An I(2) Cointegration Model with Piecewise Linear Trends

Kurita, Takamitsu; Nielsen, Heino Bohn; Rahbek, Anders Christian

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Kurita, T., Nielsen, H. B., & Rahbek, A. C. (2009). An I(2) Cointegration Model with Piecewise Linear Trends: Likelihood Analysis and Application. Department of Economics, University of Copenhagen.

Discussion Papers
Department of Economics
University of Copenhagen

No. 09-13

An I(2) Cointegration Model With Piecewise Linear Trends: Likelihood Analysis And Application

Takamitsu Kurita, Heino Bohn Nielsen,
and Anders Rahbek

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (online)

AN I(2) COINTEGRATION MODEL WITH PIECEWISE LINEAR TRENDS: LIKELIHOOD ANALYSIS AND APPLICATION

Takamitsu Kurita*, Heino Bohn Nielsen† and Anders Rahbek‡

July 6, 2009

Abstract: This paper presents likelihood analysis of the I(2) cointegrated vector autoregression with piecewise linear deterministic terms. Limiting behavior of the maximum likelihood estimators are derived, which is used to further derive the limiting distribution of the likelihood ratio statistic for the cointegration ranks, extending the result for I(2) models with a linear trend in Nielsen and Rahbek (2007) and for I(1) models with piecewise linear trends in Johansen, Mosconi, and Nielsen (2000). The provided asymptotic theory extends also the results in Johansen, Juselius, Frydman, and Goldberg (2009) where asymptotic inference is discussed in detail for one of the cointegration parameters. To illustrate, an empirical analysis of US consumption, income and wealth, 1965 – 2008, is performed, emphasizing the importance of a change in nominal price trends after 1980.

Keywords: Cointegration, I(2), Piecewise linear trends, Likelihood analysis, US consumption.
JEL Classification: C32.

1 Introduction

This paper presents the complete asymptotic likelihood analysis of the I(2) cointegrated vector autoregression (VAR) with piecewise linear trends, i.e. a model where the slopes of the deterministic trends and the equilibrium means are allowed to change at q known breakpoints. Our aim is to provide the asymptotic analysis with a focus on making inference on the cointegration ranks and testing hypotheses on the cointegrating parameters based on likelihood ratio (LR) statistics. Thus we derive in Theorem 2 the asymptotic distributions of the maximum likelihood estimators (MLEs) of the parameters based on

The authors thank David F. Hendry, Søren Johansen and Bent Nielsen for helpful comments.

* Faculty of Economics, Fukuoka University, Bunkei Center Building, 8-19-1 Nanakuma, Johnanku, Fukuoka, 814-0180, Japan. Supported by grant JSPS KAKENHI (19830111)

† Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark.

‡ Department of Economics, University of Copenhagen and CREATES.

a normalization suitable for deriving the limit distribution of the rank test statistic, see Corollary 2, and LR statistics on the cointegration parameters (τ^*, see below) also discussed in Johansen *et al.* (2009). We thereby extend the analysis in Nielsen and Rahbek (2007), where cointegration rank testing is considered for I(2) models including a constant linear trend and level, and the analysis in Johansen *et al.* (2000), where the I(1) cointegration rank test is considered for models with piecewise linear trends. The paper complements the results in Johansen *et al.* (2009) by presenting limiting behavior of all estimators, also necessary for the results therein.

A main issue in the asymptotic analysis is the role of so-called impulse dummies as induced in the model by the inclusion of changing linear trends and levels, in addition to impulse dummies included in the econometric analysis to improve the fit. We demonstrate that the parameters loading the impulse dummies are inconsistent, but bounded in probability, which again implies that they play no role in the asymptotic distribution of the MLEs or the rank test statistic.

Empirically, the I(2) model with piecewise linear trends appears to be highly relevant. Many OECD countries have experienced pronounced shifts in inflation rates since the 1960'ties, leading to smooth changes in the trend slopes of nominal variables, and time series for nominal variables over the post-World War II period seems to be well described as autoregressive processes integrated of order two, I(2), see *inter alia* Juselius (1998; 1999), Diamandis, Georgoutsos, and Kouretas (2000), Banerjee, Cockerell, and Russell (2001), Fliess and MacDonald (2001), Nielsen (2002), Bacchiocchi and Fanelli (2005), and Nielsen and Bowdler (2003) for applications of the cointegrated I(2) model. More abrupt changes in trend slopes are often related to new institutional regimes, and a simple modelling alternative in that case would be is to allow deterministic changes in trend growth for nominal variables. In the light of visible changes in mean growth rates, deterministic changes in trend slopes are undoubtedly a more relevant alternative to the hypothesis of double unit roots than a constant trend. More generally, it is known to be extremely important to have a relevant deterministic specification of the model before the presence of unit roots is tested, see *inter alia* Perron (1989).

To illustrate the use of the I(2) VAR with piecewise linear trends, the methodology is applied to quarterly observations of nominal variables for US consumption, income and wealth, 1965 – 2008. We find a significant difference in the trend slope before and after 1981, a break that can be attrributed to a shift in policy focus following the stagflation period and the recession in 1981. Based on the LR test we find clear evidence of I(2) trends in the nominal variables, also when we allow for the deterministic change in the trend. In the model with a piecewise linear trend we accept homogeneity between nominal variables; this excludes money illusion in the long-run, and facilitates a nominal-to-real transformation from I(2) to I(1), Kongsted (2005), so that the equilibrium relationships may be formulated in real magnitudes for consumption, income, and wealth together with an interest rate and inflation. Homogeneity, and hence the validity of the theoretically relevant I(2)-to-I(1) transformation, is strongly rejected in a constant trend model.

The organization of this paper is as follows. Section 2 introduces the relevant representations of the VAR model for I(2) processes in the presence of changing linear trends. Section 3 then investigates the limiting behavior of the MLEs and the LR statistic for cointegration ranks. Finally, Section 4 presents the empirical illustration. Proofs are given in the appendix.

Throughout use is made of the following notation: for any $p \times r$ matrix α of rank r, $r < p$, let $\alpha_\perp$ indicate a $p \times (p-r)$ matrix whose columns form a basis of the orthogonal complement of span(α). Set $\bar{\alpha} = \alpha\,(\alpha'\alpha)^{-1}$ such that $\overline{\alpha}\alpha' = \alpha\overline{\alpha}'$ is the orthogonal projection matrix onto span(α). The symbols $\overset{D}{\to}$ and $\overset{P}{\to}$ are used to indicate weak convergence and convergence in probability respectively. Finally, we use $[x]$ to denote the largest integer smaller than x, $x \in \mathbb{R}$, and $1\,(A)$ the indicator function which equals one if A is true, zero otherwise.

2 The Model

2.1 The I(2) model with no deterministic terms

To introduce the notation consider initially the unrestricted VAR model with $k \geq 2$ lags and parametrized conveniently for I(2) analysis of the p-dimensional X_t,

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \Psi \Delta^2 \mathbb{X}_{t-1} + \epsilon_t, \quad t = 1, 2, ..., T. \tag{1}$$

Here Π and Γ are $(p \times p)$-dimensional matrices, $\Psi\Delta^2\mathbb{X}_{t-1} = \sum_{i=1}^{k-2} \Psi_i \Delta^2 X_{t-i}$, with Ψ_i $(p \times p)$ matrices and ϵ_t is a $p-$dimensional i.i.d. $N_p(0, \Omega)$ sequence, $\Omega > 0$. Furthermore, the initial values $X_0, \Delta X_0$, and $\Delta^2\mathbb{X}_0$ are conditioned upon. The I(2) model, $H(r, s)$, is then defined by two reduced rank restrictions,

$$\Pi = \alpha\beta' \text{ and } \alpha_\perp' \Gamma \beta_\perp = \xi\eta', \tag{2}$$

with α and β $(p \times r)$ dimensional matrices, ξ and η are $(p-r) \times s$ matrices with $r \leq p$ and $s \leq p - r$. The two reduced rank restrictions lead to the following reparametrization for likelihoodbased estimation,

$$\Delta^2 X_t = \alpha[\rho'\tau' X_{t-1} + \psi' \Delta X_{t-1}] + \alpha_{\perp\Omega}\kappa'\tau'\Delta X_{t-1} + \Psi\Delta^2\mathbb{X}_{t-1} + \epsilon_t, \tag{3}$$

where ρ is $((r+s)\times r)$ dimensional, τ is $(p\times(r+s))$, ψ is $(p\times r)$, and κ is $((r+s) \times (p-r))$. Finally, $\alpha_{\perp\Omega} = \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}$ is $(p \times (p-r))$ dimensional.

To interpret the parameters and the dynamics of X_t we need the following assumption:

Assumption 1 *Assume that the characteristic polynomial,* $A(z) = I_p\,(1-z)^2 - \Pi z + \Gamma\,(1-z)\,z - \sum_{i=1}^{k-2} \Psi_i\,(1-z)^2 z^i$, *has exactly* $2(p-r) - s$ *roots at* $z = 1$ *and the remaining roots outside the unit circle.*

Under Assumption 1, $\Delta^2 X_t$, $s_t = \rho'\tau' X_t + \psi'\Delta X_t$ and $\tau'\Delta X_t$ all have a stationary representation and hence X_t is a (multi-)cointegrated I(2) process, see also Johansen (1997).

The original parameters in (1), imposing the reduced rank restrictions in (2), can be derived from the parameters in (3) as follows: First write write $\alpha_\perp = (\alpha_{\perp 1}, \alpha_{\perp 2})$ and $\beta_\perp = (\beta_{\perp 1}, \beta_{\perp 2})$ where $\alpha_{\perp 1} = \overline{\alpha}_\perp \xi$, $\beta_{\perp 1} = \overline{\beta}_\perp \eta$, $\alpha_{\perp 2} = (\alpha, \alpha_{\perp 1})_\perp$, $\beta_{\perp 2} = (\beta, \beta_{\perp 1})_\perp$. Then it holds that $\tau = (\beta, \beta_{\perp 1})$, $\beta = \tau\rho$, $\psi' = -\alpha_{\Omega^{-1}}\Gamma$, with $\alpha_{\Omega^{-1}} = \Omega^{-1}\alpha \left(\alpha'\Omega^{-1}\alpha\right)^{-1}$, and $\kappa' = -\alpha_\perp'\Gamma(\overline{\beta}, \overline{\beta}_{\perp 1}) = -(\alpha_\perp'\Gamma\overline{\beta}, \xi)$, using the skew-projection identity,

$$\alpha\alpha_{\Omega^{-1}}' + \alpha_{\perp\Omega}\alpha_\perp' = I_p. \tag{4}$$

Furthermore, the parameters α, ρ, τ, ψ, κ, Ψ and Ω are all freely varying and estimates are obtained by a switching algorithm: For fixed τ, the parameters $\alpha_\perp$ and α can be obtained by solving an eigenvalue problem and the remaining parameters can be found from ordinary linear regression. For fixed values of these parameters, τ can be estimated by generalized least squares, see Johansen (1997) for more details.

2.2 Deterministic terms

Our focus will be on the inclusion of piecewise linear trends in the I(2) model. Specically, we allow for a linear deterministic trend and q changes in the trend slopes and equilibrium levels. The deterministic terms enter the model to allow piecewise linear trends in all directions of the process, including the multi-cointegrating relationships, and are restricted to avoid quadratic and higher order trends.

