

Discussion Papers  
Department of Economics  
University of Copenhagen

No. 10-30

An Invariance Property of the Common Trends  
under Linear Transformations of the Data

Søren Johansen, Katarina Juselius

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 (E)

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

Søren Johansen<sup>†</sup>  
University of Copenhagen  
and CREATES

Katarina Juselius<sup>‡</sup>  
University of Copenhagen

October 31, 2010

## 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. We first show that  $Y_t$  has an infinite order VAR representation in terms of its prediction errors, which are a linear process in the prediction error for  $X_t$ . We then apply this result to show that the limit of the common trends for  $Y_t$  are linear functions of the common trends for  $X_t$ .

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.

---

\*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.

<sup>†</sup>Corresponding author. Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K. Denmark, Email: [Soren.Johansen@econ.ku.dk](mailto:Soren.Johansen@econ.ku.dk)

<sup>‡</sup>Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K. Denmark, Email: [Katarina.Juselius@econ.ku.dk](mailto: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 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

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

where  $\varepsilon_{xt}$  is the i.i.d.  $(0, \Omega_x)$   $p$ -dimensional prediction error. Under  $I(1)$  conditions the solution is given by the Granger representation

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

where  $C_x = \beta_{x\perp} (\alpha'_{x\perp} \Gamma_x \beta_{x\perp})^{-1} \alpha'_{x\perp}$ ,  $\Gamma_x = I_p - \sum_{i=1}^k \Gamma_{xi}$  and  $A_x$  depends on initial values so that  $\beta_x' A_x = 0$ , and the coefficient matrices  $C_{nx}^*$  are exponentially decreasing, see Johansen (1996).

The linear transformation  $Y_t = a'X_t$ , therefore, has the representation

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

which shows that the common trends of the  $m$ -dimensional process  $Y_t$  can be described easily in terms of the  $p$ -dimensional prediction errors for  $X_t$ . If  $Y_t$  is a finite order CVAR, with  $m$ -dimensional prediction errors  $\varepsilon_{yt}$ , we find the corresponding Granger representation

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

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

$$a' C_x \sum_{i=1}^t \varepsilon_{xi} = C_y \sum_{i=1}^t \varepsilon_{yi}. \quad (5)$$

These equations can be solved for  $\alpha'_{y\perp} \sum_{i=1}^t \varepsilon_{yi}$  by multiplication by  $\alpha'_{y\perp} \Gamma_y C_y$ , and therefore the common trends,  $\alpha'_{y\perp} \sum_{i=1}^t \varepsilon_{yi}$ , of  $Y_t$  are linearly related to the common trends,  $\alpha'_{x\perp} \sum_{i=1}^t \varepsilon_{xi}$ , of  $X_t$ .

It turns out that this simple argument is almost correct, but that (5) only holds as an approximation in the sense that the difference is stationary, that is, the two random walk components of  $Y_t$  are cointegrated, see Figure 2. Only by normalizing by  $T^{-1/2}$  and passing to the limit do we find a relation between the limiting Brownian motions.

The problem with this argument is that in general  $Y_t$  is not a finite order VAR process and therefore the  $e_{yt}$  estimated by fitting a finite order VAR do not estimate prediction errors. In order to equate the two Granger representations and take the limit we need to understand the process  $\varepsilon_{yt}$  defined as the  $m$ -dimensional prediction error from the linear process  $Y_t$ , and discuss how to estimate it by fitting a finite order VAR to  $Y_t$ .

We therefore first apply the prediction theory of stationary processes to find an infinite order VAR representation of  $Y_t$ , which defines the correct prediction errors  $\varepsilon_{yt}$  as a linear process in  $\varepsilon_{xt}$  and we also get a VMA representation, or Granger representation of  $Y_t$ , in terms of these. Note, however, that  $\varepsilon_{yt}$  need not be i.i.d.

In this way we can 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:  $\alpha'_{x\perp} W_x$ .

