An asymptotic invariance property of the common trends under linear transformations of the data*

Søren Johansen†
University of Copenhagen
and CREATES

Katarina Juselius‡
University of Copenhagen

January 16, 2012

Abstract

It is well known that if X_t is a nonstationary process and Y_t is a linear function of X_t, then cointegration of Y_t implies cointegration of X_t. We want to find an analogous result for common trends if X_t is generated by a finite order VAR with i.i.d. $(0, \Omega_x)$ errors ε_{xt}. We first show that Y_t has an infinite order VAR representation in terms of its white noise prediction errors, ε_{yt}, which are a linear process in ε_{xt}, the prediction error for X_t. We then apply this result to show that the limit of the common trends for Y_t generated by ε_{yt}, are linear functions of the common trends for X_t, generated by ε_{xt}.

We illustrate the findings with a small analysis of the term structure of interest rates.

Keywords: Cointegration vectors, common trends, prediction errors

JEL Classification: C32.

*We would like to thank the referee for a careful reading and for many useful comments. The first author wants to thank Søren Tolver Jensen for very helpful discussions on prediction theory and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support.

†Corresponding author. Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K. Denmark, Email: `Soren.Johansen@econ.ku.dk`

‡Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K. Denmark, Email: `Katarina.Juselius@econ.ku.dk`

1 Introduction and motivation

It is well known that if X_t is a $p-$dimensional $I(1)$ process and if the $m-$dimensional linear transformation $Y_t = a'X_t$, $m < p$, is cointegrated, that is, $\beta_y' Y_t$ is stationary for some $\beta_y \neq 0$, then X_t is cointegrated with cointegration vector $a\beta_y$, because $\beta_y' a' X_t = \beta_y' Y_t$ is stationary. Thus cointegration in the small system, Y_t, implies cointegration in the large system, X_t, but not necessarily the other way.

We want to investigate if a similar result holds for common trends. We discuss this in the context of an $I(1)$ cointegrated vector autoregressive process X_t, generated by

$$\Delta X_t = \alpha_x \beta_x' X_{t-1} + \sum_{i=1}^{k} \Gamma_{xi} \Delta X_{t-i} + \varepsilon_{xt}, \tag{1}$$

where ε_{xt} is i.i.d. $(0, \Omega_x)$ and α_x and β_x are $p \times r$. For $I(1)$ processes the solution of (1) is given by the Granger representation

$$X_t = C_x \sum_{i=1}^{t} \varepsilon_{xi} + \sum_{n=0}^{\infty} C_{xn}^* \varepsilon_{xt-n} + A_x, \tag{2}$$

see Johansen (1996, Theorem 4.2). This formulation covers the case $r = p$, where $C_x = 0$ so that X_t is stationary, and $r = 0$, where C_x has full rank, so that X_t is not cointegrating.

The m-dimensional $Y_t = a'X_t$ has a vector moving average (VMA) representation in terms of the $p-$dimensional i.i.d. sequence ε_{xt},

$$Y_t = a'C_x \sum_{i=1}^{t} \varepsilon_{xi} + \sum_{n=0}^{\infty} a'C_{xn}^* \varepsilon_{xt-n} + a'A_x. \tag{3}$$

We define the $\mathcal{L}_2$ space $\mathcal{L}_{2t}^x = \mathcal{L}_2(\beta_x' X_t, \Delta X_s, s \leq t)$ spanned by $\{\beta_x' X_t, \Delta X_s, s \leq t\}$ and the linear projection $\mathcal{P}_{t-1}^x$ onto $\mathcal{L}_{2,t-1}^x$. Then $\mathcal{P}_{t-1}^x(\Delta X_t) = \alpha_x \beta_x' X_{t-1} + \sum_{i=1}^k \Gamma_{xi} \Delta X_{t-i}$ and the prediction error of ΔX_t with respect to $\mathcal{L}_{2,t-1}^x$ is $\varepsilon_{xt} = \Delta X_t - \mathcal{P}_{t-1}^x(\Delta X_t)$. Similarly we define the prediction error of ΔY_t with respect to $\mathcal{L}_{2,t-1}^y$ as $\varepsilon_{yt} = \Delta Y_t - \mathcal{P}_{t-1}^y(\Delta Y_t)$.

If Y_t were a finite order CVAR, with $m-$dimensional prediction errors $\varepsilon_{yt} = \Delta Y_t - \mathcal{P}_{t-1}^y(\Delta Y_t)$, we would find the corresponding Granger representation

$$Y_t = C_y \sum_{i=1}^{t} \varepsilon_{yi} + \sum_{n=0}^{\infty} C_{yn}^* \varepsilon_{yt-n} + A_y, \tag{4}$$

and it is tempting to conclude that the nonstationary part of the two representations (3) and (4) should be the same

$$a'C_x \sum_{i=1}^{t} \varepsilon_{xi} = C_y \sum_{i=1}^{t} \varepsilon_{yi}, \tag{5}$$

and hence, multiplying by $\alpha'_{y\perp}\Gamma_y$, we find $\alpha'_{y\perp}\Gamma_y a' C_x \sum_{i=1}^t \varepsilon_{xi} = \alpha'_{y\perp}\Gamma_y C_y \sum_{i=1}^t \varepsilon_{yi} = \alpha'_{y\perp}\sum_{i=1}^t \varepsilon_{yi}$. Therefore the common trends, $\alpha'_{y\perp}\sum_{i=1}^t \varepsilon_{yi}$, of Y_t based on $\mathcal{L}^y_{2,t-1}$ are a linear function of the common trends, $\alpha'_{x\perp}\sum_{i=1}^t \varepsilon_{xi}$, of X_t based on $\mathcal{L}^x_{2,t-1}$.