Let therefore $D_t = (D_{0t}, D_{1t}..., D_{qt})'$ denote a generic $(q+1)$-dimensional deterministic linearly trending variable, and set $d_t = \Delta D_t$. Make the following assumption:

Assumption 2 *For the deterministic $(q+1)-$dimensional linear trend $D_t = (D_{0t}, ..., D_{qt})'$ assume that with $u \in [0,1]$, $T^{-1}D_{[Tu]} \to D_u$ on the space of $(q+1)$-dimensional cadlag functions on [0,1], where*

$$T^{-1}D_{i[Tu]} \to D_{iu}, \quad as \quad T \to \infty, \ i = 0, 1, ..., q.$$

Furthermore, it is assumed that, as $T \to \infty$,

$$T^{-3}\sum_{t=1}^{T} D_t D_t' \to \int_0^1 D_u D_u' du,$$

which is a positive difinite $(q+1) \times (q+1)$ matrix.

Set $D_{0t} = t$ and hence $D_{0u} = u$, that is, the first component of D_t is throughout a linear trend, while D_{it} for $i = 1, ..., q$ allow q linearly independent changing linear trends. A changing trend slope at say $t = T_1$ with $1 < T_1 < T$, can be represented by defining $D_{1t} = (t - (T_1 - 1))\, 1\,(t \geq T_1)$, such that $D_{1u} = (u - u_1)\, 1\,(u \geq u_1)$, with $u_1 \in]0, 1[$ satisfying $[Tu_1] = T_1$. Thus while T_1 denotes the time point of a change in the discrete time interval $[1, T]$, u_1 denotes the corresponding (limiting) fraction in the continuous time interval $[0, 1]$. Likewise for general T_i, $i = 1, 2, ..., q$.

With $d_t = \Delta D_t$ we have $d_{[Tu]} \to d_u \neq 0$ by Assumption 2. In terms of D_{1t} just defined, we have for example $d_{1t} = \Delta D_{1t} = 1\,(t \geq T_1)$ and hence $d_{1u} = 1\,(u \geq u_1)$.

2.2.1 Constant linear trend

The case of $D_t = D_{0t} = t$, which allows for a linear trend in all linear combinations of the I(2) process X_t, is analyzed in Rahbek, Kongsted, and Jørgensen (1999) and Nielsen and Rahbek (2007) and it is briefly reviewed here before introducing the changing linear trends, see also Paruolo (2000) for other specifications.

Let $D_t = t$, $d_t = 1$ and define $X_t^* = (X_t', D_t)'$. Then the I(2) model with a linear trend is conveniently given by,

$$\Delta^2 X_t = \alpha[\rho'\tau^{*\prime} X_{t-1}^* + \psi^{*\prime}\Delta X_{t-1}^*] + \alpha_{\perp\Omega}\kappa'\tau^{*\prime}\Delta X_{t-1}^* + \Psi\Delta^2\mathbb{X}_{t-1} + \epsilon_t \qquad (5)$$

where $\tau^* = (\tau', \tau_D')'$ $((p+1)\times(r+s))$, while $\psi^* = \left(\psi', \psi_d'\right)'$ $((p+1)\times r)$ and the remaining parameters are as in (3).

Under Assumption 1, it was shown in Rahbek *et al.* (1999: Theorem 2.1) that indeed X_t in (5) is an I(2) process with the representation,

$$X_t = C_2 \sum_{s=1}^{t}\sum_{i=1}^{s}\epsilon_i + C_1\sum_{i=1}^{t}\epsilon_i + \gamma_D D_t + \gamma_d d_t + C_0(L)\,\epsilon_t, \qquad (6)$$

$$C_2 = \beta_{\perp 2}(\alpha_{\perp 2}'\Theta\beta_{\perp 2})^{-1}\alpha_{\perp 2}', \quad \beta' C_1 = \bar{\alpha}'\Gamma C_2, \quad \beta_{\perp 1}' C_1 = \bar{\alpha}_{\perp 1}'(I - \Theta C_2).$$

Here $C_0(L)\,\epsilon_t = \sum_{i=1}^{\infty} C_i^0 \epsilon_{t-i}$ is a stationary mean-zero I(0) process with exponentially decaying coefficients[1] and $\Theta = \Gamma\overline{\beta}\overline{\alpha}'\Gamma + I_p - \sum_{i=1}^{k-2}\Psi_i$. The coefficients γ_D and γ_d for the trend and level, respectively, depend on τ_D and ψ_d as well as on the initial values of the process.

It follows from (6) that $\tau^{*\prime}X_t^* = \tau' X_t + \tau_D' D_t$ is I(1) whereas the $(p - r - s)$ linear combinations $\beta_{\perp 2}' X_t$ are I(2). In other words, $\tau^{*\prime}\Delta X_t^*$ and $\beta_{\perp 2}'\Delta^2 X_t$ are mean zero stationary, or I(0), processes in addition to the r mean-zero stationary linear combinations given by,

$$s_t^* = \rho'\tau^{*\prime}X_t + \psi^{*\prime}\Delta X_t^*.$$

2.2.2 Changing linear trend

One may view the resulting model in (5) as derived from the I(2) model with no deterministic terms in (3), replacing X_t by $(X_t', D_t, d_t)' = (X_t', t, 1)'$, and likewise for ΔX_t and $\Delta^2 X_{t-j}$. This would however lead to an overparametrized model, and instead by the analysis in Rahbek *et al.* (1999), this results in the model in (5) with X_t replaced by $X_t^* = (X_t', D_t)'$, and ΔX_t replaced by $\Delta X_t^* = (\Delta X_t', d_t)'$. Note that $\Delta^2 D_t = \Delta d_t = 0$ and $\Delta^2 X_{t-j}^* = \Delta^2 X_{t-j}$ enters unchanged.

[1] $||C_i^0|| < \nu^i$ with $0 \leq \nu < 1$.

Consider next extending D_t to include the additional q changing linear trends from Section 2.2. Initially, observe that in this case, with a changing linear trend such as $D_{1t} = (t - (T_1 - 1)) 1 (t \geq T_1)$, then $\Delta^2 D_{1t} = 1 (t = T_1) = \delta_{1t}$, say, where $\delta_{1t} \neq 0$. That is, the second order difference of the changing trend is an impulse dummy. Likewise with $\Delta^2 d_{1t} = \Delta\delta_{1t} \neq 0$, where $d_{1t} = \Delta D_{1t}$. Note that including $(\delta_{1t}, \Delta\delta_{1t})'$ as an unrestricted regressor is equivalent to include $(\delta_{1t}, \delta_{1t-1})'$, and below we include impulse dummies and not their differences. Introduce for that purpose δ_t, which is an m-dimensional variable of impulse dummies,

$$\delta_t = (\delta_{1t}, ..., \delta_{mt})', \quad \text{where } \delta_{it} = 1 (t = T_i) \text{ for some } T_i,\ 1 < T_i < T, \quad i = 1, 2, ..., m.$$

Then, similar to the constant trend case, we extend the I(2) model to allow for changing linear trends by including $X_t^* = (X_t', D_t')'$, $\Delta X_t^* = (\Delta X_t', d_t')'$ but now also impulse dummies in δ_t in the model, denoted $H^D (r, s)$:

$$\Delta^2 X_t = \alpha[\rho'\tau^{*\prime} X_{t-1}^* + \psi^{*\prime} \Delta X_{t-1}^*] + \alpha_{\perp\Omega}\kappa'\tau^{*\prime}\Delta X_{t-1}^* + \Psi\Delta^2\mathbb{X}_{t-1} + \Psi_\delta\delta_t + \epsilon_t, \tag{7}$$

where $\tau^* = (\tau', \tau_D')'$ is $((p + q + 1) \times (r + s))$, and $\psi^* = \left(\psi', \psi_d'\right)'$ is $((p + q + 1) \times r)$. The remaining parameters are as in (3), except the additional Ψ_δ $(p \times m)$ parameter. Note that the inclusion of the impulse dummies in δ_t as unrestricted regressors implies that $\hat{\epsilon}_{T_i} = 0$, where $\hat{\epsilon}_t$ are the estimated residuals in (7).

In empirical models, the m impulse dummies in δ_t are included for two different reasons. First of all, δ_t includes the impulse dummies resulting from the q changing linear trends (and levels). Specifically, with the example of D_{1t}, this leads to the inclusion of δ_{1t-j}, $j = 0, ..., k - 1$, which are k impulse dummies for $t = T_1 + j$. With q changing linear trends, a total of qk impulse dummies should thus be included in δ_t. As noted above, the k corresponding estimated residuals $\hat{\epsilon}_{T_1}, ..., \hat{\epsilon}_{T_1+(k-1)}$ all equal zero, and the inclusion of these k dummies is therefore equivalent to conditioning on $X_{T_1+(k-1)}, \Delta X_{T_1+(k-1)}$ and $\Delta\mathbb{X}^2_{T_1+(k-1)}$ in estimation. In addition to these qk *induced* impulse dummies, we allow for further impulse dummies δ_{it}, and hence $m \geq qk$. The additional impulse dummies entered as unrestricted regressors are sometimes referred to as *innovation dummies* and are common in empirical I(2) analyses since they often lead to a better empirical fit of the model within sample. We demonstrate below that they play no role in the asymptotic analysis, and the precise specification of δ_t is not important asymptotically. Likewise for so-called transitory impulse dummies, defined as $\Delta\delta_{it} = 1(t = T_1) - 1(t = T_1 + 1)$.

It follows directly by Rahbek *et al.* (1999) that the representation of X_t is identical to (6), with the only exception that now ϵ_t is replaced by $\epsilon_t^\delta = \epsilon_t + \Psi_\delta\delta_t$. We thus immediately get that under Assumption 1, X_t in (7) has the representation,

$$X_t = C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} \epsilon_i^\delta + C_1 \sum_{i=1}^{t} \epsilon_i^\delta + \gamma_D D_t + \gamma_d d_t + C_0 (L) \epsilon_t^\delta. \tag{8}$$

This was also used in Johansen *et al.* (2009: proof of Lemma 1) where a generic infinite sum of impulse dummies is introduced to faciliate the interpretation. Define here such a generic infinite sum,

$$\lambda_t = C_\delta(L)\,\delta_t \tag{9}$$

with $C_\delta(z) = \sum_{i=0}^{\infty} C_i^\delta z^i$, C_i^δ exponentially decreasing, and δ_t impulse dummies. The idea is that λ_t vanishes asymptotically as noted above and in this sense unimportant for the representation of X_t. For example, $C_0(L)\,\epsilon_t^\delta$ contains such a vanishing term, $C_0(L)\,\Psi_\delta\delta_t (= \lambda_t)$.

Thus, from (8) it holds that X_t is an I(2) process with broken linear trends and levels, and that $\Delta^2 X_t - E\left(\Delta^2 X_t\right)$ is I(0) with $E\left(\Delta^2 X_t\right) = \lambda_t$, a generic infinite sum of impulse dummies. Likewise, $\tau^{*\prime}\Delta X_t^* = \tau'\Delta X_t + \tau_D' d_t$ is I(0) except for $E\left(\tau^{*\prime}\Delta X_t^*\right) = \lambda_t$. Finally, the r linear combinations given by $\rho'\tau^{*\prime}X_t + \psi^{*\prime}\Delta X_t^*$ are I(0) except for $E\left(\rho'\tau^{*\prime}X_t + \psi^{*\prime}\Delta X_t^*\right) = \lambda_t$. Thus in this sense the interpretation remains identical to the linear trend case, except for the additional asymptotically vanishing infinite sums of impulse dummies, again generically referred to as λ_t. In the empirical application below, we illustrate the role of the impulse dummies and the interpretation of the deterministic terms.

3 Likelihood Inference

3.1 Estimation

Under $H^D(r,s)$, ML estimators in (7) are obtained by the usual switching algorithm described above for the I(2) model with no deterministic terms. Note that the loading the impulse dummies, Ψ_δ, is estimated from single observations only, and hence are bounded but inconsistent, see Theorem 1.