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  $\alpha_x$  and  $\beta_x$  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

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

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

$$(1 - z)\Pi_x(z)^{-1} = (1 - z) \frac{\text{adj} \Pi_x(z)}{\det \Pi_x(z)} = C_x + (1 - z)C_x^*(z), \quad (6)$$

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  $Y_t = a'X_t$  is not a finite order autoregressive process, but the processes

$$U_{1t} = \beta'_y Y_t = \beta'_y a' \sum_{n=0}^{\infty} C_{xn}^* \varepsilon_{xt-n}, \quad (7)$$

$$U_{2t} = \beta'_{y\perp} \Delta Y_t = \beta'_{y\perp} a' (C_x \varepsilon_{xt} + \sum_{n=0}^{\infty} C_{xn}^* \Delta \varepsilon_{xt-n}), \quad (8)$$

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

$$\Phi(z) = \begin{pmatrix} \beta'_y a' C_x^*(z) \\ \beta'_{y\perp} a' (C_x + (1-z)C_x^*(z)) \end{pmatrix}, \quad (9)$$

and note that

$$\Phi(L)X_t = U_t = (U'_{1t}, U'_{2t})',$$

so that the spectral density 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.

**Lemma 1** *The rational function  $\Phi(z)$  is of rank  $m$  for  $|z| < 1 + \delta$ .*

*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  $\Omega_u$ , so that the spectral density for  $U_t$  has the representation*

$$\phi_u(\lambda) = \frac{1}{2\pi} A(e^{i\lambda}) \Omega_u A'(e^{-i\lambda}). \quad (10)$$

*Moreover, we find the prediction error decomposition (VMA)*

$$U_t = \sum_{n=0}^{\infty} A_n \varepsilon_{ut-n}, \quad (11)$$

*in terms of the white noise  $\varepsilon_{ut}$  which gives the VAR representation of  $U_t$ :*

$$\varepsilon_{ut} = \sum_{n=0}^{\infty} B_n U_{t-n}. \quad (12)$$

*Here the prediction error  $\varepsilon_{ut}$  is a white noise process with  $\text{Var}(\varepsilon_{ut}) = \Omega_u$ ,  $A_0 = I_m$ , 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  $\text{rank}(\Phi(z)) = m$  for  $|z| < 1 + \delta$ , we assume we have  $z_0$  with  $|z_0| < 1 + \delta$ , and  $\text{rank}(\Phi(z_0)) < m$ . Then we can find  $v = (v'_1, v'_2)' \in \mathbb{R}_m$  so that  $v'\Phi(z_0) = 0$ , or

$$v'_1\beta'_y a' C_x^*(z_0) + v'_2\beta'_{y\perp} a' (C_x + (1 - z_0)C_x^*(z_0)) = 0. \quad (13)$$

We show that  $v = 0$ .

*Case 1:* If  $z_0 = 1$ , we multiply (13) by  $\alpha_x$  from the right and find  $v'_1\beta'_y a' C_x^*(1)\alpha_x = 0$  because  $C_x\alpha_x = 0$ . Because  $a\beta_y$  is a cointegrating relation for  $X_t$  we have  $a\beta_y = \beta_x\kappa_1$  for some matrix  $\kappa_1$ , 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  $\beta'_x C_x^*(1)\alpha_x = -I_{r_x}$ . This implies that  $v'_1\beta'_y a' C_x^*(1) = v'_1\kappa'_1\beta'_x C_x^*(1) = 0$ , and therefore  $v'_2\beta'_{y\perp} a' C_x = 0$ . 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  $v_2 = 0$ .

*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  $\Pi_x(z)$  is a polynomial, 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  $\text{rank}(\Phi(z)) = m$  for  $|z| < 1 + \delta$ .

It follows from (6) that the spectral density of  $U_t$ ,  $\phi_u(\lambda)$ , is a rational function of the form

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

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

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

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  $\Omega_u$  is positive definite.

From this result and a similar one for  $B(z) = A(z)^{-1}$ , follow the two representations (11) and (12). ■

We next apply the VMA representation (11) and the VAR representation (12) to get similar results for  $Y_t$ .