It turns out that this simple argument is almost correct, but that (5) only holds in the sense that the difference is stationary, that is, the two random walk components of Y_t are cointegrated, see Figure 2. Their difference does not go to zero with t, only by normalizing by $T^{-1/2}$ and passing to the limit do we find a relation between the limiting Brownian motions. Note that that this does not give any relation between the cyclic components of (3) and (4).

In general Y_t, however, is not a finite order VAR process and therefore the residuals estimated by fitting a finite order VAR do not estimate prediction errors of ΔY_t with respect to $\mathcal{L}^y_{2,t-1}$, moreover one cannot conclude (5) from (3) and (4).

In order to sort this out, we therefore first apply the prediction theory of stationary processes to find an infinite order VAR representation of Y_t, in terms of its white noise prediction errors ε_{yt}, and a corresponding VMA representation, or Granger representation of Y_t, in terms of ε_{yt}. Note, however, that ε_{yt} is a white noise and need not be i.i.d. as we have assumed about ε_{xt}. The exception is of course if ε_{xt} is i.i.d. Gaussian, then also ε_{yt} are i.i.d. Gaussian.

By applying the Functional Limit Theorem to the two cointegrating common trends we can then deduce from (3) and (4) that there is a linear mapping from the limiting common trends, $\alpha'_{x\perp}W_x(u)$, of the large system onto those of the small system: $\alpha'_{y\perp}W_y(u)$.

We illustrate the ideas and findings in an empirical analysis of monthly US interest rates 1987:1 to 2006:1.

2 The process X_t

Let X_t be given by (1) and define the generating matrix polynomial of a complex argument z :

$$\Pi_x(z) = (1-z)I_p - \alpha_x\beta'_x z - \sum_{i=1}^k \Gamma_{xi}(1-z)z^i,$$

and assume that α_x and β_x are $p \times r_x$ of rank $r_x \leq p$.

Under the conditions that the roots of $\det \Pi_x(z) = 0$ satisfy either $|z| > 1$ or $z = 1$, we define $1+\delta > 1$ as the absolute value of the smallest root different from 1:

$$1+\delta = \min\{|z| : \det(\Pi_x(z)) = 0, z \neq 1\}, \tag{6}$$

and $\Gamma_x = I_p - \sum_{i=1}^k \Gamma_{xi}$ and assume $\det(\alpha'_{x\perp}\Gamma_x\beta_{x\perp}) \neq 0$ so that X_t is $I(1)$ and

$$C_x = \beta_{x\perp}(\alpha'_{x\perp}\Gamma_x\beta_{x\perp})^{-1}\alpha'_{x\perp}$$

is well defined. Under these assumptions the polynomial $\Pi_x(z)$ can be inverted in the sense that

$$(1-z)\Pi_x^{-1}(z) = (1-z)\frac{\operatorname{adj}(\Pi_x(z))}{\det(\Pi_x(z))} = C_x + (1-z)C_x^*(z), \tag{7}$$

and $C_x^*(z)$ are rational functions on $\{z : |z| < 1+\delta\}$ satisfying $\beta' C_x^*(1)\alpha = -I_{r_x}$, see Johansen (2009, Theorem 3). These results can be translated into the Granger representation (2).

3 The process $Y_t = a'X_t$

In general $Y_t = a'X_t$, where a is $p \times m$, of rank $m < p$, is not a finite order autoregressive process, but the processes

$$U_{1t} = \beta_y' Y_t = \beta_y' a' \sum_{n=0}^{\infty} C_{xn}^* \varepsilon_{xt-n}, \tag{8}$$

$$U_{2t} = \beta_{y\perp}' \Delta Y_t = \beta_{y\perp}' a'(C_x \varepsilon_{xt} + \sum_{n=0}^{\infty} C_{xn}^* \Delta\varepsilon_{xt-n}), \tag{9}$$

of dimensions r_y and $m-r_y$ respectively are stationary linear processes in the $p-$dimensional prediction errors ε_{xt}. We define the $m \times p$ matrix function

$$\Phi(z) = \begin{pmatrix} \beta_y' a' C_x^*(z) \\ \beta_{y\perp}' a'(C_x + (1-z)C_x^*(z)) \end{pmatrix}, \tag{10}$$

and note that $\Phi(L)\varepsilon_{xt} = U_t = (U_{1t}', U_{2t}')'$, so that the spectral density of U_t is

$$\phi_u(\lambda) = \frac{1}{2\pi}\Phi(e^{i\lambda})\Omega_x \Phi'(e^{-i\lambda}).$$

We first show that U_t is an invertible linear process in its prediction errors $\varepsilon_{ut} = U_t - \mathcal{P}_{t-1}^u(U_t)$, where $\mathcal{P}_{t-1}^u(U_t)$ is the linear projection onto $\mathcal{L}_2^u = \mathcal{L}_2(U_s, s<t) = \mathcal{L}_2^y$.