3.2 The rank test statistic

For determination of the cointegration ranks, r and s, we consider the LR statistic for $H^D(r,s)$ against the unrestricted alternative $H^D(p) = H^D(p,0)$, and it is defined by,

$$Q_{(r,s)}^{LR} = -T\log\left|\check{\Omega}\hat{\Omega}^{-1}\right|,$$

where $\check{\Omega}$ and $\hat{\Omega}$ denote the covariance matrices estimated under $H^D(r,s)$ and $H^D(p)$, respectively.

3.3 Asymptotics

When reporting results for the asymptotics of the parameter estimators emphasis will be on the parameters $\tau^* = (\tau', \tau_D')'$, $\psi^* = \left(\psi', \psi_d'\right)'$ and ρ. The parameters α, κ, Ψ and Ω have the same asympotic behaviour as in the model with no deterministic terms analysed in Johansen (1997). As shown the remaining parameter Ψ_δ plays no role for the asymptotic analysis, and we also note in this respect that $\hat{\Psi}_\delta$ is not consistent. We start by providing the necessary results for parameters in the I(2) model which are of theorectical interest, as their asymptotic distributions are used to derive the limiting behaviour of the LR statistics for rank and linear hypotheses respectively.

3.3.1 Theoretical parameters

In the following $\hat{\theta}$ denotes the ML estimator of a parameter θ, while θ_0 denotes the true value. Furthermore, the parameters β, τ and $\alpha_\perp$ under $H^D(r,s)$ are normalized on $\bar{\beta}_0$, $\bar{\tau}_0$ and $\bar{\alpha}_{0\perp\Omega}$ respectively such that $\bar{\beta}_0'\beta = I_r$, $\bar{\tau}_0'\tau = I_{r+s}$, and $\bar{\alpha}_{\perp\Omega 0}'\alpha_\perp = I_{p-r}$. These are *theoretically* convenient normalisations which ensure identification of all parameters in the model. Note in particular that $\rho = \bar{\tau}_0'\beta$ which is $(r+s)\times r$. Define next the parameters,

$$\begin{array}{rclrcl} B_0' &=& (\psi-\psi_0)'\,\bar{\beta}_{\perp 20} & B_1' &=& (\beta-\beta_0)'\,\bar{\beta}_{\perp 10} \\ B_D' &=& \rho'\,(\tau_D-\tau_{D0})' & & & \\ B_2' &=& (\beta-\beta_0)'\,\bar{\beta}_{\perp 20} = \rho'\,(\tau-\tau_0)'\,\bar{\beta}_{\perp 20} & C_0' &=& \rho_\perp'\,(\tau-\tau_0)'\,\bar{\beta}_{\perp 20} \\ B_d' &=& (\psi_d-\psi_{d0})' - (\psi-\psi_0)'\,\bar{\tau}_0\tau_{D0}' & C_D' &=& \rho_\perp'\,(\tau_D-\tau_{D0})' \end{array} \tag{10}$$

Note that B_0, B_1, B_2 and C are identical to the definitions in Johansen (1997), while B_D, B_d and C_D are new parameters corresponding to the deterministic terms.

We first turn to consistency of the just defined parameters, with the proof given in appendix:

Theorem 1 *For the model $H^D(r,s)$ under Assumption 1 the ML estimators exist with probability tending to one, and using the definitions in (10),*

$$\left(T^{1/2}\hat{B}_0', T^{1/2}\hat{B}_1', T^{3/2}\hat{B}_2', T\hat{B}_D', \hat{B}_d'\right)' \xrightarrow{P} 0 \quad \textit{and} \quad \left(T^{1/2}\hat{C}_0', \hat{C}_D'\right)' \xrightarrow{P} 0 \tag{11}$$

as $T\to\infty$. Moreover, $T^{1/2}(\hat{\rho}-\rho_0) \xrightarrow{P} 0$, and $\hat{\alpha}, \hat{\kappa}, \hat{\Psi}$ and $\hat{\Omega}$ are consistent. Finally, $\hat{\Psi}_\delta = O_P(1)$.

Theorem 1 establishes also rates of convergence and the next theorem gives the asymptotic distributions these estimators. To report these some definitions are needed first. Define first for X_u, Y_u and Z_u of dimension p_x, p_y and p_z defined on the unit interval $u\in[0,1]$,

$$\begin{array}{rcl} X_u|\,Y &=& X_u - \int_0^1 X_sY_s'ds\left(\int_0^1 Y_sY_s'ds\right)^{-1}Y_u, \\ S(X,Y,Z) &=& \int_0^1 dXY'\left(\int_0^1 Y_uY_u'du\right)^{-1}\int_0^1 Y dZ,' \\ R(Y,Z) &=& \left(\int_0^1 Y_uY_u'du\right)^{-1}\int_0^1 Y dZ'. \end{array} \tag{12}$$

And next define the process H_u by,

$$H_u = \left(H_{0u}', H_{1u}', H_{2u}'\right)' = \left(V_u'C_2'\beta_{\perp 2}, V_u'C_1'\beta_{\perp 1}, \int_0^u V_s'dsC_2'\beta_{\perp 2}\right)', \tag{13}$$

with V_u a Brownian motion on $u\in[0,1]$ with covariance Ω_0. Furthermore, define

$$V_{1u} = \left(\alpha_0'\Omega_0^{-1}\alpha_0\right)^{-1}\alpha_0'\Omega_0^{-1}V_u \tag{14}$$

$$V_{2u} = \left(\phi_0'\Omega_0^{-1}\phi_0\right)^{-1}\phi_0'\Omega_0^{-1}V_u \tag{15}$$

where $\phi_0' = \bar{\rho}_{\perp 0}'\kappa_0\alpha_{\perp\Omega 0}'$.

Theorem 2 *For the model $H^D(r,s)$ under Assumption 1,*

$$\left(T\hat{B}_0', T\hat{B}_1', T^2\hat{B}_2', T^{3/2}\hat{B}_D', T^{1/2}\hat{B}_d'\right)' \xrightarrow{D} B^\infty = \left(B_0^{\infty\prime}, B_1^{\infty\prime}, B_2^{\infty\prime}, B_D^{\infty\prime}, B_d^{\infty\prime}\right)' = R(H^*, V_1)$$
$$\left(T\hat{C}_0', T^{1/2}\hat{C}_D'\right)' \xrightarrow{D} C^\infty = (C_0^{\infty\prime}, C_D^{\infty\prime})' = R(H_0^*, V_2)$$

as $T \to \infty$. Here $H_u^* = (H_u', D_u', d_u')'$ and $H_{0u}^* = (H_{0u}', d_u')'$ with H_u defined in (13) and V_1, V_2 are defined in (14)-(15). Moreover, D_u and d_u are defined in Section 2.2.

Finally, $T(\hat{\rho} - \rho_0) \xrightarrow{D} \bar{\tau}_0'\beta_{\perp 10}B_1^\infty$ while the remaining parameters are asymptotically Gaussian. In particular, $T^{1/2}\left(\hat{\theta}^{z_0} - \theta_0^{z_0}\right) \xrightarrow{D} N_{p\times(2r+s+p)}\left(0, \Omega_0 \otimes \Sigma_{00}^{-1}\right)$, where θ^{z_0} defined in (27) in the appendix is the coefficient for the (asymptotically) stationary relations Z_{0t} in (26) of the model, and $\Sigma_{00} = Var(Z_{0t})$.

3.3.2 Asymptotics for hypotheses on individual parameters

From Theorem 2 limiting distributions for $\hat{\tau}^*$ and $\hat{\psi}^*$, normalized on known constants rather than as here the true parameters, are straightforward to derive using the definitions in (10) analogous to Johansen (1997: Theorems 3, 4 and 5). Likewise for other parameters in the model by exploiting their definitions in terms of the parameters B_D and C_D in Theorem 2. To examplify we derive the limiting distribution of $\hat{\tau}^* = \left(\hat{\tau}', \hat{\tau}_D'\right)'$ when τ is normalized by a known constant $p \times (r+s)$ dimensional matrix a say, that is

$$\hat{\tau}_a^* = \hat{\tau}^*\left(a'\hat{\tau}\right)^{-1} = \begin{pmatrix} \hat{\tau}_a \\ \hat{\tau}_{Da} \end{pmatrix}. \tag{16}$$

With a such that $a'\tau_0 = I_{r+s}$ we have the following corollary:

Corollary 1 *For $\hat{\tau}_a^*$ defined in (16) it follows that,*

$$\begin{pmatrix} T\bar{\beta}_{\perp 20}'(\hat{\tau}_a - \tau_0) \\ \sqrt{T}(\hat{\tau}_{Da} - \tau_{D0}) \end{pmatrix} \xrightarrow{D} C^\infty \bar{\rho}_{\perp 0}',$$

which is mixed Gaussian.

The proof is given in the appendix. An immediate implication of the result is that likelihood ratio tests for linear hypotheses as applied in Section 4 of the form $\tau^* = H\phi$, with H a known $((p+q+1)\times h)$-dimensional matrix, $r+s \leq h \leq p+q+1$ and ϕ $(h \times (r+s))$-dimensional, are asymptotically χ^2 distributed. A thorough discussion of hypothesis testing on the I(2) cointegration parameters τ as well as β is given in Johansen (2006) and Boswijk (2000) for the I(2) model with no deterministics, which in Johansen *et al.* (2009) are applied for a general discussion on χ^2-based inference on $\beta^* = \tau^*\rho$ and τ^* in the extended model here. Note in this respect, that Johansen *et al.* (2009) consider in particular the distribution of β^* under general and empirically relevant identifying restrictions. These results may also be derived from our Theorems 1 and 2.

3.3.3 Rank Test Asymptotics

From the results in Theorem 2, and using Nielsen and Rahbek (2007), we get the asymptotic distribution of the rank test statistic as a corollary, with proof given in the appendix:

Corollary 2 *Under Assumption 1, then as $T \to \infty$,*

$$Q_{(r,s)}^{LR} \overset{D}{\to} Q_r^{\infty} + Q_{(r,s)}^{\infty}, \tag{17}$$

where $Q_r^{\infty} = tr\left\{S\left(W, G^r, W\right)\right\}$ and $Q_{(r,s)}^{\infty} = tr\left\{S\left(W_2, G^{(r,s)}, W_2\right)\right\}$. Here $W = (W_1', W_2')'$ is a $(p-r)$dimensional standard Brownian motion, where W_1 is s-dimensional and W_2 is $(p-r-s)$dimensional. Furthermore, $G_u^r = \left(\left(W_{1u}', \int_0^u W_{2v}' dv, D_u'\right) \middle| G^{(r,s)}\right)$ and $G_u^{(r,s)} = (W_{2u}', d_u')'$ with $u \in [0,1]$.

The asymptotic distribution in (17) depends on q and the timing of the changing trend slopes, $(u_1, u_2, ..., u_q)$, and for empirical applications the distributions have to be simulated. Below we simulate this for a particular empirical example.

4 Empirical Illustration

To illustrate the theoretical results we conduct an empirical analysis of US quarterly consumption data, $1964 - 2008$. We consider the $p = 5$ dimensional vector:

$$X_t = (c_t, y_t, w_t, p_t, R_t)', \tag{18}$$

where c_t is nominal private consumption, y_t is nominal disposable income after tax, w_t is nominal wealth including financial wealth and housing equity, while p_t represents the price level measured as the consumption deflator. These variables are all transformed by natural logs. To capture interest rate effects on savings, we include the annual bond yield, R_t, divided by 4 to be comparable to a quarterly inflation rate, Δp_t. See Appendix B for details of the data. Similar data sets for real rather than nominal variables have been analyzed in *inter alia* Lettau and Ludvigson (2001) and Palumbo, Rudd, and Whelan (2006).