**Lemma 2** *The stationary process  $\Delta Y_t$  has the prediction error (VMA) decomposition*

$$\Delta Y_t = \sum_{n=0}^{\infty} C_{yn} \varepsilon_{yt-n} = C_y \varepsilon_{yt} + \sum_{n=0}^{\infty} C_{yn}^* \Delta \varepsilon_{yt-n}, \quad (14)$$

where  $\varepsilon_{yt}$  is  $m$ -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  $\Delta Y_t$  has an infinite order CVAR representation

$$\Delta Y_t = \alpha_y \beta'_y Y_{t-1} + \sum_{n=1}^{\infty} \Gamma_{yn} \Delta Y_{t-n} + \varepsilon_{yt}, \quad (15)$$

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

$$C_y = \beta_{y\perp} (\alpha'_{y\perp} \Gamma_y \beta_{y\perp})^{-1} \alpha'_{y\perp}. \quad (16)$$

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

$$\varepsilon_{yt} = \sum_{n=0}^{\infty} K_n \varepsilon_{xt-n} = K \varepsilon_{xt} + \sum_{n=0}^{\infty} K_n^* \Delta \varepsilon_{xt-n}, \quad (17)$$

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.

Note that  $\varepsilon_{yt}$  is a white noise sequence but not necessarily an i.i.d. sequence.

**Proof.** *Proof of (14):* We have a representation (11) 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 (11) that

$$\Delta Y_t = \overline{\beta}_y \Delta U_{1t} + \overline{\beta}_{y\perp} U_{2t} = (\overline{\beta}_y \Delta, \overline{\beta}_{y\perp}) \sum_{n=0}^{\infty} A_n (\beta_y, \beta_{y\perp})' \varepsilon_{yt-n} = \sum_{n=0}^{\infty} C_{yn} \varepsilon_{yt-n}, \quad (18)$$

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 (14).

*Proof of (15):* Similarly we use (12) to find a VAR representation for  $Y_t$ . We define the  $\mathcal{L}_2$  space

$$\mathcal{L}_2\{U_s, s \leq t\} = \mathcal{L}_2\{\beta_y' Y_s, \beta_{y\perp}' \Delta Y_s, s \leq t\},$$

and note that because  $\beta_y' Y_{t-n} = \beta_y' Y_{t-1} - \sum_{v=1}^{n-1} \Delta \beta_y' Y_{t-v}$ , we can eliminate  $\beta_y' Y_{t-n}$  for  $n = 0, 2, 3, \dots$  and find

$$\mathcal{L}_2\{U_s, s \leq t\} = \mathcal{L}_2\{\beta_y' Y_{t-1}, \Delta Y_s, s \leq t\},$$

so that a linear function of  $U_s, s \leq t$  is a linear function of  $\beta_y' Y_{t-1}$  and  $\Delta Y_t, \Delta Y_{t-1}, \dots$ . Then from (12) we find

$$\begin{aligned} \varepsilon_{yt} &= (\overline{\beta}_y, \overline{\beta}_{y\perp}) \varepsilon_{ut} = (\overline{\beta}_y, \overline{\beta}_{y\perp}) \sum_{n=0}^{\infty} B_n \begin{pmatrix} \beta_y' Y_{t-n} \\ \beta_{y\perp}' \Delta Y_{t-n} \end{pmatrix} \\ &= (\overline{\beta}_y, \overline{\beta}_{y\perp}) \left[ B_0 \begin{pmatrix} \beta_y' \Delta Y_t \\ \beta_{y\perp}' \Delta Y_t \end{pmatrix} + \sum_{n=0}^{\infty} B_n \begin{pmatrix} \beta_y' (Y_{t-n} - Y_{t-1}) \\ \beta_{y\perp}' \Delta Y_{t-n} \end{pmatrix} + \sum_{n=0}^{\infty} B_n \begin{pmatrix} \beta_y' Y_{t-1} \\ 0 \end{pmatrix} \right] \end{aligned} \quad (19)$$

Thus the coefficient of  $\Delta Y_t$  is  $(\overline{\beta_y}, \overline{\beta_{y\perp}})B_0(\beta_y, \beta_{y\perp})' = I_m$  and the coefficient of  $\beta_y' Y_{t-1}$  is  $(\overline{\beta_y}, \overline{\beta_{y\perp}}) \sum_{n=0}^{\infty} B_n(I_{r_y}, 0)' = -\alpha_y$ , say. Then we can write (19) as

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

for suitable exponentially decreasing coefficients  $\Gamma_{yn}, n = 1, \dots$ . This proves (15).