Lemma 1 *The rational function $\Phi(z)$ is of rank m for $|z| < 1+\delta$, see (6). It follows that there exists an $m \times m$ function $A(z) = \sum_{n=0}^{\infty} A_n z^n$ of full rank for $|z| < 1+\delta$ with real exponentially decreasing coefficients, $A_0 = I_m$, and an $m \times m$ positive definite symmetric matrix Ω_u, so that the spectral density for U_t has the representation*

$$\phi_u(\lambda) = \frac{1}{2\pi}A(e^{i\lambda})\Omega_u A'(e^{-i\lambda}). \tag{11}$$

Moreover, we find the prediction error decomposition (VMA)

$$U_t = \sum_{n=0}^{\infty} A_n \varepsilon_{ut-n}, \tag{12}$$

in terms of the white noise $\varepsilon_{ut} = U_t - \mathcal{P}^u_{t-1}(U_t)$, which inverted gives the infinite order VAR representation of U_t :

$$\varepsilon_{ut} = \sum_{n=0}^{\infty} B_n U_{t-n}. \tag{13}$$

Here the prediction error ε_{ut} is a white noise process with $Var(\varepsilon_{ut}) = \Omega_u$, $A_0 = I_m$, and $A = \sum_{n=0}^{\infty} A_n$ has full rank. The function $B(z) = \sum_{n=0}^{\infty} B_n z^n = A(z)^{-1}$ is defined for $|z| < 1+\delta$ with exponentially decreasing coefficients, $B_0 = I_m$, and $B = \sum_{n=0}^{\infty} B_n = A^{-1}$.

Proof. To prove that $rank(\Phi(z)) = m$ for $|z| < 1 + \delta$, we assume we have z_0 with $|z_0| < 1 + \delta$, and $rank(\Phi(z_0)) < m$. Then we can find $v = (v_1', v_2')' \in \mathbb{R}_m$ so that

$$v_1'\beta_y' a' C_x^*(z_0) + v_2'\beta_{y\perp}' a'(C_x + (1 - z_0)C_x^*(z_0)) = 0. \tag{14}$$

We show that $v = 0$.

Case 1: If $z_0 = 1$, we multiply (14) by α_x from the right and find $v_1'\beta_y' a' C_x^*(1)\alpha_x = 0$ because $C_x\alpha_x = 0$. Because $a\beta_y$ are cointegrating relations for X_t we have $a\beta_y = \beta_x\kappa_1$ for some matrix κ_1 of rank r_y, and $0 = v_1'\beta_y' a' C_x^*(1)\alpha_x = v_1'\kappa_1'\beta_x' C_x^*(1)\alpha_x = -v_1'\kappa_1'$ because $\beta_x' C_x^*(1)\alpha_x = -I_{r_x}$, so that $v_1 = 0$ and therefore from (14), $v_2'\beta_{y\perp}' a' C_x = 0$. But then $a\beta_{y\perp} v_2$ is a cointegrating vector for X_t and $\beta_{y\perp} v_2$ a cointegrating vector for Y_t, which implies that $v_2 = 0$, and hence $v = 0$.

Case 2: If $z_0 \neq 1$, then $(1 - z_0) \neq 0$, and because $\beta_y' a' C_x = 0$ we find

$$0 = v'\Phi(z_0) = [(1 - z_0)^{-1} v_1'\beta_y' a' + v_2'\beta_{y\perp}' a'][C_x + (1 - z_0)C_x^*(z_0)].$$

Now $C_x + (1 - z_0)C_x^*(z_0) = (1 - z_0)\Pi_x(z_0)^{-1}$ has full rank because $\det(\Pi_x(z_0)) \neq 0$, and therefore

$$(1 - z_0)^{-1} v_1'\beta_y' a' + v_2'\beta_{y\perp}' a' = 0.$$

But $\beta_y' a'$ and $\beta_{y\perp}' a'$ are linearly independent which implies that $v_1 = 0$ and $v_2 = 0$.

This proves that $v = 0$, and $rank(\Phi(z)) = m$ for $|z| < 1 + \delta$.

It follows from (7) that the spectral density of U_t, $\phi_u(\lambda)$, is a rational function of the form

$$\phi_u(\lambda) = \frac{\sum_{n=-q}^{q} G_n e^{in\lambda}}{\sum_{n=-q}^{q} g_n e^{in\lambda}},$$

for $m \times m$ matrices $G_n = G_{-n}$ and real $g_n = g_{-n}$, where the roots of both numerator and denominator are greater than $1 + \delta$. From Hannan (1970, Theorem 10', page 66 and page 129) such a function can be written as

$$\frac{\sum_{n=-q}^{q} G_n e^{in\lambda}}{\sum_{n=-q}^{q} g_n e^{in\lambda}} = \frac{1}{2\pi}(\sum_{n=0}^{\infty} A_n e^{in\lambda})\Omega_u(\sum_{n=0}^{\infty} A_n' e^{-in\lambda}),$$

where $A(z) = \sum_{n=0}^{\infty} A_n z^n$ is regular of full rank for $|z| < 1 + \delta$, and A_n is exponentially decreasing, $A_0 = I_m$, and Ω_u is positive definite.

From this result and a similar one for $B(z) = A(z)^{-1}$, follow the two representations (12) and (13). ■

We next apply the VMA representation (12) and the VAR representation (13) to get similar results for Y_t.