The time series are presented in Figure 1. In graph (A), the developments of the nominal variables, c, y and w, are quite parallel, although the wealth variable, w, fluctuates more. Recent discussions have referred to this as signs of 'bubbles' in asset and house prices. The price index, p, has increased less over time, but seems to share a similar smooth stochastic trend. We note that the trend slope seems to change just after 1980, both in the price deflator and in the nominal measures, and in the empirical analysis we allow for a deterministic change in the trend slope in $1981:2$. The shift in trend slope reflects a change in policy focus following the stagflation period in late 1970'ties. The US entered a severe recession in July 1981 partly initiated by a contractionary monetary policy to dampen inflation, cf. the inflation rate (Δp) and bond yield (b) in graph (B).

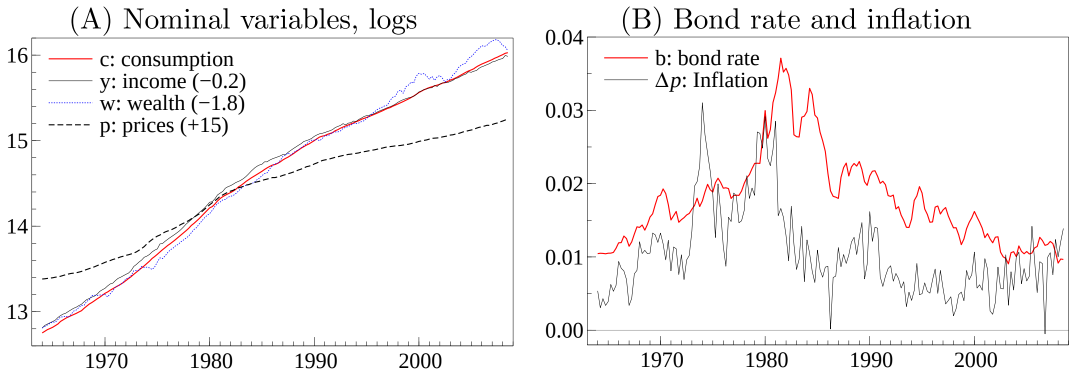

Figure 1: *US data for the empirical analysis, 1964-2008. The time series in graph (A) have been shifted to have comparable means.*

After the recovery of the US economy through 1982, the inflation rate stayed at more moderate values than the previous decade. Also note that inflation is clearly persistent, emphasizing the presence of I(2) trends in the data.

Statistical Analysis. The empirical analysis is based on a VAR with $k = 3$ lags and the effective sample contains the $T = 175$ observations from $1965 : 1$ to $2008 : 3$, hence conditioning on observations for $1964 : 2$, $1964 : 3$, and $1964 : 4$. The model incorporates in addition to the standard constant and linear trend term, a change in levels and trend slopes in $1981 : 2$, and hence three impulse dummies as well. The likelihood function of the unrestricted model seems to accounts for the main features of the data, and the hypotheses of no autocorrelation of order one and two are not rejected with $\chi^2(25)$ and $\chi^2(50)$ statistics of $\xi_{AR(1)} = 36$ and $\xi_{AR(2)} = 63$, respectively. There are several outlying residuals in the model, however, associated with special events and large shocks in the sample period, and the Jarque Bera test for the null hypothesis of Gaussian residuals is rejected with a $\chi^2(10)$ statistic of $\xi_{JB} = 178$. We will refer to this as the *baseline model* in the following.

To account for a number of the large shocks in the sample period, and to restore normality of the residuals, we also consider a version of the model that includes nine additional impulse dummies in δ_t, defined to take the value one in $1972 : 4$, $1974 : 1$, $1975 : 2$, $1980 : 2$, $1982 : 4$, $1984 : 2$, $1993 : 1$, $1999 : 4$, and $2008 : 2$, respectively. For this, the *augmented model*, the above hypotheses for no-autocorrelation and Gaussianity are not rejected ($\xi_{AR(1)} = 34$, $\xi_{AR(2)} = 58$, and $\xi_{JB} = 17$). Recall that the additional unrestricted impulse dummies do not change the asymptotic distributions of estimators and test statistics, and as we illustrate below, they only marginally change the finite sample results; in fact, all main conclusions of the empirical analysis are unchanged.

$p-r$	s	Mean	Variance	Quantiles .50	.80	.85	.90	.95	$\Gamma_{.95}$
5	0	209.660	294.066	208.913	223.838	227.433	232.067	239.016	238.64
5	1	178.362	259.383	177.648	191.724	195.142	199.490	206.014	205.65
5	2	151.031	221.336	150.372	163.273	166.399	170.566	176.708	176.31
5	3	127.583	190.067	126.927	138.917	141.814	145.576	151.364	151.07
5	4	108.089	163.135	107.485	118.601	121.281	124.696	130.066	129.92
5	5	92.253	141.580	91.671	101.901	104.441	107.779	112.899	112.66
4	0	144.721	204.845	143.982	156.606	159.633	163.557	169.449	169.04
4	1	119.242	175.203	118.586	130.074	132.980	136.650	142.119	141.82
4	2	97.787	147.598	97.158	107.853	110.375	113.689	118.774	118.59
4	3	79.946	123.639	79.353	88.999	91.419	94.512	99.278	99.07
4	4	66.078	102.288	65.433	74.238	76.529	79.418	83.791	83.55
3	0	91.270	134.829	90.672	100.737	103.303	106.570	111.380	111.17
3	1	71.688	108.736	71.025	80.182	82.473	85.388	89.847	89.66
3	2	55.865	86.523	55.309	63.450	65.491	68.152	72.128	71.99
3	3	43.869	69.429	43.214	50.625	52.520	54.928	58.672	58.41
2	0	49.588	73.772	48.976	56.496	58.404	60.943	64.804	64.51
2	1	35.549	56.247	34.897	41.548	43.225	45.502	48.828	48.72
2	2	25.347	41.418	24.705	30.483	31.960	33.901	36.889	36.78
1	0	19.109	29.493	18.480	23.424	24.678	26.369	28.958	28.83
1	1	10.825	18.675	10.162	14.160	15.203	16.597	18.843	18.77

Table 1: *Asymptotic distribution simulated from Corollary 2 with $q = 1$ and $u_1 = 0.377$. Based on random walks with 2000 steps and 50.000 replications. $\Gamma_{.95}$ is the 95% quantile of the approximating Γ distribution.*

Cointegration Ranks. To make inference on the cointegration ranks we first simulate the asymptotic distribution in (17) for the current $q = 1$ and $u_1 = 0.377$. This is done by replacing the Brownian motion W_u with a random walk with 2000 steps, replacing D_u with a discrete time trend function, and replacing $d_t = \Delta D_t$ with the corresponding discrete step function. The simulation here is based on 50.000 replications and moments and quantiles are reported in Table 1. To calculate tail probabilities for the test statistics below the asymptotic distribution is also approximated by a $\Gamma-$distribution with the simulated mean and variance, see Doornik (1998), which closely reproduces the simulated quantiles, see Table 1.

Table 2 reports the LR statistics for the cointegration ranks for the baseline model, together with the asymptotic tail probabilities derived from the $\Gamma-$approximation. The hypotheses $H^D(r, s)$ are tested sequentially against $H^D(p)$ based on the partial nesting structure. All models with $r = 0$ and $r = 1$ are safely rejected. In the row for $r = 2$ the reductions to the models $H^D(2, 1)$ and $H^D(2, 2)$ have tail probabilities around 10%, and we note that in the augmented model with 9 additional impulse dummies the tail probabilities for the LR statistics for the two candiate models are 8% and 14%, respectively. The two potentially preferred models are nested, $H^D(2, 1) \subset H^D(2, 2)$, and can be compared

r	$-2\log Q\left(H^D(r,s)\,\middle|\,H^D(p)\right)$					
0	442.8 [.00]	342.4 [.00]	260.8 [.00]	202.4 [.00]	167.7 [.00]	154.6 [.00]
1		248.5 [.00]	180.1 [.00]	129.3 [.01]	104.0 [.02]	94.3 [.01]
2			124.8 [.00]	85.4 [.10]	67.2 [.11]	58.8 [.05]
3				54.3 [.28]	42.5 [.18]	31.5 [.17]
4					22.1 [.27]	12.4 [.32]
$p-r-s$	5	4	3	2	1	0

Table 2: *Likelihood ratio tests for the cointegration ranks, (r,s). The numbers in brackets are tail probabilities derived from the $\Gamma-$approximation of the simulated distribution in Table 1.*

directly using a LR test. It follows directly from the result in Corollary 2 that the likelihood ratio statistic for $H^D(2,1) \mid H^D(2,2)$, calculated from the estimated covariances as

$$Q^{LR}_{(2,1)|(2,2)} = -T\log\left|\hat{\Omega}_{(2,1)}\hat{\Omega}^{-1}_{(2,2)}\right|,$$

has the limiting distribution of the maximum eigenvalue of $S\left(W_2, G^{(r,s)}, W_2\right)$, see Nielsen (2007), which is easily simulated. For the baseline model the statistic is 18.2, corresponding to a tail probability of 7%, while the augmented model produces a test of 20.5 and a tail probability of 3%, showing that the reduction to the model $H^D(2,1)$ is marginal. Furthermore, the model $H^D(2,2)$ with $p-r-s=1$ I(2) trend is most easily reconsiled with economic theory and together with the statistical evidence we take this model as the preferred in the following, noting, however, that it could be interesting also to consider the economic implications of second I(2) trend in the data.

Note that there are strong indications of an I(2) trend in the data, even after allowing for a deterministic shift in the trend. For the baseline model with the hypothesis $H^D(2,2)$ imposed, the characteristic polynomial has four unit roots and the inverses of the remaining 11 roots are given by,

$$0.67 \pm 0.21 \cdot i;\ -0.42 \pm 0.05 \cdot i;\ -0.11 \pm 0.38 \cdot i;\ 0.28 \pm 0.26 \cdot i;\ 0.33 \pm 0.08 \cdot i;\ -0.22,$$

which all have absolute values smaller than one, and hence there are no indications of additional unit roots.

Testing Homogeneity. Based on the preferred model $H^D(2,2)$, we first investigate if the change in the linear trend implied by D_{1t} is needed, or equivalently, we test the restriction of a common deterministic trend coefficient in all cointegrating relationships, $\tau' X_t$, in the two sub-samples. We formulate this as,

$$\mathcal{H}_0 : \tau^* = \begin{pmatrix} I_6 \\ 0_{(1\times 6)} \end{pmatrix}\varphi,$$

with φ unrestricted, imposing a zero row in τ^*. The LR statistic for $\mathcal{H}_0 \mid H^D(2,2)$ is given by 17.8 corresponding to a zero tail probability in the asymptotic $\chi^2(4)$ distribution. This emphasizes the relevance of the changing trend. In the augmented model with 9 additional impulse dummies the corresponding statistic is 29.6, confirming this.

An important hypothesis is that the common I(2) trend loads into the nominal variables with equal coefficients, so that the real variables, $c_t - p_t$, $y_t - p_t$, and $w_t - p_t$, are first order non-stationary, I(1). Economically, this hypothesis implies that money illusion is excluded in the long-run, and the hypothesis would allow a nominal-to-real transformation from the I(2) vector X_t to a vector of I(1) variables, e.g.

$$Y_t = (c_t - p_t, y_t - p_t, w_t - p_t, \Delta p_t, R_t)',$$

see Kongsted (2005). Given homogeneity, the subsequent I(1) cointegration analysis of Y_t can be conducted without loss of information and the polynomially cointegrating relationships are embedded as usual cointegrating relationships in the I(1) cointegration model, see Kongsted and Nielsen (2004). Often this hypothesis is imposed *a priori*, see e.g. the analyses of real consumption variables in Lettau and Ludvigson (2001) and Palumbo *et al.* (2006), but here we want to explicitly test the hypothesis of homogeneity.

