

Article

Cointegration between Trends and Their Estimators in State Space Models and Cointegrated Vector Autoregressive Models

Søren Johansen * and Morten Nyboe Tabor

Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, 1353 Copenhagen K, Denmark; morten.nyboe.tabor@econ.ku.dk

* Correspondence: soren.johansen@econ.ku.dk; Tel.: +45-35-323-071

Academic Editors: Katarina Juselius and Marc S. Paoletta

Received: 1 March 2017; Accepted: 17 August 2017; Published: 22 August 2017

Abstract: A state space model with an unobserved multivariate random walk and a linear observation equation is studied. The purpose is to find out when the extracted trend cointegrates with its estimator, in the sense that a linear combination is asymptotically stationary. It is found that this result holds for the linear combination of the trend that appears in the observation equation. If identifying restrictions are imposed on either the trend or its coefficients in the linear observation equation, it is shown that there is cointegration between the identified trend and its estimator, if and only if the estimators of the coefficients in the observation equations are consistent at a faster rate than the square root of sample size. The same results are found if the observations from the state space model are analysed using a cointegrated vector autoregressive model. The findings are illustrated by a small simulation study.

Keywords: cointegration of trends; state space models; cointegrated vector autoregressive models

JEL Classification: C32

1. Introduction and Summary

This paper is inspired by a study on long-run causality, see Hoover et al. (2014). Causality is usually studied for a sequence of multivariate i.i.d. variables using conditional independence, see Spirtes et al. (2000) or Pearl (2009). For stationary autoregressive processes, causality is discussed in terms of the variance of the shocks, that is, the variance of the i.i.d. error term. For nonstationary cointegrated variables, the common trends play an important role for long-run causality. In Hoover et al. (2014), the concept is formulated in terms of independent common trends and their causal impact coefficients on the nonstationary observations. Thus, the emphasis is on independent trends, and how they enter the observation equations, rather than on the variance of the measurement errors.

The trend is modelled as an m -dimensional Gaussian random walk, starting at T_0 ,

$$T_{t+1} = T_t + \eta_{t+1}, t = 0, \dots, n - 1, \quad (1)$$

where η_t are i.i.d. $N_m(0, \Omega_\eta)$, that is, Gaussian in m dimensions with mean zero and $m \times m$ variance $\Omega_\eta > 0$. This trend has an impact on future values of the p -dimensional observation y_t modelled by

$$y_{t+1} = BT_t + \varepsilon_{t+1}, t = 0, \dots, n - 1, \quad (2)$$

where ε_t are i.i.d. $N_p(0, \Omega_\varepsilon)$ and $\Omega_\varepsilon > 0$. It is also assumed that the ε_s and η_t are independent for all s and t . In the following the joint distribution of $T_1, \dots, T_n, y_1, \dots, y_n$ conditional on a given value of T_0 is considered.

The observations are collected in the matrices Y_n , $p \times n$, and ΔY_n , $p \times (n - 1)$, which are defined as

$$Y_n = (y_1, \dots, y_n), \text{ and } \Delta Y_n = (y_2 - y_1, \dots, y_n - y_{n-1}).$$

The processes y_t and T_t are obviously nonstationary, but the conditional distribution of Y_n given T_0 is well defined. We define

$$\begin{aligned} E_t T_t &= E(T_t | Y_t, T_0), \\ V_t &= \text{Var}_t(T_t) = \text{Var}(T_t | Y_t, T_0). \end{aligned}$$

Then the density of Y_n conditional on T_0 is given by the prediction error decomposition

$$p(Y_n | T_0) = p(y_1 | T_0) \prod_{t=1}^{n-1} p(y_{t+1} | Y_t, T_0),$$

where y_{t+1} given (Y_t, T_0) is p dimensional Gaussian with mean and variance

$$\begin{aligned} E_t y_{t+1} &= B E_t T_t, \\ \text{Var}_t(y_{t+1}) &= B V_t B' + \Omega_\varepsilon. \end{aligned}$$