*Proof of (16):* We find from (18) that

$$C_y = \sum_{n=0}^{\infty} C_{yn} = (0, \beta_{y\perp}) A(\beta_y, \beta_{y\perp})' \quad (20)$$

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

From (15) we find

$$\alpha'_{y\perp} \Delta Y_t = \sum_{n=1}^{\infty} \alpha'_{y\perp} \Gamma_{yn} \Delta Y_{t-n} + \alpha'_{y\perp} \varepsilon_{yt}, \quad (21)$$

and inserting Identifying (14) we can identify coefficients to  $\varepsilon_{yt}$  in an expansion in terms of  $\varepsilon_{yt}, \Delta \varepsilon_{yt}, \Delta \varepsilon_{yt-1}, \dots$ . We find the identity

$$\alpha'_{y\perp} C_y = \sum_{n=1}^{\infty} \alpha'_{y\perp} \Gamma_{yn} C_y + \alpha'_{y\perp} \text{ or } \alpha'_{y\perp} \Gamma_y C_y = \alpha'_{y\perp}, \quad (22)$$

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  $\kappa$  has rank  $m - r_y$  and equals  $(\alpha'_{y\perp} \Gamma_y \beta_{y\perp})^{-1} \alpha'_{y\perp}$ . This proves (16).

*Proof of (17):* Finally we see from (7) and (8) that  $U_t$  is a linear process in  $\varepsilon_{xt}$ , and from (12) that  $\varepsilon_{yt}$  is a linear process in  $U_t$ , both with exponentially decreasing coefficients. It therefore also holds that the white noise  $\varepsilon_{yt}$  is a linear process in  $\varepsilon_{xt}$  with exponentially decreasing coefficients, which we write as (17) for suitable coefficients  $K_n$  with  $K = A\Phi(1)$  of rank  $m$ . ■

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

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

$$T^{-1/2} \left( \sum_{t=1}^{[Tu]} \varepsilon_{xt}, \sum_{t=1}^{[Tu]} \varepsilon_{yt} \right) \implies (W_x(u), W_y(u)), \quad (23)$$

where  $W_y$  and  $W_x$  are Brownian motions and  $W_y(u) = KW_x$ , see (17).



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

$$a\beta_y = \beta_x\kappa_1, \quad (24)$$

$$\alpha'_{y\perp}W_y = \kappa_2\alpha'_{x\perp}W_x, \quad (25)$$

for some matrices  $\kappa_1$  and  $\kappa_2$ .

**Proof.** We have seen in (17) that  $\varepsilon_{yt}$  is a linear process in the i.i.d. process  $\varepsilon_{xt}$ . Hence we can apply the functional limit theorem which proves (23). The proof of (24) is trivial.

Finally we find from (3) and (4) the two different representations of  $Y_t$  in terms of common trends,

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

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 (23), we find a simpler expression

$$C_y W_y(u) = a' C_x W_x(u),$$

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 (15),

$$\Delta Y_t = \alpha_y \beta'_y Y_{t-1} + \sum_{n=1}^{\infty} \Gamma_{yn} \Delta Y_{t-n} + \varepsilon_{yt}, \quad (26)$$

suggests fitting a  $k$ th order CVAR, and analyse the properties of the estimators  $\hat{\alpha}_y^{(k)}$ ,  $\hat{\beta}_y^{(k)}$ , and  $\hat{\Gamma}_{yn}^{(k)}$ ,  $n = 1, \dots, k$  for  $k \rightarrow \infty$  with  $T$ .