Equal loadings to the single stochastic I(2) trend corresponds to $\beta_{\perp 2}$ being proportional to $B = (1,1,1,1,0)'$, see (8). From the baseline case and from the augmented model , the estimated counterparts are given by

$$\begin{aligned} \hat{\beta}_{\perp 2}^{\text{baseline}} &= (1,\ 0.918,\ 1.436,\ 1.163,\ 0.018)' \\ \hat{\beta}_{\perp 2}^{\text{augmented}} &= (1,\ 0.914,\ 1.263,\ 1.130,\ 0.029)', \end{aligned}$$

respectively. The results suggest that the unrestricted estimates under $H^D(2,2)$ are quite close to homogeneous, except slightly larger coefficients to w and p. Noting that $\beta_{\perp 2} = \tau_\perp$, the homogeneity restriction can be formally tested as

$$\mathcal{H}_1 : \tau^* = \begin{pmatrix} B_\perp & 0_{(5\times 2)} \\ 0_{(2\times 4)} & I_2 \end{pmatrix} \varphi,$$

where $B_\perp$ is a 5×4 matrix and φ is 7×6 with unrestricted parameters. The LR statistics for $\mathcal{H}_1$ are 5.3 and 3.4 in the two specifications, corresponding to tail probabilities of 0.26 and 0.49 in the asymptotic $\chi^2(4)$ distribution.

For comparison, the homogeneity restriction in $\mathcal{H}_1$ has also been tested in the model with no change in the linear trend (which is therefore misspecified). Without allowing for the changing trend slopes, the LR statistic is 20.4, and this would lead to a firm rejection of homogeneity of the stochastic trends, and, hence, a rejection of the economically relevant nominal-to-real transformation. Thus, the baseline model is well-specified and leads to sound economic interpretations of the dynamics, while the misspecified model where a change in the trend is not allowed for, leads to the reverse.

Deterministic Terms. To illustrate the role of the deterministic components, i.e. the effects of the trend with a changing slope in 1981 : 2, the $k = 3$ impulse dummies induced by the changing trend slope, and the nine additional innovation dummies included to account for outliers, Figure 2 shows the stable combinations, $\hat{s}_t = \hat{\beta}' X_t + \hat{\psi}' \Delta X_t$, $\hat{\beta}'_{\perp 1} \Delta X_t$, and $\hat{\beta}'_{\perp 2} \Delta^2 X_t$, together with their deterministic components. The deterministic components of the data are calculated as the terms in (8) involving the deterministic variables, D_t, d_t, δ_t, and the initial values, X_0, ΔX_0, and $\Delta^2 X_0$:

$$C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} \Psi_\delta \delta_i + C_1 \sum_{i=1}^{t} \Psi_\delta \delta_i + \gamma_D D_t + \gamma_d d_t + C_0 (L) \Psi_\delta \delta_i, \tag{19}$$

where γ_D and γ_d contain also the effects of the initial values of the process. We recall that the innovation dummies (δ_t) enter the dynamics in the same way as the innovations (ϵ_t), and accumulate (once and twice) to produce level shifts and changing trend slopes in the data.

Graph (A) and (B) show the $r = 2$ multi-cointegrating relationships, $\hat{s}_t = \hat{\beta}' X_t + \hat{\psi}' \Delta X_t$. Here we have normalized so that $\hat{s}_{1t}$ has unit coefficient to c_t and excludes the interest rate, b_t, while $\hat{s}_{2t}$ is normalized on b_t and excludes consumption, c_t. Regarding the deterministic components, we first note the marked linear trends in equilibrium. The break in 1981 : 2 allows for a shift in the equilibrium level and in the slope of the linear trend. For the chosen normalization, the consumption relation, $\hat{s}_{1t}$, has approximately a constant trend slope, suggesting *co-breaking* between the trend breaks of individual variables. The changing trend slope is clearly important for the interest rate relation, $\hat{s}_{2t}$, however. Regarding the impulse dummies, we note that the k induced dummies play the role of conditioning on observations for 1981 : 2, 1981 : 3, and 1981 : 4, and the effect is comparable to the initial values, 1964 : 2, 1964 : 3, and 1964 : 4. In addition, Figure 2 highlights the observations modelled by innovation impulse dummies. From the accumulation in (19), with $\beta' C_2 = 0$ and $\beta' C_1 \neq 0$, the impulse dummies give at most level shifts in $\hat{\beta}' X_t$, but the accumulated effects cancel in the multi-cointegrating relations producing only exponential decreasing effects, λ_t.

The I(1) directions of the data, $\hat{\beta}'_{\perp 1} X_t$, also contain trends with a changing slope (and a level shift) in 1981 : 2, and the first differences, $\hat{\beta}'_{\perp 1} \Delta X_t$, are reported graph (C) and (D). We note that the changing trend in $\hat{\beta}'_{\perp 1} X_t$ gives a change in the growth rates in the graphs in 1981 : 2. From (19), with $\beta'_{\perp 1} C_2 = 0$ and $\beta'_{\perp 1} C_1 \neq 0$, the innovation dummies produce level shifts in $\hat{\beta}'_{\perp 1} X_t$, but they are eliminated in the graph by first differencing. Note that the first differencing produce a slightly more complex behavior of λ_t, which, by the way, is the same as the dynamic effect of hte normal innovations, ϵ_t.

Finally, the I(2) direction, $\hat{\beta}'_{\perp 2} X_t$, contains a linear trend and the changing trend slope in 1981 : 2. Furthermore, since $\beta'_{\perp 2} C_2 \neq 0$, the innovation dummies produce changing slopes at nine additional points in time. Graph (E) shows the stationary transformation, $\hat{\beta}'_{\perp 2} \Delta^2 X_t$. This has mean zero (from the double difference of the linear trends) apart from the exponential effects of innovational dummies, λ_t.

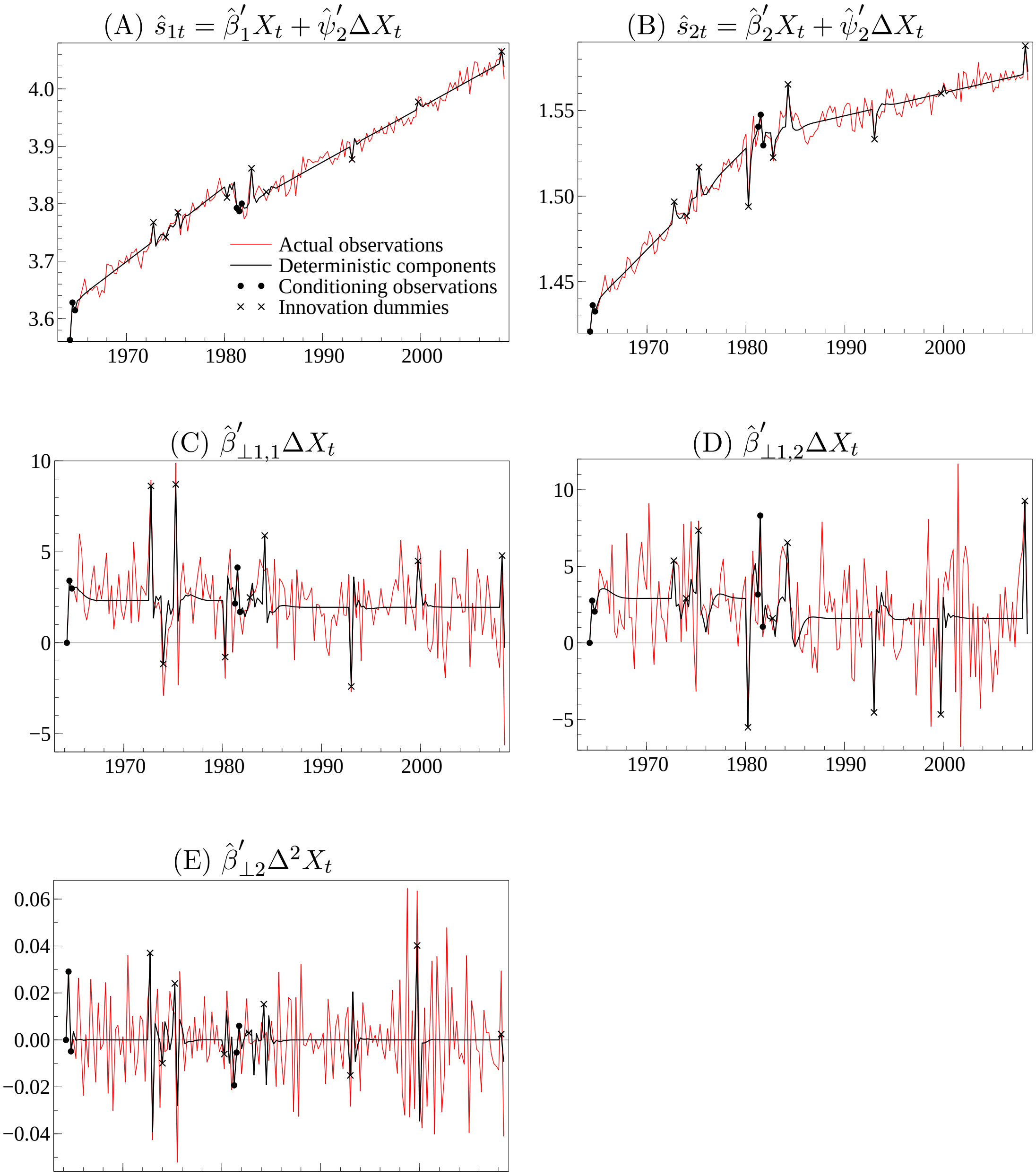

Figure 2: *Stationary linear combinations of the data based on the estimated augmented model, i.e.* $\hat{s}_t = \hat{\beta}' X_t + \hat{\psi}' \Delta X_t$, $\hat{\beta}_{\perp 1}' \Delta X_t$, *and* $\hat{\beta}_{\perp 2}' \Delta^2 X_t$*, and their deterministic components. The deterministic parts are the terms in (8) depending on* D_t, d_t, δ_t*, and the initial values, see (19).*

For empirical applications a choice must be made between allowing an innovation dummy, producing changing trend slopes in the data that co-break by assumption, or allowing also changing trend in the equilibrium relationships. Economically, this amounts to choosing between large shocks that follow the usual dynamics of the normal innovations versus genuine regime shifts. In the application above this choice was based on *a priori* reasoning and the graphical appearance of the data.

Software Implementation. The empirical analysis above was carried out in Ox, see Doornik (2002). Ox code for the I(2) rank test and for simulating the asymptotic distribution in the case of changing trend slopes can be obtained from the authors. The cointegrated I(2) model and the likelihood ratio test for the cointegration ranks are also implemented in the software CATS in RATS, see Dennis (2006).