In this model it is clear that y_t and T_t cointegrate, that is, $y_{t+1} - B T_{t+1} = \varepsilon_{t+1} - B \eta_{t+1}$ is stationary, and the same holds for T_t and the extracted trend $E_t T_t = E(T_t | y_1, \dots, y_t, T_0)$. Note that in the statistical model defined by (1) and (2) with parameters B, Ω_η , and Ω_ε , only the matrices $B \Omega_\eta B'$ and Ω_ε are identified because for any $m \times m$ matrix ζ of full rank, $B \zeta^{-1}$ and $\zeta \Omega_\eta \zeta'$ give the same likelihood, by redefining the trend as ζT_t .

Let $\hat{E}_t T_t$ be an estimator of $E_t T_t$. The paper investigates whether there is cointegration between $E_t T_t$ and $\hat{E}_t T_t$ given two different estimation methods: A simple cointegrating regression and the maximum likelihood estimator in an autoregressive representation of the state space model.

Section 2, on the probability analysis of the data generating process, formulates the model as a common trend state space model, and summarizes some results in three Lemmas. Lemma 1 contains the Kalman filter equations and the convergence of $\text{Var}(T_t | y_1, \dots, y_t, T_0)$, see Durbin and Koopman (2012), and shows how its limit can be calculated by solving an eigenvalue problem. Lemma 1 also shows how y_t can be represented in terms of its prediction errors $v_j = y_j - E_{j-1} y_j$, $j = 1, \dots, t$. This result is used in Lemma 2 to represent y_t in steady state as an infinite order cointegrated vector autoregressive model, see (Harvey 2006, p. 373). Section 3 discusses the statistical analysis of the data and the identification of the trends and their loadings. Two examples are discussed. In the first example, only B is restricted and the trends are allowed to be correlated. In the second example, B is restricted but the trends are uncorrelated, so that also the variance matrix is restricted. Lemma 3 analyses the data from (1) and (2) using a simple cointegrating regression, see Harvey and Koopman (1997), and shows that the estimator of the coefficient B suitably normalized is n -consistent.

Section 4 shows in Theorem 1 that the spread between $B E_t T_t$ and its estimator $\hat{B} \hat{E}_t T_t$ is asymptotically stationary irrespective of the identification of B and T_t . Then Theorem 2 shows that the spread between $E_t T_t$ and its estimator $\hat{E}_t T_t$ is asymptotically stationary if and only if B has been identified so that the estimator of B is superconsistent, that is, consistent at a rate faster than $n^{1/2}$.

The findings are illustrated with a small simulation study in Section 5. Data are generated from (1) and (2) with $T_0 = 0$, and the observations are analysed using the cointegrating regression discussed in Lemma 3. If the trends and their coefficients are identified by the trends being independent, the trend extracted by the state space model does not cointegrate with its estimator. If, however, the trends are identified by restrictions on the coefficients alone, they do cointegrate.

2. Probability Analysis of the Data Generating Model

This section contains first two examples, which illustrate the problem to be solved. Then a special parametrization of the common trends model is defined and some, mostly known, results are given in Lemmas 1 concerning the Kalman filter recursions. Lemma 2 is about the representation of the steady state solution as an autoregressive process. All proofs are given in the Appendix.

2.1. Two Examples

Two examples are given which illustrate the problem investigated. The examples are analysed further by a simulation study in Section 5.