Lemma 2 *The stationary process ΔY_t has the prediction error (VMA) decomposition*

$$\Delta Y_t = \sum_{n=0}^{\infty} C_{yn}\varepsilon_{yt-n} = C_y\varepsilon_{yt} + \sum_{n=0}^{\infty} C_{yn}^*\Delta\varepsilon_{yt-n}, \tag{15}$$

where $\varepsilon_{yt} = Y_t - \mathcal{P}_{t-1}^{y}(Y_t)$ is an $m-$dimensional white noise and $C_{y0} = I_m$, $C_y = \sum_{n=0}^{\infty} C_{yn}$ and C_{yn} and C_{yn}^* are exponentially decreasing.

Moreover ΔY_t has an infinite order CVAR representation

$$\Delta Y_t = \alpha_y\beta_y' Y_{t-1} + \sum_{n=1}^{\infty}\Gamma_{yn}\Delta Y_{t-n} + \varepsilon_{yt}, \tag{16}$$

where Γ_{yn} are exponentially decreasing and for $\Gamma_y = I_m - \sum_{n=1}^{\infty}\Gamma_{yn}$ we find

$$C_y = \beta_{y\perp}(\alpha_{y\perp}'\Gamma_y\beta_{y\perp})^{-1}\alpha_{y\perp}'. \tag{17}$$

Finally the white noise prediction errors ε_{yt} are linear processes in the i.i.d. errors ε_{xt} :

$$\varepsilon_{yt} = \sum_{n=0}^{\infty} K_n\varepsilon_{xt-n} = K\varepsilon_{xt} + \sum_{n=0}^{\infty} K_n^*\Delta\varepsilon_{xt-n}, \tag{18}$$

where $K = \sum_{n=0}^{\infty} K_n$ is of rank m, and K_n and K_n^* are exponentially decreasing $m \times p$ matrices.

Proof. *Proof of (15):* We have a representation (12) of $U_t = (Y_t'\beta_y, \Delta Y_t'\beta_{y\perp})$ but need a representation for $\Delta Y_t = \overline{\beta_y}\Delta U_{1t} + \overline{\beta_{y\perp}}U_{2t}$, where we used the notation $\overline{\beta_y} = \beta_y(\beta_y'\beta_y)^{-1}$ and similarly for $\overline{\beta_{y\perp}}$, so that

$$I_m = \overline{\beta_y}\beta_y' + \overline{\beta_{y\perp}}\beta_{y\perp}' = (\overline{\beta_y}, \overline{\beta_{y\perp}})(\beta_y, \beta_{y\perp})'.$$

We define $\varepsilon_{yt} = (\overline{\beta_y}, \overline{\beta_{y\perp}})\varepsilon_{ut}$, and find from (12) that

$$\Delta Y_t = \overline{\beta_y}\Delta U_{1t} + \overline{\beta_{y\perp}}U_{2t} = (\overline{\beta_y}\Delta, \overline{\beta_{y\perp}})\sum_{n=0}^{\infty} A_n(\beta_y, \beta_{y\perp})'\varepsilon_{yt-n} = \sum_{n=0}^{\infty} C_{yn}\varepsilon_{yt-n}, \tag{19}$$

say, where $C_{y0} = (\overline{\beta_y}, \overline{\beta_{y\perp}})A_0(\beta_y, \beta_{y\perp})' = I_m$ and C_{yn} and C_{yn}^* decrease exponentially. This proves (15).

Proof of (16): Similarly we use (13) to find a VAR representation for Y_t. In (13) we define

$$\begin{aligned}\varepsilon_{yt} &= (\overline{\beta_y}, \overline{\beta_{y\perp}})\varepsilon_{ut} = (\overline{\beta_y}, \overline{\beta_{y\perp}}) \sum_{n=0}^{\infty} B_n \begin{pmatrix} \beta_y' Y_{t-n} \\ \beta_{y\perp} \Delta Y_{t-n} \end{pmatrix} \\ &= (\overline{\beta_y}, \overline{\beta_{y\perp}}) \left[B_0 \begin{pmatrix} \beta_y' \Delta Y_t \\ \beta_{y\perp}' \Delta Y_t \end{pmatrix} + \sum_{n=1}^{\infty} B_n \begin{pmatrix} \beta_y'(Y_{t-n} - Y_{t-1}) \\ \beta_{y\perp} \Delta Y_{t-n} \end{pmatrix} + (\sum_{n=0}^{\infty} B_n) \begin{pmatrix} \beta_y' Y_{t-1} \\ 0 \end{pmatrix} \right]\end{aligned} \tag{20}$$

Thus the coefficient of ΔY_t is $(\overline{\beta_y}, \overline{\beta_{y\perp}})B_0 \left(\beta_y, \beta_{y\perp}\right)' = I_m$ and the coefficient of $\beta_y' Y_{t-1}$ is $(\overline{\beta_y}, \overline{\beta_{y\perp}}) \sum_{n=1}^{\infty} B_n \left(I_{r_y}, 0\right)' = -\alpha_y$, say. Then we can write (20) as

$$\varepsilon_{yt} = \Delta Y_t - \sum_{n=1}^{\infty} \Gamma_{yn} \Delta Y_{t-n} - \alpha_y \beta_y' Y_{t-1},$$

for suitable exponentially decreasing coefficients $\Gamma_{yn}, n = 1, \ldots$. This proves (16).

Proof of (17): We find from (19) that

$$C_y = \sum_{n=0}^{\infty} C_{yn} = (0, \overline{\beta_{y\perp}}) A (\beta_y, \beta_{y\perp})' \tag{21}$$

which has rank $m - r_y$ and satisfies $\beta_y' C_y = 0$, so that $C_y = \beta_{y\perp} \kappa$ where κ has rank $m - r_y$. We next find κ.