A Asymptotics

A.1 Proof of Theorem 1

The I(2) model in (7) is a regression model with nonlinear parameters. To analyze this, it is as in Johansen (1997: Theorem A1) useful to initially analyze a linear regression model with regressors as in (7). With V_t p-dimensional, write the linear regression model as,

$$V_t = \theta^0 Z_{0t}^{\lambda} + \theta^1 Z_{1t}^{\lambda} + \theta^2 Z_{2t}^{\lambda} + \theta^D Z_{Dt} + \theta^d Z_{dt} + \theta^{\delta}\delta_t + \epsilon_{vt}(\theta), \tag{20}$$

where for $t = 1, 2, ..., T$, $\epsilon_{vt}(\theta)$ is $N_p(0, \Omega)$ distributed, conditional on the regressors Z_{it}^{λ}, and past V_t and Z_{it}^{λ}. The p_{z_i}-dimensional regressors Z_{it}^{λ} are – apart from an asymptotically vanishing term λ_t defined in (9) in terms of impulse dummies δ_t – mean-zero I(i) processes for $i = 0, 1, 2$. Specifically, with the p_η-dimensional η_t independent of ϵ_{vt} and i.i.d.$(0, \Sigma_\eta)$ distributed, $Z_{it}^{\lambda} = Z_{it} + \lambda_t$, where $\Delta^i Z_{it} = C^i(L)\eta_t = \sum_{j=0}^{\infty} C_j^i \eta_{t-i}$ and the coefficients C_j^i exponentially decreasing. Furthermore, $Z_{Dt} = D_t$ which is p_D-dimensional and which satisfies Assumption 2, and $Z_{dt} = d_t$, with $d_t = \Delta D_t$. Finally δ_t is a p_δ-dimensional impulse dummy regressor with entries $\delta_{it} = 1(t = T_i)$, $1 < T_i < T$, and $T_i = [Tu_i]$ with $u_i \in]0, 1[$.

Lemma 1 *Set $\theta^Z = (\theta^0, \theta^1, \theta^2, \theta^D, \theta^d)$, and $\theta = (\theta^Z, \theta^\delta) \in \Theta \subset \mathbb{R}^n$, where Θ is closed and $\Omega > 0$ varies freely. Then for the MLE $\hat{\theta}$ it holds that*

$$N_T^{-1}\left(\hat{\theta}^Z - \theta_0^Z\right) \xrightarrow{P} 0, \tag{21}$$

as $T \to \infty$ and with $N_T =$blockdiag$(I_{p_{z_0}}, T^{-1/2}I_{p_{z_1}}, T^{-3/2}I_{p_{z_2}}, T^{-1}I_{p_D}, I_{p_D})$. Furthermore, $\left(\hat{\theta}^\delta - \theta_0^\delta\right) = O_P(1)$.

Proof: Define $Z_t^{\lambda} = \left(Z_{0t}^{\lambda\prime}, Z_{1t}^{\lambda\prime}, Z_{2t}^{\lambda\prime}, Z_{Dt}', Z_{dt}'\right)'$, $Z_t = (Z_{0t}', Z_{1t}', Z_{2t}', Z_{Dt}', Z_{dt}')$ and set $\epsilon_{vt} = \epsilon_{vt}(\theta_0)$. Moreover, use the notation that for any p_x and p_y dimensional time series X_t and Y_t respectively,

$$M_{yx} = \tfrac{1}{T}\sum_{t=1}^{T} Y_t X_t' \tag{22}$$

Next, note that

$$\begin{aligned}\left(\hat{\theta}^Z - \theta_0^Z\right) &= M_{\epsilon z^\lambda . \delta} M_{z^\lambda z^\lambda . \delta}^{-1} \\ &= \left(M_{\epsilon z^\lambda} - M_{\epsilon\delta}M_{\delta\delta}^{-1}M_{\delta z^\lambda}\right)\left(M_{z^\lambda z^\lambda} - M_{z^\lambda\delta}M_{\delta\delta}^{-1}M_{\delta z^\lambda}\right)^{-1}\end{aligned}$$

By definition of the p_δ-dimensional impulse dummy δ_t, and the generic λ_t defined in (9), standard limit arguments immediately give, $N_T M_{z^\lambda z^\lambda . \delta} N_T = N_T M_{zz} N_T + o_P(1)$. That is, the OLS correction for δ_t is asymptotically negliable, and moreover, $Z_{it}^{\lambda} = Z_{it} + \lambda_t$ behaves asymptotically as Z_{it} for $i = 0, 1$ and 2. Hence,

$$N_T M_{z^\lambda z^\lambda . \delta} N_T = N_T M_{zz} N_T + o_P(1) \xrightarrow{D} \begin{pmatrix} \Sigma_{00} & 0 \\ 0 & \int_0^1 F_u F_u' du \end{pmatrix}, \tag{23}$$

where $\Sigma_{00} = V(Z_{0t}) = \sum_{i=0}^{\infty} C_i^0 \Sigma_{\eta 0} C_i^{0\prime}$ and $F_u = (F_{1u}', F_{2u}', D_u', d_u')'$. Here $F_{1u} = C^1(1) W_u^\eta$ and $F_{2u} = C^2(1) \int_0^s W_s^\eta ds$, with W_u^η a Brownian motion with variance $\Sigma_{\eta 0}$. Similarly,

$$T^{1/2} N_T M_{z^\lambda \epsilon \cdot \delta} = T^{1/2} N_T M_{z\epsilon} + o_P(1) \xrightarrow{D} \left(N_{p_{z_0} \times p}(0, \Sigma_{00} \otimes \Omega_0), \int_0^1 F dW^{\epsilon\prime} \right), \tag{24}$$

where W_u^ϵ is a p-dimensional Brownian motion with variance Ω_0. Collecting terms (21) holds. Note that it is essential for the results that the asymptotically stationary $Z_{0t}^\lambda = Z_{0t} + \lambda_t$ regressor has mean zero apart from the generic λ_t defined in (9) which is asymptotically vanishing. If not, e.g. the blockdiagonality in (23), which corresponds to the limiting information, would not apply.

Finally, with each entry δ_{it} in the p_δ-dimensional δ_t of the form $\delta_{it} = 1(t = T_i)$, it follows that $\hat{\theta}^\delta = M_{\epsilon(\hat{\theta})\delta} M_{\delta\delta}^{-1} = \left(M_{v\delta} - \hat{\theta}^Z M_{z^\lambda \delta} \right) M_{\delta\delta}^{-1}$, or

$$\left(\hat{\theta}^\delta - \theta_0^\delta \right) = \left(\epsilon_{vT_1} - \left(\hat{\theta}^Z - \theta_0^Z \right) Z_{T_1}, ..., \epsilon_{vT_{p_\delta}} - \left(\hat{\theta}^Z - \theta_0^Z \right) Z_{T_{p_\delta}} \right) = \left(\epsilon_{vT_1}, ..., \epsilon_{vT_{p_\delta}} \right) + o_P(1),$$

from which $\hat{\theta}^\delta = O_P(1)$ and inconsistency holds. □

Proof of Theorem 1: Rewrite the I(2) model $H^D(r,s)$ in (7) as in (20),

$$\Delta^2 X_t = \theta^0 Z_{0t}^\lambda + \theta^1 Z_{1t}^\lambda + \theta^2 Z_{2t}^\lambda + \theta^D Z_{Dt} + \theta^d Z_{dt} + \theta^\delta \delta_t + \epsilon_t(\theta), \tag{25}$$

where $Z_{Dt} = D_{t-1}$, $Z_{dt} = d_{t-1}$, $Z_{2t}^\lambda = \beta_{\perp 20}' X_{t-1}$, $\theta^\delta \delta_t = \Psi_\delta \delta_t$ and

$$\begin{aligned} Z_{0t}^\lambda &= \begin{pmatrix} \beta_0' X_{t-1} + \psi_0' \Delta X_{t-1} + \rho_0' \tau_{D0}' D_{t-1} + \psi_{D0}' d_{t-1} \\ \tau_0' \Delta X_{t-1} + \tau_{D0}' d_{t-1} \\ \Delta^2 \mathbb{X}_{t-1} \end{pmatrix} \\ Z_{1t}^\lambda &= \begin{pmatrix} \beta_{\perp 20}' \Delta X_{t-1} \\ \beta_{\perp 10}' X_{t-1} + \beta_{\perp 10}' \bar{\tau}_0 \tau_{D0}' D_{t-1} \end{pmatrix}. \end{aligned} \tag{26}$$

Recall that $\rho = \bar{\tau}_0' \beta$, such that $\rho = \rho_0 + \bar{\tau}_0' \beta_{\perp 10} B_1 = \rho(B_1)$, implying $\rho_\perp = \rho_\perp(B_1)$ as well. Using the definitions in (10), the parameters θ^0, θ^1 and θ^2 are given by (27),

$$\begin{aligned} \theta^0 &= \left(\alpha, \alpha(\psi - \psi_0)' \bar{\tau}_0 + \alpha_{\perp\Omega} \kappa', \Psi \right) \\ \theta^1 &= \left(\alpha B_0' + \alpha_{\perp\Omega} \kappa' \left[\bar{\rho}_\perp(B_1) C_0' + \bar{\rho}(B_1) B_2' \right], \alpha B_1' \right) \\ \theta^2 &= \alpha B_2', \end{aligned} \tag{27}$$

while the parameters for the deterministic regressors are given by, $\theta^\delta = \Psi_\delta$,

$$\theta^D = \alpha B_D' \quad \text{and} \quad \theta^d = \left(\alpha B_d' + \alpha_{\perp\Omega} \kappa' \left[\bar{\rho}_\perp(B_1) C_D' + \bar{\rho}(B_1) B_D' \right] \right). \tag{28}$$

Applying our Lemma 1, the proof is identical to the proof of Theorem 2 in Johansen (1997), apart from the θ^D and θ^d parameters in (28), and hence B_D, B_d and C_D in (10). As $\hat{\theta}^D = \hat{\alpha} \hat{B}_D'$, with $\hat{\alpha}$ consistent, we can conclude by Lemma 1 that $T \hat{B}_D \xrightarrow{P} 0$. Next, with θ^d defined in (28), multiply by $\hat{\alpha}_\perp$ to see that $\hat{C}_D \xrightarrow{P} 0$, as $\hat{\rho}_\perp, \hat{\rho}$, $\hat{\kappa}$, $\hat{\Omega}$, $\hat{\alpha}_\perp$ and $\hat{B}_1$ and $\hat{B}_D$ are consistent. Likewise, multiplying by $\hat{\alpha}' \hat{\Omega}^{-1}$, gives $\hat{B}_d \xrightarrow{P} 0$. □

A.2 Proof of Theorem 2

The proof proceeds basically as in the proof of Lemma 1 in Johansen (1997), apart from the additional deterministic terms here. Thus, in terms of the parametrization in (25) note initially that the parameters α, $\theta_2^0, \Psi, \Psi_\delta, \Omega, B_0, B_1, B_2, B_D, B_d, C_0$ and C_D are all freely varying, where

$$\theta_2^0 = \alpha \left(\psi - \psi_0\right)' \bar{\tau}_0 + \alpha_{\perp\Omega}\kappa'. \tag{29}$$

Clearly $\alpha, \Psi, \Psi_\delta$ and Ω are trivial to obtain from these, as noted above $\rho = \rho\left(B_1\right)$, while $\beta = \beta\left(B_1, B_2\right)$, $\psi = \psi\left(\theta_2^0, B_0, \alpha, \Omega\right)$, $\kappa = \kappa\left(\theta_2^0, \alpha, \Omega\right)$, and $\tau = \tau\left(C_0, B_2\right)$, see also Johansen (1997: equation (48)). For the remaining new parameters τ_D and ψ_d note first that τ_D can be found from $\rho = \rho\left(B_1\right), B_D$ and C_D as

$$\left(\tau_D - \tau_{D0}\right)' = \bar{\rho}\left(B_1\right) B_D' + \bar{\rho}_\perp\left(B_1\right) C_D'. \tag{30}$$

Next,

$$\left(\psi_d - \psi_{d0}\right)' = B_d' + \left(\psi - \psi_0\right)' \bar{\tau}_0 \tau_{D0}'. \tag{31}$$

With $\theta^Z = \left(\theta^0, \theta^1, \theta^2, \theta^D, \theta^d\right)$, $\theta^\delta = \Psi_\delta$ and $\theta = \left(\theta^Z, \theta^\delta\right)$ the log-likelihood function is given by,

$$L_T\left(\theta, \Omega\right) = -\tfrac{1}{2}\left[T \log\left|\Omega\right| + tr\{\Omega^{-1}\sum_{t=1}^{T} \epsilon_t\left(\theta\right)\epsilon_t\left(\theta\right)'\}\right], \tag{32}$$

where with $\epsilon_t = \epsilon_t\left(\theta_0\right)$,

$$\begin{aligned}
\epsilon_t\left(\theta\right) &= \Delta^2 X_t - \theta^0 Z_{0t}^\lambda - \theta^1 Z_{1t}^\lambda - \theta^2 Z_{2t}^\lambda - \theta^D Z_{Dt} - \theta^d Z_{dt} - \theta^\delta \delta_t \\
&= \epsilon_t - \left(\theta^0 - \theta_0^0\right) Z_{0t}^\lambda - \theta^1 Z_{1t}^\lambda - \theta^2 Z_{2t}^\lambda \\
&\quad - \left(\theta^D - \theta_0^D\right) Z_{Dt} - (\theta^d - \theta_0^d) Z_{dt} - (\theta^\delta - \theta_0^\delta)\delta_t.
\end{aligned}$$

The limiting distribution of $\hat{\theta}^Z$ is found by considering an asymptotic expansions of the score evaluated at $\hat{\theta}$. Introduce therefore the notation $dL_T(\hat{A}; dA) = dL_T\left(\theta, \Omega; dA\right)|_{\theta=\hat{\theta}}$ for the differential[2] of the log-likelihood function in (32) in the direction dA, where A is a matrix (or vector) valued parameter in θ, and the differential is evaluated at $\theta = \hat{\theta}$.