Example 1. In the first example the two random walks T_{1t} and T_{2t} are allowed to be dependent, so Ω_η is unrestricted, and identifying restrictions are imposed only on their coefficients B . The equations are

$$\begin{aligned} y_{1,t+1} &= T_{1t} + \varepsilon_{1,t+1}, \\ y_{2,t+1} &= T_{2t} + \varepsilon_{2,t+1}, \\ y_{3,t+1} &= b_{31}T_{1t} + b_{32}T_{2t} + \varepsilon_{3,t+1}. \end{aligned} \quad (3)$$

for $t = 0, \dots, n - 1$. Thus, $y_t = (y_{1t}, y_{2t}, y_{3t})'$, $T_t = (T_{1t}, T_{2t})'$, and

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ b_{31} & b_{32} \end{pmatrix}. \quad (4)$$

Moreover, $\Omega_\eta > 0$ is 2×2 , $\Omega_\varepsilon > 0$ is 3×3 , and both are unrestricted positive definite. Simulations indicate that $E_t y_{t+1} - \hat{E}_t y_{t+1} = BE_t T_t - \hat{B} \hat{E}_t T_t$ is stationary, and this obviously implies that the same holds for the first two coordinates $E_t T_{1t} - \hat{E}_t T_{1t}$ and $E_t T_{2t} - \hat{E}_t T_{2t}$. ■

Example 2. The second example concerns two independent random walks T_{1t} and T_{2t} , and the three observation equations

$$\begin{aligned} y_{1,t+1} &= T_{1t} + \varepsilon_{1,t+1}, \\ y_{2,t+1} &= b_{21}T_{1t} + T_{2t} + \varepsilon_{2,t+1}, \\ y_{3,t+1} &= b_{31}T_{1t} + b_{32}T_{2t} + \varepsilon_{3,t+1}. \end{aligned} \quad (5)$$

In this example

$$B = \begin{pmatrix} 1 & 0 \\ b_{21} & 1 \\ b_{31} & b_{32} \end{pmatrix}, \Omega_\eta = \text{diag}(\sigma_1^2, \sigma_2^2), \quad (6)$$

and $\Omega_\varepsilon > 0$ is 3×3 and unrestricted positive definite. Thus the nonstationarity is caused by two independent trends. The first, T_{1t} , is the cause of the nonstationarity of y_{1t} , whereas both trends are causes of the nonstationarity of (y_{2t}, y_{3t}) . From the first equation it is seen that y_{1t} and T_{1t} cointegrate. It is to be expected that also the extracted trend $E_t T_{1t}$ cointegrates with T_{1t} , and also that $E_t T_{1t}$ cointegrates with its estimator $\hat{E}_t T_{1t}$. This is all supported by the simulations. Similarly, it turns out that

$$E_t y_{2,t+1} - \hat{E}_t y_{2,t+1} = b_{21}E_t T_{1t} - \hat{b}_{21}\hat{E}_t T_{1t} + E_t T_{2t} - \hat{E}_t T_{2t},$$

is asymptotically stationary. In this case, however, $E_t T_{2t} - \hat{E}_t T_{2t}$ is not asymptotically stationary, and the paper provides an answer to why this is the case. ■

The problem to be solved is why in the first example cointegration was found between the extracted trends and their estimators, and in the second example they do not cointegrate. The solution to the problem is that it depends on the way the trends and their coefficients are identified. For

some identification schemes the estimator of B is n -consistent, and then stationarity of $E_t T_t - \hat{E}_t T_t$ can be proved. But if identification is achieved by imposing restrictions also on the covariance of the trends, as in Example 2, then the estimator for B is only $n^{1/2}$ -consistent, and that is not enough to get asymptotic stationarity of $E_t T_t - \hat{E}_t T_t$.

2.2. Formulation of the Model as a Common Trend State Space Model

The common trend state space model with constant coefficients is defined by

$$\begin{aligned} \alpha_{t+1} &= \alpha_t + \eta_t, \\ y_t &= B\alpha_t + \varepsilon_t, \end{aligned} \tag{7}$$

$t = 1, \dots, n$, see Durbin and Koopman (2012) or Harvey (1989), with initial state α_1 . Here α_t is the unobserved m -dimensional state variable and y_t the p -dimensional observation and B is $p \times m$ of rank $m < p$. The errors ε_t and η_t are as specified in the discussion of the model given by (1) and (2).

