta\widetilde{Y}_t &= \widetilde{\alpha}\widetilde{\beta}'\widetilde{Y}_{t-1} + \widetilde{\Gamma}_1\Delta\widetilde{Y}_{t-1} + \widetilde{\alpha}\psi + \widetilde{\epsilon}_t \\
\widetilde{X}_t &= \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \\ 0 \end{pmatrix} t
\end{aligned}$$

by the reduced rank regression and GLS iterative algorithm.

The results are summarized in table 1C. It is apparent, that results for the average angle and the rank determination are almost indistinguishable from the baseline case, implying that the efficiency loss from not imposing the restriction is minimal. We also report the rejection frequency of the test statistic for the hypothesis H_1, which is asymptotically $\chi^2(1)$. The size is fairly close to the nominal for all sample lengths.

4 Conclusions

This paper has derived the restrictions that apply to a transformed model obtained by a minimal I(2)-to-I(1) transformation. The relationship between the parameters of an I(2) vector autoregression and the transformed model is characterized, including the coefficients of polynomially cointegrating relationships.

Unrestricted reduced rank regression in the transformed model, which is common in applied work, is shown to yield only a minor loss of efficiency compared to imposing the restrictions in the simulation experiment. Most efficiency is gained by imposing the absence of redundant lags in the differenced I(2) process, which is a fairly simple restriction in terms of the restricted estimation procedure. Imposing the restriction on the trend coefficients, which requires a more involved iterative estimation algorithm, leads to a very marginal efficiency gain. It appears fairly safe to ignore this restriction in applied work.

In conclusion, a properly transformed vector autoregression provides a practical and effective means for inference on the parameters of the I(2) model.

References

[1] Banerjee, A., Cockerell, L., and Russell, B. (2001), An I(2) analysis of inflation and the markup, *Journal of Applied Econometrics*, 16, 221-240.

[2] Doornik, J.A. (2001), Object-Oriented Matrix Programming Using Ox, 4rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

[3] Doornik, J., Hendry, D. and Nielsen, B. (1998), Inference in Cointegrating Models: UK M1 Revisited, *Journal of Economic Surveys*, 12, 533-572.

[4] Hansen, P. R., and Johansen, S (1998): *Workbook on cointegration*, Oxford University Press.

[5] Johansen, S. (1992), A Representation of Vector Autoregressive Processes Integrated of Order 2, *Econometric Theory*, 8, 188-202.

[6] Johansen, S. (1995), A Statistical Analysis of Cointegration for I(2) Variables, *Econometric Theory*, 11, 25-59.

[7] Johansen, S. (1996), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, 2nd edition, Oxford University Press, Oxford.

[8] Kongsted, H. C. (1998), An I(2) Cointegration Analysis of Small-Country Import Price Determination, Discussion paper 98-22, December 1998, Institute of Economics, University of Copenhagen.

[9] Kongsted, H. C. (1999), Testing the nominal-to-real transformation, Mimeo, Institute of Economics, University of Copenhagen.

[10] Larsson, Rolf and Mattias Villani (2001), A distance measure between cointegration spaces, *Economics Letters*, 70, 21-27.

[11] Nielsen, Bent and Anders Rahbek (2000), Similarity Issues in Cointegration Analysis, *Oxford Bulletin of Economics and Statistics*, 62:1, 5-22.

[12] Nielsen, Heino (2002), Cointegration Analysis in the Presence of Outliers, Mimeo, Institute of Economics, University of Copenhagen.

[13] Rahbek, A., Kongsted, H. C., and Jørgensen, C. (1999), Trend Stationarity in the I(2) Cointegration Model, *Journal of Econometrics*, 90, 265-289.

[14] Saikkonen, Pentti and Helmut Lütkepohl (2000), Trend Adjustment Prior to Testing for the Cointegration Rank of a VAR Process, *Journal of Time Series Analysis*, 21:4, 435-456.

[15] Tsay, Ruey S. , Daniel Peña and Alan E. Pankratz (2000), Outliers in multivariate time series, *Biometrika*, 87:4, 789-804.

Table 1: Simulation results, 1000 replications.

			Rejection frequency (per cent)		
T	Average angle	Difference in angle to baseline	LR test VAR restriction	Trace test $r = 0$	Trace test $r \leq 1$
	A. Baseline: Reduced rank regression				
25	4.852	-	-	66.90	10.20
50	1.208	-	-	97.30	9.30
75	0.728	-	-	100.00	7.30
100	0.493	-	-	100.00	6.60
150	0.317	-	-	100.00	6.20
200	0.234	-	-	100.00	4.50
	B. Lag restriction imposed				
25	2.666	-2.187	28.70	100.00	13.90
50	1.059	-0.149	12.30	100.00	9.40
75	0.665	-0.063	7.40	100.00	7.20
100	0.465	-0.028	7.70	100.00	7.10
150	0.306	-0.012	6.30	100.00	6.00
200	0.226	-0.008	6.00	100.00	4.50
	C. Trend restriction imposed				
25	4.746	-0.107	9.00	62.40	8.50
50	1.205	-0.003	7.20	97.00	8.50
75	0.727	-0.001	6.70	100.00	6.40
100	0.493	-0.000	4.80	100.00	6.20
150	0.317	-0.000	4.90	100.00	5.80
200	0.234	-0.000	5.50	100.00	4.00