Set $B' = \left(B_0', B_1', B_2', B_D', B_d'\right)$, $B_T' = \left(TB_0', TB_1', T^2 B_2', T^{3/2} B_D', T^{1/2} B_d'\right)$ and define accordingly $Z_{Bt} = \left(Z_{1t}^{\lambda\prime}, Z_{2t}^{\lambda\prime}, Z_{Dt}', Z_{dt}'\right)'$. Moreover, corresponding to the order of magnitudes of the processes in Z_{Bt}, set N_T^B =blockdiag$\left(T^{-1/2} I_{p-r}, T^{-3/2} I_{p-r-s}, T^{-1} I_{q+1}, I_{q+1}\right)$. Then by definition $dL_T(\hat{B}, dB) = 0$, and with $\hat{B}_T$ inserted for $\hat{B}$, one finds that

$$tr\left\{\left[\alpha_0'\Omega_0^{-1}\left(T^{1/2} M_{\epsilon z_B} N_T^B\right) - \alpha_0'\Omega_0^{-1}\alpha_0 \hat{B}_T' N_T^B M_{z_B z_B} N_T^B\right] dB\right\} = o_P\left(1\right). \tag{33}$$

This is the equivalent of Johansen (1997: equation (55)), and holds as there by applying limiting arguments in terms of Z_{0t}^λ, Z_{1t}^λ and Z_{2t}^λ which, apart from asymptotically vanishing λ_t terms, are I(0), I(1) and I(2) respectively – see the proof of Lemma 1. A further

[2] See Magnus and Neudecker (1999) for the theory of matrix differential calculus.

difference is the inclusion of the $\left(\theta^{\delta}-\theta_{0}^{\delta}\right)\delta_{t}$ term in the residual $\epsilon_{t}(\theta)$. In (33), we have in particular used that,

$$\alpha_{0}^{\prime}\Omega_{0}^{-1}\left(\hat{\theta}^{\delta}-\theta_{0}^{\delta}\right)T^{1/2}M_{\delta z_{B}}N_{T}^{B}=o_{P}(1),$$

which holds since $T^{1/2}M_{\delta z_{B}}N_{T}^{B}=o_{P}(1)$ as δ_{t} contain alone impulse dummies, and as $(\hat{\theta}^{\delta}-\theta_{0}^{\delta})=O_{P}(1)$, see Theorem 1.

Similar to Johansen (1997) one may also note that differentials of θ^{1} and θ^{d} in the direction dB_{1} do not matter asymptotically as they are multiplied by either of C_{0},B_{2},C_{D} or B_{D} and hence by Theorem 1 converge in probability to zero. Likewise the differentials of θ^{1} in the direction dB_{2} and of θ^{d} in the direction dB_{D} do not matter asymptotically. Moreover, the definitions in (27) and (28) have been used, in addition to the consistency results of Theorem 1 to see that,

$$\begin{aligned}
\hat{\alpha}^{\prime}\hat{\Omega}^{-1}T\hat{\theta}^{1}&=\alpha_{0}^{\prime}\Omega_{0}^{-1}\alpha_{0}\left(T\hat{B}_{0}^{\prime},T\hat{B}_{1}^{\prime}\right)+o_{P}(1)\\
\hat{\alpha}^{\prime}\hat{\Omega}^{-1}T^{2}\hat{\theta}^{2}&=\alpha_{0}^{\prime}\Omega_{0}^{-1}\alpha_{0}\left(T^{2}\hat{B}_{2}^{\prime}\right)+o_{P}(1)\\
\hat{\alpha}^{\prime}\hat{\Omega}^{-1}T^{3/2}\hat{\theta}^{D}&=\alpha_{0}^{\prime}\Omega_{0}^{-1}\alpha_{0}\left(T^{3/2}\hat{B}_{D}^{\prime}\right)+o_{P}(1)\\
\hat{\alpha}^{\prime}\hat{\Omega}^{-1}T^{1/2}\hat{\theta}^{d}&=\alpha_{0}^{\prime}\Omega_{0}^{-1}\alpha_{0}(T^{1/2}\hat{B}_{d}^{\prime})+o_{P}(1).
\end{aligned}$$

Next, by (33) and (23), then in the limit as $T\to\infty$, with B^{∞} denoting the limiting distribution of $\hat{B}_{T}$,

$$\alpha_{0}^{\prime}\Omega_{0}^{-1}\left(\textstyle\int_{0}^{1}dVH^{*\prime}\right)=\alpha_{0}^{\prime}\Omega_{0}^{-1}\alpha_{0}B^{\infty\prime}\textstyle\int_{0}^{1}H_{u}^{*}H_{u}^{*\prime}du, \tag{34}$$

from which the first result in Theorem 2 follows. Note that $H_{u}^{*}=(H_{u}^{\prime},D_{u}^{\prime},d_{u}^{\prime})^{\prime}$ is defined in Theorem 2 in terms of H_{u} in (13), limit of the deterministic terms and the p-dimensional Brownian motion V_{u} with covariance Ω_{0}.

For the asymptotics of $\hat{C}_{0}$ and $\hat{C}_{D}$, set similar to above $C^{\prime}=(C_{0}^{\prime},C_{D}^{\prime})$, $C_{T}^{\prime}=\left(TC_{0}^{\prime},T^{1/2}C_{D}^{\prime}\right)$ and define $Z_{Ct}=(Z_{1t}^{\prime}(I_{p-r-s},0),Z_{dt})^{\prime}$, that is $Z_{Ct}=\left(\Delta X_{t-1}^{\prime}\beta_{20},Z_{dt}\right)^{\prime}$. Moreover, set N_{T}^{C} =blockdiag$\left(T^{-1/2}I_{p-r-s},I_{q+1}\right)$ corresponding to the order of magnitude of Z_{Ct}. By definition $dL_{T}(\hat{C},dC)=0$, and with $\hat{C}_{T}$ inserted for $\hat{C}$, and similar to (33),

$$tr\left\{\left[\phi_{0}^{\prime}\Omega_{0}^{-1}\left(T^{1/2}M_{\epsilon z_{C}}N_{T}^{C}\right)-\phi_{0}^{\prime}\Omega_{0}^{-1}\phi_{0}\hat{C}_{T}^{\prime}N_{T}^{C}M_{z_{C}z_{C}}N_{T}^{C}\right]dC\right\}=o_{P}(1), \tag{35}$$

where $\phi=\alpha_{\perp\Omega}\kappa^{\prime}\bar{\rho}_{\perp}=\Omega\alpha_{\perp}\left(\alpha_{\perp}^{\prime}\Omega\alpha_{\perp}\right)^{-1}\kappa^{\prime}\bar{\rho}_{\perp}$, cf. (15). This is the equivalent of Johansen (1997: p.461) and holds as above by standard limiting arguments, the fact that δ_{t} is asymptotically negliable, and the definitions in (27) and (28), in addition to the consistency results of Theorem 1. In particular, it has been used that $\phi^{\prime}\Omega^{-1}\alpha=0$ such that,

$$\begin{aligned}
\hat{\phi}^{\prime}\hat{\Omega}^{-1}T\hat{\theta}^{1}&=\phi_{0}^{\prime}\Omega_{0}^{-1}\phi_{0}\left(T\hat{C}_{0}^{\prime}\right)(I_{p-r-s},0_{p-r-s\times s})+o_{P}(1)\\
\hat{\alpha}^{\prime}\hat{\Omega}^{-1}T^{1/2}\hat{\theta}^{d}&=\phi_{0}^{\prime}\Omega_{0}^{-1}\phi_{0}(T^{1/2}\hat{C}_{D}^{\prime})+o_{P}(1).
\end{aligned}$$

Next, by (35) and (23), then in the limit as $T \to \infty$, with C^∞ denoting the limiting distribution of $\hat{C}_T$,

$$\phi_0'\Omega_0^{-1}\left(\int_0^1 dV H_0^{*\prime}\right) = \phi_0'\Omega_0^{-1}\phi_0 C^{\infty\prime}\int_0^1 H_{0u}^* H_{0u}^{*\prime} du, \tag{36}$$

from which the second result in Theorem 2 follows using the definition of ϕ. Note that $H_{0u}^* = (H_{0u}', H_{du}')'$ is defined in Theorem 2 in terms of H_{0u} in (13).

As in Johansen (1997: p.461-462) the asymptotic distribution of $\hat{\rho}$ follows from the identity,

$$\hat{\rho} = \bar{\tau}_0'\hat{\beta} = \rho_0 + \bar{\tau}_0'\beta_{\perp 10}\hat{B}_1, \tag{37}$$

while $T^{1/2}\left(\hat{\theta}^0 - \theta_0^0\right) \xrightarrow{D} N_{p\times(2r+s+p)}\left(0, \Omega_0 \otimes \Sigma_{00}^{-1}\right)$,where $\Sigma_{00} = V(Z_{0t})$. □

A.3 Proof of Corollary 1

The proof consists of two parts. In the first we apply Theorem 2 to find the asymptotic distribution of $(\hat{\tau}^* - \tau_0^*)$, which is then used in the second part where a Taylor expansion of $(\hat{\tau}_a^* - \tau_0^*)$ is applied.

Part 1: Theorem 2 implies that the asymptotic distribution of $(\hat{\tau}^* - \tau_0^*)$ is given by,

$$\begin{pmatrix} T\bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0) \\ \sqrt{T}(\hat{\tau}_D - \tau_{D0}) \end{pmatrix} = \begin{pmatrix} T\hat{C}_0 \\ \sqrt{T}\hat{C}_D \end{pmatrix} \bar{\rho}_{\perp 0}' + o_P(1) \xrightarrow{D} C^\infty \bar{\rho}_{\perp 0}' = R(H_0^*, V_2)\bar{\rho}_{\perp 0}'. \tag{38}$$

To see this note that, $\bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0) = \bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0)(\rho_0\bar{\rho}_0' + \rho_{\perp 0}\bar{\rho}_{\perp 0}')$. Using the identities Johansen (1997: p.461 and p.459), together with the derived consistency of $\hat{\rho}$ in Theorem 1 here, one finds[3]

$$\bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0)\rho_0 = \hat{B}_2 - \hat{C}_0\hat{B}_1 + o_P\left(T^{-2}\right) \quad \text{and} \quad \bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0)\rho_{\perp 0} = \hat{C}_0 + o_P\left(T^{-1}\right).$$

Likewise,

$$(\hat{\tau}_D - \tau_{D0})\rho_0 = \hat{B}_D - \hat{C}_D\hat{B}_1 + o_P\left(T^{-3/2}\right) \quad \text{and} \quad (\hat{\tau}_D - \tau_{D0})\rho_{\perp 0} = \hat{C}_D + o_P\left(T^{-1/2}\right),$$

and collecting terms (38) holds.