Defining $T_t = \alpha_{t+1}, t = 0, \dots, n$, gives the model (1) and (2). Note that in this notation $E_t T_t = E_t \alpha_{t+1}$ is the predicted value of the trend α_{t+1} , which means that it is easy to formulate the Kalman filter.

The Kalman filter calculates the prediction $a_{t+1} = E_t \alpha_{t+1}$ and its conditional variance $P_{t+1} = \text{Var}_t(\alpha_{t+1})$ by the equations

$$a_{t+1} = a_t + P_t B' (B P_t B' + \Omega_\varepsilon)^{-1} (y_t - E_{t-1}(y_t)), \tag{8}$$

$$P_{t+1} = P_t + \Omega_\eta - P_t B' (B P_t B' + \Omega_\varepsilon)^{-1} B P_t, \tag{9}$$

starting with $a_1 = \alpha_1$ and $P_1 = 0$.

The recursions (8) and (9) become

$$E_{t+1} T_{t+1} = E_t T_t + K'_t (y_{t+1} - E_t y_{t+1}), \tag{10}$$

$$V_{t+1} = \Omega_\eta + V_t - K'_t B V_t, \tag{11}$$

$t = 0, \dots, n - 1$ starting with $E_1 T_1 = T_0$ and $V_1 = \Omega_\eta$, and defining the Kalman gain

$$K'_t = V_t B' (B V_t B' + \Omega_\varepsilon)^{-1}. \tag{12}$$

Lemma 1 contains the result that V_{t+1} converges for $t \rightarrow \infty$ to a finite limit V , which can be calculated by solving an eigenvalue problem. Equation (11) is an algebraic Ricatti equation, see Chan et al. (1984), where the convergence result can be found. The recursion (10) is used to represent y_{t+1} in terms of its cumulated prediction errors $v_{t+1} = y_{t+1} - E_t y_{t+1}$, as noted by Harvey (2006, Section 7.3.2).

Lemma 1. Let $V_t = \text{Var}(T_t | Y_t)$ and $E_t T_t = E(T_t | Y_t)$.

(a) The recursion for V_t , (11), can be expressed as

$$V_{t+1} = \Omega_\eta + V_t - V_t (V_t + \Omega_B)^{-1} V_t \rightarrow V, t \rightarrow \infty, \tag{13}$$

where $\Omega_B = \text{Var}(\bar{B}' \varepsilon_t | B' \varepsilon_t)$ for $\bar{B} = B(B'B)^{-1}$. Moreover,

$$I_m - K'_t B = I_m - V_t B' (B V_t B' + \Omega_\varepsilon)^{-1} B \rightarrow I_m - K' B = \Omega_B (V + \Omega_B)^{-1}, t \rightarrow \infty, \tag{14}$$

which has positive eigenvalues less than one, such that $I_m - K' B$ is a contraction, that is, $(I_m - K' B)^n \rightarrow 0, n \rightarrow \infty$.

(b) The limit of V_t can be found by solving the eigenvalue problem

$$|\lambda \Omega_B - \Omega_\eta| = 0,$$

for eigenvectors W and eigenvalues $(\lambda_1, \dots, \lambda_m)$, such that $W'\Omega_B W = I_m$ and $W'\Omega_\eta W = \text{diag}(\lambda_1, \dots, \lambda_m)$. Hence, $W'VW = \text{diag}(\tau_1, \dots, \tau_m)$ for

$$\tau_i = \frac{1}{2}\{\lambda_i + (\lambda_i^2 + 4\lambda_i)^{1/2}\}. \tag{15}$$

(c) Finally, using the prediction error, $v_{t+1} = y_{t+1} - E_t y_{t+1}$, it is found from (10) that

$$E_t T_t = T_0 + \sum_{j=1}^t K'_{j-1} v_j, \text{ and } y_{t+1} = v_{t+1} + B(T_0 + \sum_{j=1}^t K'_{j-1} v_j). \tag{16}$$