Saikkonen (1992) analysed this problem for the triangular form of the VAR with the added assumption that the prediction errors  $\varepsilon_{yt}$  were in fact independent. In order to apply his results we therefore assume, in the asymptotic analysis below, that  $\varepsilon_{xt}$  is

i.i.d.  $N(0, \Omega_x)$ , so that also  $\varepsilon_{yt}$  are i.i.d.  $N_m(0, \Omega_y)$ . The triangular form requires for a given rank, that a matrix  $c$  is known for which we can assume  $c'\beta_y$  has full rank, so that  $c'_\perp Y_t$  is not cointegrated. If we define  $\theta = -\bar{c}'_\perp \beta_y$  so that  $\beta_y = \bar{c} - c_\perp \theta'$ , then the processes  $Y_{1t} = \bar{c}' Y_t$  and  $Y_{2t} = c'_\perp Y_t$  are cointegrated because  $\beta'_y Y_t = Y_{1t} - \theta Y_{2t}$ , and the equations (26) can be written in triangular representation

$$\begin{aligned} Y_{1t} &= \theta Y_{2t} + v_{1t} \\ \Delta Y_{2t} &= v_{2t} \end{aligned}$$

where the stationary process  $v_t$  is given by

$$v_t = \begin{pmatrix} \beta'_y Y_t \\ c'_\perp \Delta Y_t \end{pmatrix}.$$

It is seen in the same way as in Lemma 1, where  $c = \beta_y$ , that the error process  $v_t$  has an infinite VAR representation with a nonsingular long-run impact matrix.

If we apply the usual reduced rank (QMLE) for estimation of the parameters in a CVAR of order  $k$  for  $Y_t$ , it follows from Saikkonen (1992) that provided  $k^3/T \rightarrow 0$  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 (26) in the form

$$\Delta^2 Y_t = \alpha_y \beta'_y Y_{t-1} - \Gamma_y \Delta X_{t-1} + \sum_{n=1}^{\infty} \Gamma_{yn}^* \Delta^2 Y_{t-n} + \varepsilon_{yt}, \quad (27)$$

their results show that the matrix  $\Gamma_y$  is estimated consistently estimating a finite order CVAR to (27). This shows that usual asymptotic inference is possible both for the cointegrating rank of  $Y_t$  and for the long-run matrix  $C_y$ , and that

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

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

**Proof.** Because  $\hat{C}_y^{(k)} \hat{\alpha}_y^{(k)} \beta'_y = 0$  we get  $\hat{C}_y^{(k)} \hat{\varepsilon}_{yi}^{(k)} = \hat{C}_y^{(k)} (\Delta^2 Y_i - \widehat{\Delta^2 Y}_i^{(k)})$  which becomes

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

Summing to  $[Tu]$  and normalizing with  $T^{-1/2}$ , the last term converges to  $C_y W_y(u)$  and the remaining terms tend to zero. The result of Saikkonen and Lütkepohl (1996) is

that  $\|\alpha_y - \hat{\alpha}_y^{(k)}\|$ ,  $\|\hat{\Gamma}_{yn}^{*(k)} - \Gamma_{yn}^*\|$  and  $\|\hat{\Gamma}_y^{*(k)} - \Gamma_y^*\|$  are  $O_P((k/T)^{1/2})$ . We therefore find that

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

and

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

Finally, because  $\|\Delta^2 Y_{t-n}\|_2 \leq c$  we have, because the matrices  $\Gamma_{yn}^*$  are exponentially decreasing, that there is a  $\rho < 1$ , so that

$$\|\hat{C}_y^{(k)} \sum_{n=k+1}^{\infty} \Gamma_{yn}^* T^{-1/2} \sum_{t=1}^{[Tu]} \Delta Y_{t-n}\|_2 \leq c T^{1/2} \sum_{n=k+1}^{\infty} \|\Gamma_{yn}^*\|_2 \leq c T^{1/2} \sum_{n=k+1}^{\infty} \rho^n \leq c T^{1/2} \rho^k \rightarrow 0.$$

■

## 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 four interest rates of different maturities; the federal funds rate  $i_{ff}$ , and the treasury bills rates for maturity 6 months, 3 years and 10 years, denoted  $i_{6m}$ ,  $i_{3y}$ ,  $i_{10y}$  respectively. To obtain more straightforward results on weak exogeneity the analysis is based on an equivalent transformation into two spreads in the short end of the term structure and two long interest rates.