From (16) we find

$$\alpha_{y\perp}' \Delta Y_t = \sum_{n=1}^{\infty} \alpha_{y\perp}' \Gamma_{yn} \Delta Y_{t-n} + \alpha_{y\perp}' \varepsilon_{yt}, \tag{22}$$

and expanding both sides as in terms of $\varepsilon_{yt}, \Delta\varepsilon_{yt}, \Delta\varepsilon_{yt-1}, \ldots$, we find from (15) that

$$\alpha_{y\perp}' C_y = \sum_{n=1}^{\infty} \alpha_{y\perp}' \Gamma_{yn} C_y + \alpha_{y\perp}' \text{ or } \alpha_{y\perp}' \Gamma_y C_y = \alpha_{y\perp}', \tag{23}$$

where $\Gamma_y = I_m - \sum_{n=1}^{\infty} \Gamma_{yn}$. Now insert $C_y = \beta_{y\perp} \kappa$ and we find $\alpha_{y\perp}' \Gamma_y \beta_{y\perp} \kappa = \alpha_{y\perp}'$ which shows that κ has rank $m - r_y$ and equals $(\alpha_{y\perp}' \Gamma_y \beta_{y\perp})^{-1} \alpha_{y\perp}'$. This proves (17).

Proof of (18): We see from (8) and (9) that U_t is a linear process in ε_{xt}, and from (13) that ε_{yt} is a linear process in U_t, both with exponentially decreasing coefficients. It therefore also holds that the white noise ε_{yt} is a linear process in ε_{xt} with exponentially decreasing coefficients, which we write as (18) for suitable coefficients K_n with $K = B\Phi(1)$ of rank m. ■

Now we can apply the functional limit theorem to the two cumulated predictions errors and prove the main result. We use $\Longrightarrow$ to denote convergence in distribution on $D[0,1]$ or $D^2[0,1]$.

Theorem 3 *For the cumulated prediction errors it holds that*

$$T^{-1/2}(\sum_{t=1}^{[Tu]}\varepsilon_{xt}, \sum_{t=1}^{[Tu]}\varepsilon_{yt}) \Longrightarrow (W_x(u), W_y(u)), \tag{24}$$

where W_y and W_x are Brownian motions and $W_y(u) = KW_x(u)$, see (18).

The relations between cointegration and common trends for Y_t and X_t are then given by

$$a\beta_y = \beta_x\kappa_1, \tag{25}$$
$$\alpha'_{y\perp}W_y = \kappa_2\alpha'_{x\perp}W_x, \tag{26}$$

for some matrices κ_1 and κ_2.

Proof. We have seen in (18) that ε_{yt} is a linear process in the i.i.d. process ε_{xt}. Hence we can apply the functional limit theorem which proves (24). The proof of (25) is trivial.

Finally we find from (3) and (4) the two different representations of Y_t in terms of ε_{yt} common trend and the ε_{xt} common trend,

$$Y_t = C_y\sum_{i=1}^{t}\varepsilon_{yi} + \sum_{n=0}^{\infty}C_n^*\varepsilon_{yt-n} + G_y = a'C_x\sum_{i=1}^{t}\varepsilon_{xi} + \sum_{n=0}^{\infty}a'C_{xn}^*\varepsilon_{xt-n} + a'G_x.$$

It is not so easy to disentangle the random walk part from the stationary part of these expressions, but if we divide by $T^{-1/2}$ and pass to the limit for $t = [Tu]$, and use (24), we find a simpler expression

$$C_yW_y(u) = a'C_xW_x(u),$$

and multiplying by $\alpha'_{y\perp}\Gamma_y$ we find

$$\alpha'_{y\perp}W_y = \alpha'_{y\perp}\Gamma_y a'\beta_{x\perp}(\alpha'_{x\perp}\Gamma_x\beta_{x\perp})^{-1}\alpha'_{x\perp}W_x = \kappa_2\alpha'_{x\perp}W_x.$$

■

4 Estimation of the infinite order CVAR for Y_t

The representation of Y_t as the solution of an infinite order CVAR, see (16), suggests fitting a CVAR of lag order $k = k_T$, and analyse the properties of the estimators $\hat{\alpha}_y^{(k)}, \hat{\beta}_y^{(k)}$, and $\hat{\Gamma}_{yn}^{(k)}, n = 1, \ldots, k$ for $k = k_T \to \infty$ with T. For readability we suppress the subscript T of k in the estimators.

Saikkonen (1992) analysed this problem for the triangular form of the VAR with the added assumption that the prediction errors ε_{yt} were in fact independent. The corresponding result for white noise errors has not been developed for the infinite order

CVAR, but Raïssi (2009) found the asymptotic results for Gaussian MLE assuming only white noise errors with a mixing condition.

In order to apply the result of Saikkonen, we therefore assume in the asymptotic analysis below that ε_{xt} is i.i.d. $N(0, \Omega_x)$, so that also ε_{yt} are i.i.d. $N_m(0, \Omega_y)$.

If we apply the usual reduced rank (QMLE) for estimation of the parameters in a CVAR of order k_T for Y_t, it follows from Saikkonen (1992) that provided $k_T^3/T \to 0$ and $E|\varepsilon_{yt}|^4 < \infty$, we have $(\hat{\alpha}_y^{(k)}, \hat{\beta}_y^{(k)}) \xrightarrow{P} (\alpha_y, \beta_y)$ and that the limit distribution of the test for rank r_y has the usual limit distribution.