The prediction errors are independent Gaussian with mean zero and variances

$$\text{Var}(v_{t+1}) = \text{Var}_t(y_{t+1}) = \text{Var}_t(BT_t + \varepsilon_{t+1}) = BV_t B' + \Omega_\varepsilon \rightarrow BVB' + \Omega_\varepsilon, t \rightarrow \infty,$$

such that in steady state the prediction errors are i.i.d. $N_p(0, BVB' + \Omega_\varepsilon)$, and (16) shows that y_t is approximately an $AR(\infty)$ process, for which the reduced form autoregressive representation can be found, see (Harvey 2006, Section 7.3.2).

Lemma 2. *If the system (7) is in steady state, prediction errors v_t are i.i.d. $N(0, BVB' + \Omega_\varepsilon)$ and*

$$\Delta y_t = \Delta v_t + BK'v_{t-1}. \tag{17}$$

Applying the Granger Representation Theorem, y_t is given by

$$\Delta y_t = \alpha\beta'y_{t-1} + \sum_{i=1}^{\infty} \Gamma_i \Delta y_{t-i} + v_t. \tag{18}$$

Here $\alpha = -K_\perp(B'_\perp K_\perp)^{-1}$ and $\beta = B_\perp$.

2.3. Cointegration among the Observations and Trends

In model (1) and (2), the equation $y_{t+1} = BT_t + \varepsilon_{t+1}$ shows that y_t and T_t are cointegrated. It also holds that $T_t - E_t T_t$ is asymptotically stationary because

$$v_{t+1} = y_{t+1} - E_t y_{t+1} = BT_t + \varepsilon_{t+1} - BE_t T_t,$$

which shows that $B(T_t - E_t T_t) = v_{t+1} - \varepsilon_{t+1}$ is asymptotically stationary. Multiplying by $\bar{B}' = (B'B)^{-1}B'$, the same holds for $T_t - E_t T_t$.

In model (18) the extracted trend is

$$T_t^* = \alpha'_\perp \sum_{i=1}^t v_i = K' \sum_{i=1}^t v_i,$$

and (16) shows that in steady state, $y_{t+1} - BT_t^* = v_{t+1} + BT_0$ is stationary, so that y_t cointegrates with T_t^* . Thus, the process y_t and the trends T_t, T_t^* , and $E_t T_t$ all cointegrate, in the sense that suitable linear combinations are asymptotically stationary. The next section investigates when similar results hold for the estimated trends.

3. Statistical Analysis of the Data

In this section it is shown how the parameters of (7) can be estimated from the CVAR (18) using results of Saikkonen (1992) and Saikkonen and Lutkepohl (1996), or using a simple cointegrating regression, see (Harvey and Koopman 1997, p. 276) as discussed in Lemma 3. For both the state space model (1)–(2) and for the CVAR in (18) there is an identification problem between T_t and its coefficient B ,

or between β_{\perp} and T_t^* , because for any $m \times m$ matrix ζ of full rank, one can use $B\zeta^{-1}$ as parameter and ζT_t as trend and $\zeta\Omega_{\eta}\zeta'$ as variance, and similarly for β_{\perp} and T_t^* . In order to estimate B , T , and Ω_{η} , it is therefore necessary to impose identifying restrictions. Examples of such identification are given next.

Identification 1. Because B has rank m , the rows can be permuted such that $B' = (B'_1, B'_2)$, where B_1 is $m \times m$ and has full rank. Then the parameters and trend are redefined as

$$B^{\dagger} = \begin{pmatrix} I_m \\ B_2 B_1^{-1} \end{pmatrix} = \begin{pmatrix} I_m \\ \gamma' \end{pmatrix}, \Omega_{\eta}^{\dagger} = B_1 \Omega_{\eta} B_1', T_t^{\dagger} = B_1 T_t. \tag{19}$$

Note that $B^{\dagger} T_t^{\dagger} = B T_t$ and $B^{\dagger} \Omega_{\eta}^{\dagger} B^{\dagger'} = B \Omega_{\eta} B'$. This parametrization is the simplest which separates parameters that are n -consistently estimated, γ , from those that are $n^{1/2}$ -consistently estimated, $(\Omega_{\eta}, \Omega_{\epsilon})$, see Lemma 3. Note that the (correlated) trends are redefined by choosing T_{1t} as the trend in y_{1t} , then T_{2t} as the trend in y_{2t} , as in Example 1.