The baseline VAR model is with two lags and a constant term restricted to the cointegration space.

$$\Delta X_t = \alpha(\beta' X_{t-1} + \mu') + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad (28)$$

where  $X_t = [s_{ff6m}, s_{6m3y}, i_{3y}, i_{10y}]$ ,  $s_{ff6m} = i_{ff} - i_{6m}$ , and  $s_{6m3y} = i_{6m} - i_{3y}$ .

Empirical applications often start with an analysis of a smaller system, which is then augmented with some new variables potentially considered important. This is also the procedure here.

In the small system  $X_t = [s_{ff6m}, s_{6m3y}, i_{3y}]$ , i.e. the 10 year rate is left out. The same model specification can be used to describe the variation of the data. The trace test and the roots of the characteristic polynomial suggest one cointegration relation and, hence, two stochastic trends. The estimates of  $\beta_1$  and  $\alpha_1$  are reported in Table 1.

The small system results				
	$s_{ff6m}$	$s_{6m3y}$	$i_{3y}$	$constant$
$\beta'_1$	1	-0.15 [-3.40]	0.04 [2.66]	0.00 [0.77]
$\alpha'_1$	-0.29 [-5.80]	0.01 [0.24]	0.04 [0.48]	

Table 1: The cointegration estimates in the small model.

The cointegration relation is a combination of the two spreads and a small level effect from the 3 year rate.

The  $\alpha$  coefficients suggest that only the shortest spread is significantly adjusting. This is confirmed by the joint test of weak exogeneity of  $s_{6m3y}$  and  $i_{3y}$  (p-value 0.79) which shows that  $s_{6m3y}$  and  $i_{3y}$  can be considered weakly exogenous. Thus, the two stochastic trends can be associated with shocks to the level of the longest interest rate and the spread between the 6 month and the 3 year rate, i.e. the term structure of interest rates seems to be driven by the shocks to a level and a slope component, similar to what is reported in Giese (2008).

The system is now enlarged with the 10 year bond rate. The trace test and the roots of the characteristic polynomials suggest that the rank is two and, hence, that two stochastic trends are also driving the large system. Table 2 reports the  $\beta$  and  $\alpha$  estimates of the two just identified cointegration relations. The first cointegration relation is identified by the zero restriction on the 10 year bond rate and thus corresponds to the relation in the small system. It is notable that the 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.  $\{(i_{3m} - i_{3y}) - (i_{3y} - i_{10y})\} \sim I(0)$ .

The large system results					
	$s_{ff6m}$	$s_{6m3y}$	$i_{3y}$	$i_{10y}$	$constant$
$\beta'_1$	1	-0.16 [-3.64]	0.05 [3.14]	0.00	0.00 [0.43]
$\beta'_2$	0.00	1.00	-0.99 [-8.82]	1.00 [7.49]	-0.00 [-1.23]
$\alpha'_1$	-0.29 [-5.47]	0.08 [1.58]	0.00 [0.02]	0.07 [0.92]	
$\alpha'_2$	-0.00 [-0.13]	-0.13 [-4.32]	0.06 [1.31]	0.03 [0.71]	

Table 2: The cointegration estimates in the large system.

The joint test of weak exogeneity of  $i_{3y}$  and  $i_{10y}$  (p-value 0.15) shows that the two long rates can be considered weakly exogenous, implying that their cumulated shocks

define the two common trends. This is consistent with the estimated  $\alpha$  coefficients that are insignificant for the two longest rates. Thus, the results support an interpretation of the term structure as being described by a nonstationary level and slope effect and a stationary curvature; see, Giese (2008).

In the small model, we concluded that it was the shocks to  $i_{3y}$  and  $s_{6m3y}$  that drive the system, whereas by adding the 10 year long-term rate to the small system we now conclude that 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 component of a variable is asymptotically the same, independently of the dimension of the system, the two stochastic trends estimated from the small and the large model, respectively, have to be able to replicate this realized random walk component. Thus, what is (asymptotically) invariant is (the space spanned by) the random walks, but not an interpretation in terms of a structural shock with a given label. Only when the information set is sufficiently large so that adding more variables does not change the definition of an exogenous shock, is it possible to discuss invariance of labels. Thus, the common trends are invariant, but their interpretation depends on the information set.