In Saikkonen and Lütkepohl (1996) the short-run dynamics is investigated and if we write (16) in the form

$$\Delta^2 Y_t = \alpha_y \beta_y' Y_{t-1} - \Gamma_y \Delta Y_{t-1} + \sum_{n=1}^{\infty} \Gamma_{yn}^* \Delta^2 Y_{t-n} + \varepsilon_{yt}, \tag{27}$$

their results show that the matrix Γ_y is estimated consistently from a finite order CVAR with increasing lag length k_T. This shows that usual asymptotic inference is possible both for the cointegrating rank of Y_t and for the long-run matrix C_y.

Lemma 4 *Fitting a CVAR(k_T) to data generated by (27), where ε_{yt} is i.i.d. $N_m(0, \Omega_y)$, we find for $k_T^3/T \to 0$ that*

$$T^{-1/2} \hat{C}_y^{(k)} \sum_{i=1}^{[Tu]} \hat{\varepsilon}_{yi}^{(k)} \Longrightarrow C_y W_y(u).$$

Proof. Estimating a CVAR(k_T) we find $\hat{\varepsilon}_{yi}^{(k)}$, and because $\hat{C}_y^{(k)} \hat{\alpha}_y^{(k)} \beta_y' = 0$, we find that $\hat{C}_y^{(k)} \hat{\varepsilon}_{yi}^{(k)}$ is

$$\hat{C}_y^{(k)}(\varepsilon_{yi} + (\alpha_y - \hat{\alpha}_y^{(k)})\beta_y' Y_{i-1} + (\hat{\Gamma}_y - \Gamma_y)\Delta Y_{i-1} - \sum_{n=1}^{k} (\hat{\Gamma}_{yn}^{*(k)} - \Gamma_{yn}^*)\Delta^2 Y_{i-n} + \sum_{n=k+1}^{\infty} \Gamma_{yn}^* \Delta^2 Y_{i-n}). \tag{28}$$

Summing to $[Tu]$ and normalizing with $T^{-1/2}$, the first term converges to $C_y W_y(u)$. Saikkonen and Lütkepohl (1996) show that $||\hat{\alpha}_y^{(k)} - \alpha_y||_2$, $||\hat{\Gamma}_{yn}^{*(k)} - \Gamma_{yn}^*||_2$ and $||\hat{\Gamma}_y^{(k)} - \Gamma_y||_2$ are $O_P((k_T/T)^{1/2})$. We therefore find that the next three terms of (28) are bounded by

$$\begin{aligned} \hat{C}_y^{(k)}(\alpha_y - \hat{\alpha}_y^{(k)})[T^{-1/2} \sum_{i=1}^{[Tu]} \beta_y' Y_{i-1}] &= O_P((k_T/T)^{1/2}) \xrightarrow{P} 0 \\ T^{-1/2} ||\hat{\Gamma}_y - \Gamma_y||_2 ||Y_{[Tu]-1} - Y_{-1}||_2 &= T^{-1/2} O_P((k_T/T)^{1/2}) \xrightarrow{P} 0 \end{aligned}$$

and

$$||\hat{C}_y^{(k)}||_2 \sum_{n=1}^{k_T} ||\Gamma_{yn}^{*(k)} - \Gamma_{yn}^*||_2 T^{-1/2} ||\Delta Y_{[Tu]-n} - \Delta Y_{-n}||_2 = k_T O_P((k_T/T)^{1/2}) T^{-1/2} \xrightarrow{P} 0.$$

Finally, because $|| \sum_{t=1}^{[Tu]} \Delta^2 Y_{t-n}||_2 = ||\Delta Y_{[Tu]-n} - \Delta Y_{-n}||_2 \leq c$ and $||\Gamma_{yn}^*||_2 \leq \rho^n$ for some $0 < \rho < 1$, the last term tends to zero in probability. ∎

5 An illustration using US interest rates

We consider US monthly interest rates in the period 1987:1 to 2006:1 which defines the period when Greenspan was the chairperson of the Federal Reserve System. The data is taken from IMF's financial database and consists of the following four interest rates: the federal funds rate i_{ff}, the 6 month treasury bills rate, the 3 year and 10 year bond rates, denoted $i_{6m},\ i_{3y},\ i_{10y}$ respectively.

The baseline VAR model has two lags and a constant term restricted to the cointegration space,

$$\Delta X_t = \alpha_x(\beta_x' X_{t-1} + \mu_x') + \Gamma_{x1}\Delta X_{t-1} + \varepsilon_t, \tag{29}$$

where $X_t = [s_{ff6m}, s_{6m3y}, i_{3y}, i_{10y}]_t$, $s_{ff6m} = i_{ff} - i_{6m}$, and $s_{6m3y} = i_{6m} - i_{3y}$. The motivation for using the equivalent transformation of two spreads and two levels rather than the four interest rates is because it facilitates the interpretation of the common stochastic trends.

To illustrate the (asymptotic) invariance of common stochastic trends to extensions of the information set we first report the results from a VAR analysis when the 10 year bond rate has been left out and then put it back.