A more general parametrization, which also gives n -consistency, is defined, as in simultaneous equations, by imposing linear restrictions on each of the m columns and require the identification condition to hold, see Fisher (1966). ■

Identification 2. The normalization with diagonality of Ω_{η}^{\dagger} is part of the next identification, because this is the assumption in the discussion of long-run causality. Let $\Omega_{\eta}^{\dagger} = C_{\eta} \text{diag}(\sigma_1^2, \dots, \sigma_m^2) C_{\eta}'$ be a Cholesky decomposition of Ω_{η} . That is, C_{η} is lower-triangular with one in the diagonal, corresponding to an ordering of the variables. Using this decomposition the new parameters and the trend are

$$B^{\#} = \begin{pmatrix} C_{\eta} \\ \gamma' C_{\eta} \end{pmatrix}, \Omega_{\eta}^{\#} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2), T_t^{\#} = C_{\eta}^{-1} T_t^{\dagger}, \tag{20}$$

such that $B^{\#} T_t^{\#} = B^{\dagger} T_t^{\dagger} = B T_t$ and $B^{\#} \Omega_{\eta}^{\#} B^{\#'} = B^{\dagger} \Omega_{\eta}^{\dagger} B^{\dagger'} = B \Omega_{\eta} B'$.

Identification of the trends is achieved in this case by defining the trends to be independent and constrain how they load into the observations. In Example 2, T_{1t} was defined as the trend in y_{1t} , and T_{2t} as the trend in y_{2t} , but orthogonalized on T_{1t} , such that the trend in y_{2t} is a combination of T_{1t} and T_{2t} . ■

3.1. The Vector Autoregressive Model

When the process is in steady state, the infinite order CVAR representation is given in (18). The model is approximated by a sequence of finite lag models, depending on sample size n ,

$$\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{k_n} \Gamma_i \Delta y_{t-i} + v_t,$$

where the lag length k_n is chosen to depend on n such that k_n increases to infinity with n , but so slowly that k_n^3/n converges to zero. Thus one can choose for instance $k_n = n^{1/3} / \log n$ or $k_n = n^{1/3-\epsilon}$, for some $\epsilon > 0$. With this choice of asymptotics, the parameters α , β , $\Gamma = I_p - \sum_{i=1}^{\infty} \Gamma_i$, $\Sigma = \text{Var}(v_t)$, and the residuals, v_t , can be estimated consistently, see Johansen and Juselius (2014) for this application of the results of Saikkonen and Lutkepohl (1996).

This defines for each sample size consistent estimators $\check{\alpha}$, $\check{\beta}$, $\check{\Gamma}$ and $\check{\Sigma}$, as well residuals \check{v}_t . In particular the estimator of the common trend is $\check{T}_t^* = \check{\alpha}'_{\perp} \sum_{i=1}^t \check{v}_i$. Thus, $\check{\alpha} \check{\beta}' \xrightarrow{P} \alpha \beta'$,