Part 2: As in the proof of Theorem 4.2 in Rahbek *et al.* (1999) and Lemma 3 in Johansen *et al.* (2009), use the expansion around $\tau_{a0}^* = \tau_0^*(a'\tau_0)^{-1} = \tau_0^*$,

$$\begin{aligned} \hat{\tau}_a^* - \tau_0^* &= \left(I_{p+1+q} - \tau_0^*(a', 0)\right)(\hat{\tau}^* - \tau_0^*) + O_P\left(|\hat{\tau}^* - \tau_0^*|^2\right) \\ &= (a', 0)_\perp'\left(\tau_{\perp 0}^{*\prime}(a', 0)_\perp'\right)^{-1}\tau_{\perp 0}^{*\prime}(\hat{\tau}^* - \tau_0^*) + O_P\left(|\hat{\tau}^* - \tau_0^*|^2\right). \end{aligned}$$

Observe that,

$$\tau_{\perp 0}^{*\prime}(\hat{\tau}^* - \tau_0^*) = \begin{pmatrix} \bar{\beta}_{\perp 20}'(\hat{\tau} - \tau_0) \\ \hat{\tau}_D - \tau_{D0} \end{pmatrix},$$

[3] Using in particular, $\hat{\rho}_\perp - \rho_{\perp 0} = -\rho_0(\hat{\rho}_0'\rho_0)^{-1}(\hat{\rho} - \rho_0)'\rho_{\perp 0}$, and $\bar{\rho}_{\perp 0}'\bar{\tau}_0\beta_{\perp 10} = I$,

and

$$\left(a',0\right)'_{\perp}\left(\tau^{*\prime}_{\perp 0}\left(a',0\right)'_{\perp}\right)^{-1}=\begin{pmatrix} a_{\perp}\left(\beta'_{\perp 20}a_{\perp}\right)^{-1} & 0 \\ \tau_{D0}\bar{\tau}'_{0}a_{\perp}\left(\beta'_{\perp 20}a_{\perp}\right)^{-1} & I_{q+1}\end{pmatrix}$$

and the result holds as claimed using (38). □

A.4 Proof of Corollary 2

The results follow by mimicking the proof of Theorem 2 in Nielsen and Rahbek (2007) (NR henceforth), using the results in Theorem 2. Specifically, replacing in NR indices 'l' (for linear) by 'D', and 'c' (for constant) by 'd' the arguments are completely identical except for the role of the impulse dummy δ_t as additional regressor.

NR falls in two parts: First, asymptotics for the test of an auxillary null H_{aux} against $H^D(r,s)$, and, next, asymptotics for auxillary null against $H^D(p)$.

On H_{aux} against $H^D(r,s)$: Replace the residuals $\hat{\epsilon}_t$, $\check{\epsilon}_t$ and $\hat{\epsilon}_{t0}$ in NR by,

$$\begin{aligned}\hat{\epsilon}_t &= \Delta^2 X_t - \hat{\theta}^0 Z^{\lambda}_{0t} - \hat{\theta}^1 Z^{\lambda}_{1t} - \hat{\theta}^2 Z^{\lambda}_{2t} - \hat{\theta}^D Z_{Dt} - \hat{\theta}^d Z_{dt} - \hat{\theta}^{\delta}\delta_t \\ \hat{\epsilon}_{t0} &= \Delta^2 X_t - \hat{\theta}^0 Z^{\lambda}_{0t} - \hat{\theta}^{\delta}\delta_t \\ \check{\epsilon}_t &= \Delta^2 X_t - \check{\theta}^0 Z^{\lambda}_{0t} - \check{\theta}^{\delta}\delta_t,\end{aligned}$$

where $\hat{\theta}$ denotes the estimator under $H^D(r,s)$, while $\check{\theta}$ denotes the estimator under the auxillary null, given by $\Delta^2 X_t = \theta^0 Z^{\lambda}_{0t} + \theta^{\delta}\delta_t + \epsilon_t(\theta)$. That is, ψ, ψ_d, τ, τ_D and β fixed at their true values. All arguments remain the same as in NR, except in the study of the covariance estimated under $H^D(r,s)$,

$$\hat{\Omega} = M_{\hat{\epsilon}\hat{\epsilon}} = M_{\hat{\epsilon}_0\hat{\epsilon}_0} + X_T - Y_T - Y'_T,$$

with X_T as in NR, while

$$Y_T = \left(M_{\epsilon Z_B} - \left(\hat{\theta}^0 - \theta^0_0\right)M_{Z_0 Z_B} - \left(\hat{\theta}^{\delta} - \theta^{\delta}_0\right)M_{\delta Z_B}\right)\hat{\theta}^{Z_B\prime}.$$

Here Z_B refers to $Z_{Bt} = \left(Z^{\lambda\prime}_{1t}, Z^{\lambda\prime}_{2t}, Z'_{Dt}, Z'_{dt}\right)'$ (see proof of Theorem 2, where also the corresponding normalization matrix N^B_T is defined, which in NR-2 corresponds to D^{-1}_T there) and $\hat{\theta}^{Z_B} = \left(\hat{\theta}^1, \hat{\theta}^2, \hat{\theta}^D, \hat{\theta}^d\right)$. For the extra term in Y_T, $\left(\hat{\theta}^{\delta} - \theta^{\delta}_0\right)M_{\delta Z_B}$, it is needed that

$$T\left(\hat{\theta}^{\delta} - \theta^{\delta}_0\right)M_{\delta Z_B}\hat{\theta}^{Z_B\prime} = o_P(1).$$

But this holds as *(i)* $\left(\hat{\theta}^{\delta} - \theta^{\delta}_0\right) = O_P(1)$ by Theorem 1, *(ii)* $\sqrt{T}N^B_T\hat{\theta}^{Z_B\prime} = O_P(1)$ from Theorem 2, and *(iii)* $\sqrt{T}M_{\delta Z_B}\left(N^B_T\right)^{-1} = o_P(1)$ by definition of δ_t.

On H_{aux} against $H^D(p)$: Similar to NR, the model $H^D(p)$ is given by

$$\Delta^2 X_t = \Pi X_{t-1} + \Pi_D D_{t-1} - \Gamma\Delta X_{t-1} - \Gamma_d d_{t-1} + \Psi\Delta^2\mathbb{X}_{t-1} + \Psi_{\delta}\delta_t + \epsilon_t(\theta),$$

and as shown in the proof of Lemma 1 above, the additional regressor δ_t plays no role asymptotically, and therefore the arguments in NR remain identical. □

B Data in Section 4

The data are from the Flow of Funds Accounts (FFA) by the Federal Board of Governors and the National Income and Product Account (NIPA) from the US Department of Commerce.

Consumption is measured as the personal expenditures of households and non-profit organizations on non-durable goods and services from NIPA. The price level is measured as the corresponding implicit deflator. Income is measured as the disposable income of households and non-profit organizations from NIPA, calculated as personal income minus current taxes. Wealth is taken from the FFA and is calculated as households tangible and financial assets minus liabilities. Finally, the bond rate is the Federal funds 10-year bond rate from the US department of Commerce.

References

Bacchiocchi, E. and L. Fanelli (2005): "Testing the Purchasing Power Parity Through I(2) Cointegration Techniques." *Journal of Applied Econometrics*, 20(6):749–770.

Banerjee, A., L. Cockerell, and B. Russell (2001): "An I(2) Analysis of Inflation and the Markup." *Journal of Applied Econometrics*, 16:221–240.

Boswijk, H. P. (2000): "Mixed Normality and Ancillarity in I(2) Systems." *Econometric Theory*, 16:878–904.

Dennis, J. (2006): *CATS in RATS - Cointegration Analysis of Time Series, Version 2*. Evanston: Estima.

Diamandis, P., D. Georgoutsos, and G. Kouretas (2000): "The Monetary Model in the Presence of I(2) Components: Long-Run Relationships, Short-Run Dynamics and Forecasting of the Greek Drachma." *Journal of International Money and Finance*, 19:917–941.

Doornik, J. A. (1998): "Approximations to the Asymptotic Distribution of Cointegration Tests." *Journal of Economic Surveys*, 12(5):573–593.

——— (2002): *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press, 3rd edition.

Fliess, N. and R. MacDonald (2001): "The Instability of the Money Demand Function: an I(2) Interpretation." *Oxford Bulletin of Economics and Statistics*, 63(4):475–495.

Johansen, S. (1997): "Likelihood Analysis of the I(2) Model." *Scandinavian Journal of Statistics*, 24:433–462.

——— (2006): "Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model." *Journal of Econometrics*, 132(1):81–115.

Johansen, S., K. Juselius, R. Frydman, and M. Goldberg (2009): "Testing Hypotheses in an I(2) Model with Piecewise Linear Trends. An Analysis of the Persistent Long Swings in the Dmk/$ Rate." *Journal of Econometrics, Forthcoming.*

Johansen, S., R. Mosconi, and B. Nielsen (2000): "Cointegration in the Presence of Structural Breaks in the Deterministic Trend." *The Econometrics Journal*, 1(3):216–249.

Juselius, K. (1998): "A Structured VAR for Denmark under Changing Monetary Regimes." *Journal of Business and Economic Statistics*, 16(4):400–412.

——— (1999): "Price Convergence in the Medium and Long Run: An I(2) Analysis of Six Price Indices." In: R. F. Engle and H. White (Editors), *Cointegration, Causality, and Forecasting - A Festschrift in Honour of Clive W.J. Granger*. Oxford: Oxford University Press.

Kongsted, H. C. (2005): "Testing the Nominal-To-Real Transformation." *Journal of Econometrics*, 124:205–225.

Kongsted, H. C. and H. B. Nielsen (2004): "Analyzing I(2) Systems by Transformed Vector Autoregressions." *Oxford Bulletin of Economics and Statistics*, 66(3):379–397.

Lettau, M. and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns." *The Journal of Finance*, 56(3):815–849.

Magnus, J. R. and H. Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons.

Nielsen, H. B. (2002): "An I(2) Cointegration Analysis of Price and Quantity Formation in Danish Manufactured Exports." *Oxford Bulletin of Economics and Statistics*, 64(5):449–472.

——— (2007): "A Maximum-Eigenvalue Test for the Cointegration Ranks in I(2) VAR Models." *Economics Letters*, 94(3):445–451.

Nielsen, H. B. and C. Bowdler (2003): "Inflation Adjustment in the Open Economy: An I(2) Analysis of UK Prices." *Working Paper 2003-W5, Nuffield College, Oxford.*

Nielsen, H. B. and A. Rahbek (2007): "The Likelihood Ratio Test for Cointegration Ranks in the I(2) Model." *Econometric Theory*, 23(4):613–635.

Palumbo, M., J. Rudd, and K. Whelan (2006): "On the Relationships Between Real Consumption, Income, and Wealth." *Journal of Business and Economic Statistics*, 25(1):1–11.

Paruolo, P. (2000): "Asymptotic Efficiency of the Two Stage Estimator in I(2) Systems." *Econometric Theory*, 16:524–550.

Perron, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica*, 57(6):1361–1401.

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