In the small system, $Y_t = [s_{ff6m}, s_{6m3y}, i_{3y}]$, the trace test and the roots of the characteristic polynomial suggest one cointegration relation and, hence, two stochastic trends. The estimates of β_y and α_y are reported in Table 1. The cointegration relation is a combination of the two spreads and a small level effect from the 3 year rate.

Table 1: The small system results

The cointegration estimates				
	s_{ff6m}	s_{6m3y}	i_{3y}	*constant*
β_y'	1	-0.15 [-3.40]	0.04 [2.66]	0.00 [0.77]
α_y'	-0.29 [-5.80]	0.01 [0.24]	0.04 [0.48]	

Note: t–ratios in brackets

The α_y coefficients show that the shortest spread is significantly adjusting but not the longer spread and the 3 year bond rate, suggesting that s_{6m3y} and i_{3y} might be weakly exogenous which is strongly supported by the joint test of weak exogeneity (p-value 0.79). Hence, the two stochastic trends can be associated with cumulated shocks to the level of the longest interest rate and the spread between the 6 month and the 3 year rate. Thus, the term structure of interest rates seems to be driven by the shocks to a level and a slope component, similar to the finding in Giese (2008).

In the extended system the trace test and the roots of the characteristic polynomials suggest that the rank is two and, hence, that there are two stochastic trends. Thus, the same number of common stochastic trends are driving the small and the extended system. Table 2 reports the β_x and α_x estimates of the two just identified cointegration relations. The identifying zero restriction on 10 year bond rate in the first cointegration relation makes it comparable with the cointegration relation in the small system. It is notable that its estimated coefficients are identical for both relations, illustrating the point that cointegration in the small system implies cointegration in the large system. The second cointegration relation is identified by the zero restriction on the spread between the federal funds rate and the 6 months rate. It is notable that the second relation suggests that the 'curvature' of the term structure is stationary, i.e. $\{(i_{6m} - i_{3y}) - (i_{3y} - i_{10y})\} \sim I(0)$.

Table 2: The large system results

The cointegration estimates					
	s_{ff6m}	s_{6m3y}	i_{3y}	i_{10y}	$constant$
β'_{x1}	1	-0.16 [-3.64]	0.05 [3.14]	0.00	0.00 [0.43]
β'_{x2}	0.00	1.00	-0.99 [-8.82]	1.00 [7.49]	-0.00 [-1.23]
α'_{x1}	-0.29 [-5.47]	0.08 [1.58]	0.00 [0.02]	0.07 [0.92]	
α'_{x2}	-0.00 [-0.13]	-0.13 [-4.32]	0.06 [1.31]	0.03 [0.71]	

Note: t–ratios in brackets.

The joint test of weak exogeneity of i_{3y} and i_{10y} (p-value 0.15) suggests that the two long rates can be considered weakly exogenous, implying that their cumulated shocks define the two common stochastic trends. This is consistent with the estimated α_x coefficients that are insignificant for the two longest rates. Thus, the results support the previous small system interpretation that the term structure can be described by a nonstationary level and slope component and a stationary curvature.

In the small model it was the shocks to i_{3y} and s_{6m3y} that were found to drive the system, whereas in the large model it is the shocks to the two long rates, i_{3y} and i_{10y}, that drive the system. Thus, the 10 year bond rate has now taken over the role as a weakly exogenous variable from the spread s_{6m3y}. As the realized random walk components of a variable are asymptotically the same, independently of the dimension of the system, the two stochastic trends, whether estimated from the small or the large model, should be able to replicate this realized random walk component. Thus, what is (asymptotically) invariant is (the space spanned by) the random walks, but not the interpretation in terms of a structural shock with a given label. Only when adding more variables does not change the definition of an exogenous shock, is the information set sufficiently large to warrant invariance of labels. Thus, the common trends are asymptotically invariant, but their interpretation depends on the information set.

To illustrate how closely the two estimated stochastic trends replicate the realized random walk components of each variable Figure 1 plots each of the four variables and their random walk component. In the left hand panels, the random walk components are based on the stochastic trends from the big model, whereas in the right hand panels they are from the small model. The exogenous long-term interest rates, i_{3y} and i_{10y}, are very close to their random walk component, whereas the two spreads differ more. In particular, the shortest spread appears to be dominated by short-run variation. In all cases the random walk component seems to captures the long-term movements of the variables.

The estimated random walk component of each variable looks very similar in the two systems. This is confirmed by Figure 2, which compares the random walk component estimated by the small and the large model, respectively, for each of the first three variables. While not identical, they capture much the same pattern in the series.

6 Conclusion

We have analyzed two common trends representations of a linear transformation of a VAR process. One is in terms of the prediction errors of the VAR process, and the other in terms of the prediction errors of the transformed system. Such common trends are cointegrated and by normalizing and taking the limit we show that there is a linear transformation from the limit Brownian motions of the VAR to the limit Brownian motions of the transformed variable.

Empirical VAR models are often based on a small set of key economic variables assuming that other relevant omitted variables, if added to the model, would not affect the basic conclusions. It has long been well known that the cointegration property is invariant to extensions of the information set; if cointegration is found within a smaller set of variables it will also be found within an extended set. The paper has shown that this invariance property is also asymptotically valid for common trends. From a practical point of view this result has some important implications for how the choice of information set influences results in applied work.