To illustrate how closely the two stochastic trends replicate the realized random walk components of each variable, the left panel of Figure 1, plots each of the four variables against their random walk component as determined in the large system. As expected, the exogenous long-term interest rates,  $i_{3y}$  and  $i_{10y}$ , are very close to their random walk component, whereas this is less so for the 3 month - 3 year spread,  $s_{3m3y}$ , and the shortest spread,  $s_{ff3m}$ , appears to be dominated by short run variation. In all cases the random walk captures the long swings of the variables.

Next, the right hand panel in Figure 1 does the same for the small model. The estimated random walk component of each variable looks similar to the one obtained from the large system. To get a more precise picture of how close they are, Figure 2 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 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. Saikkonen, P. (1992) Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* 8, 1–27.
6. Saikkonen, P. and Lütkepohl. H. (1996) Infinite order cointegrated vector autoregressive processes. Estimation and inference, *Econometric Theory* 12, 814–844.

**Comparison of variables with random walks**

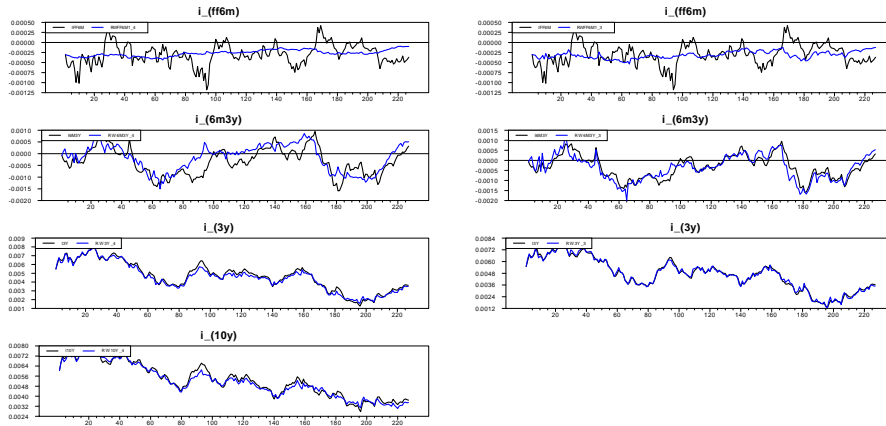


Figure 1: The plots shows the variables  $i_{ff6m}$ ,  $i_{6m3y}$ ,  $i_{3y}$ ,  $i_{10y}$  compared to the their random walk component. In the left panel the random walk is constructed from the large system and in the right hand panel from the small system.

**Comparison of random walks**

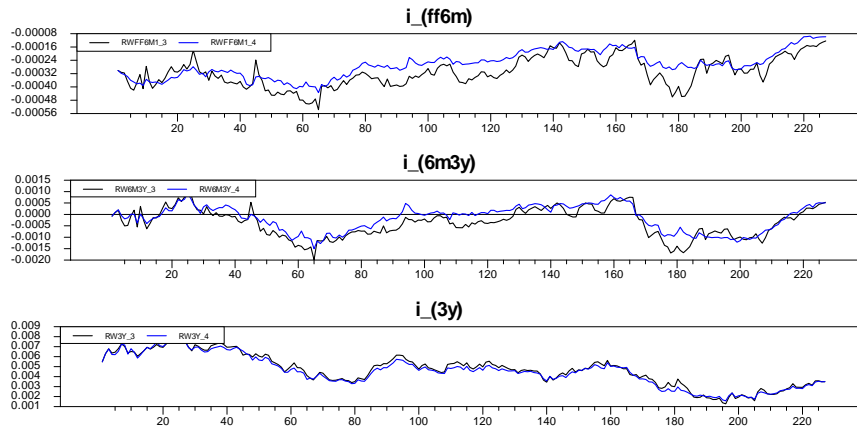


Figure 2: The plots compare the random walk component of each of the variables  $i_{ff6m}$ ,  $i_{6m3y}$ ,  $i_{3y}$  estimated from from the large and the small system. The finding that the random walk components are roughly the same in the two systems, which illustrates the invariance shown in the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