For example, if the purpose of the analysis is a trend-cycle decomposition, the stochastic trend extracted by the Beverage-Nelson decomposition, say, is asymptotically invariant to changes in the information set. Nonetheless, this is only an asymptotic result, and if the sample is short and the short-run variation of the variables is relatively large, deviations between the common trends can occur. If, on the other hand, the purpose is explanation, a structural interpretation of shocks is likely to be hazardous. This is because their interpretation may vary considerably when extending the information set even though the permanent shocks are the same. This was illustrated in the example by the finding that in the small model it was the shocks to the 6 month - 3 year spread and to the 3 year bond rate that had generated the two stochastic trends, whereas in the extended model it was the shocks to the 3 year and 10 year bond rate.

7 References

1. Giese, J.V., 2008. Level, slope, curvature: characterizing the yield curve in a cointegrated VAR model. Economics: The Open-Access, Open-Assessment E-Journal, Vol. 2, 2008-28.

2. Hannan, E.J., 1970. Multiple Time Series. John Wiley and Sons.

3. Johansen, S., 1996. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

4. Johansen, S., 2009. Representation of cointegrated autoregressive processes with application to fractional processes. Econometric Reviews 28, 121–145.

5. Raïssi, H., 2009. Testing the cointegrating rank when the errors are uncorrelated but nonindependent. Stochastic Analysis and Applications 27, 24–50.

6. Saikkonen, P., 1992. Estimation and testing of cointegrated systems by an autoregressive approximation. Econometric Theory 8, 1–27.

7. Saikkonen, P., Lütkepohl. H., 1996. Infinite order cointegrated vector autoregressive processes. Estimation and inference. Econometric Theory 12, 814–844.

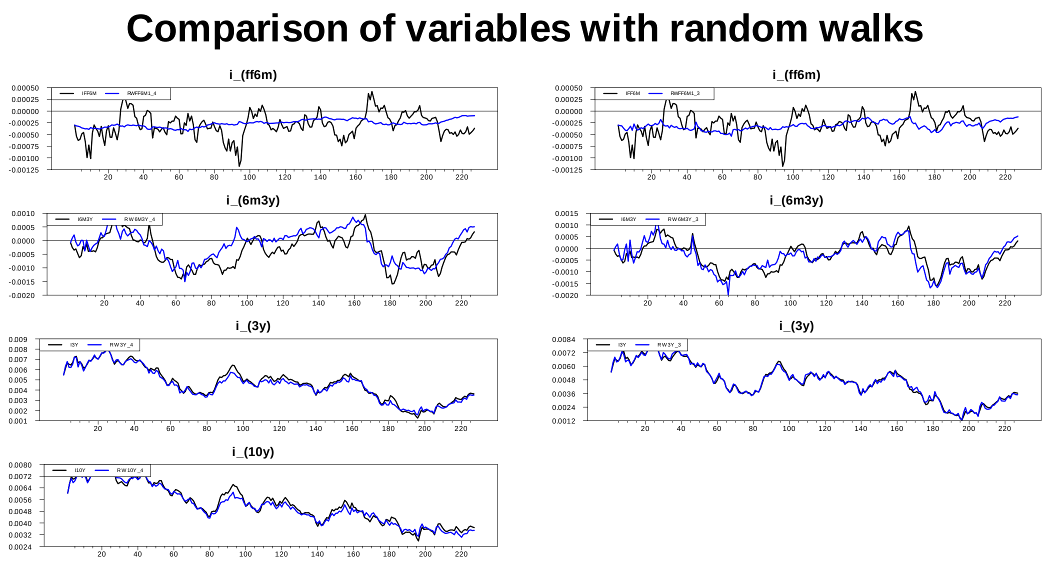

Figure 1: The plots show the variables i_{ff6m}, i_{6m3y}, i_{3y}, i_{10y} compared to the their random walk component. In the left panel the random walk is constructed from the large system and in the right hand panel from the small system.

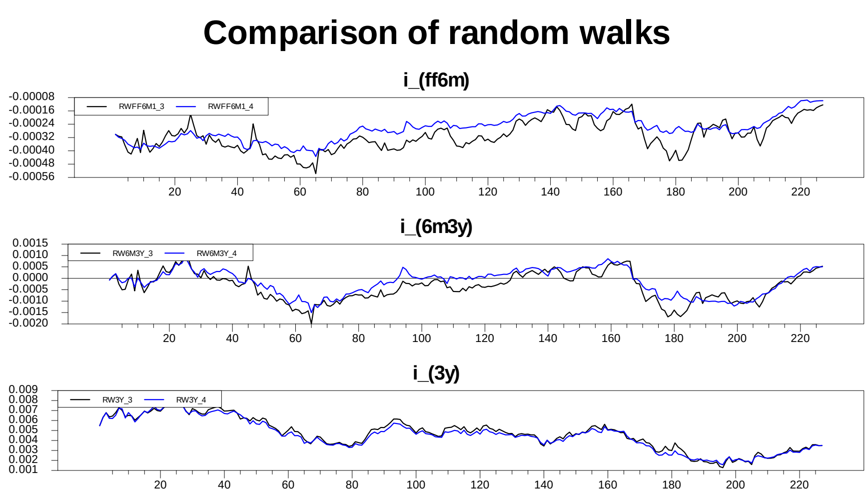

Figure 2: The plots compare the random walk component of each of the variables i_{ff6m}, i_{6m3y}, i_{3y} estimated from the large and the small system. The finding is that the random walk components are roughly the same in the two systems, which illustrates the asymptotic invariance shown in the paper for this example.