$\check{C} = \check{\beta}_\perp (\check{\alpha}'_\perp \check{\Gamma} \check{\beta}_\perp)^{-1} \check{\alpha}'_\perp \xrightarrow{P} C = BK'$ and $\check{\Sigma} \xrightarrow{P} \Sigma = BVB' + \Omega_\epsilon$. If β_\perp is identified as $(I_m, \gamma)'$, then $\check{B} = \check{\beta}_\perp \xrightarrow{P} \beta_\perp$. In steady state, the relations

$$\begin{aligned} \Omega_\eta &= VB'(BVB' + \Omega_\epsilon)^{-1}B'V = VB'\Sigma^{-1}B'V, \\ C &= BK' = BVB'(BVB' + \Omega_\epsilon)^{-1} = BVB'\Sigma^{-1}, \end{aligned}$$

hold, see (11) and Lemma 2. It follows that

$$\check{B}\check{\Omega}_\eta\check{B}' = \check{C}\check{\Sigma}\check{C}' \xrightarrow{P} B\Omega_\eta B', \text{ and } \check{\Omega}_\eta = (\check{B}'\check{B})^{-1}\check{B}'\check{C}\check{\Sigma}\check{C}'\check{B}(\check{B}'\check{B})^{-1} \xrightarrow{P} \Omega_\eta.$$

Finally, an estimator for Ω_ϵ can be found as

$$\check{\Omega}_\epsilon = \check{\Sigma} - \frac{1}{2}(\check{C}\check{\Sigma} + \check{\Sigma}\check{C}') \xrightarrow{P} BVB' + \Omega_\epsilon - \frac{1}{2}(BVB' + BVB') = \Omega_\epsilon.$$

Note that $C\Sigma$ is not a symmetric matrix in model (18), but its estimator converges in probability towards the symmetric matrix BVB' .

3.2. The State Space Model

The state space model is defined by (1) and (2). It can be analysed using the Kalman filter to calculate the diffuse likelihood function, see Durbin and Koopman (2012), and an optimizing algorithm can be used to find the maximum likelihood estimator for the parameters Ω_η , Ω_ϵ , and B , once B is identified.

In this paper, an estimator is used which is simpler to analyse and which gives an n -consistent estimator for B suitably normalized, see (Harvey and Koopman 1997, p. 276).

The estimators are functions of ΔY_n and $B'_\perp Y_n$, and therefore do not involve the initial value T_0 . Irrespective of the identification, the relations

$$Var(\Delta y_t) = B\Omega_\eta B' + 2\Omega_\epsilon, \tag{21}$$

$$Cov(\Delta y_t, \Delta y_{t+1}) = -\Omega_\epsilon, \tag{22}$$

hold, and they give rise to two moment estimators, which determine Ω_η and Ω_ϵ , once B has been identified and estimated.

Consider the identified parametrization (19), where $B = (I_m, \gamma)'$, and take $B_\perp = (\gamma', -I_{p-m})'$. Then define $z_{1t} = (y_{1t}, \dots, y_{mt})'$ and $z_{2t} = (y_{m+1,t}, \dots, y_{pt})'$, such that $y_t = (z'_{1t}, z'_{2t})'$ and $B'_\perp y_t = \gamma' z_{1t} - z_{2t} = B'_\perp \epsilon_t$, that is,

$$z_{2t} = \gamma' z_{1t} - B'_\perp \epsilon_t. \tag{23}$$

This equation defines the regression estimator $\hat{\gamma}_{reg}$:

$$\hat{\gamma}_{reg} = \left(\sum_{t=0}^{n-1} z_{1t} z'_{1t} \right)^{-1} \sum_{t=0}^{n-1} z_{1t} z'_{2t} = \gamma - \left(\sum_{t=0}^{n-1} z_{1t} z'_{1t} \right)^{-1} \sum_{t=0}^{n-1} z_{1t} \epsilon'_t B_\perp. \tag{24}$$

To describe the asymptotic properties of $\hat{\gamma}_{reg}$, two Brownian motions are introduced

$$n^{-1/2} \sum_{t=1}^{[nu]} \epsilon_t \xrightarrow{D} W_\epsilon(u) \text{ and } n^{-1/2} \sum_{t=1}^{[nu]} \eta_t \xrightarrow{D} W_\eta(u). \tag{25}$$

Lemma 3. Let the data be generated by the state space model (1) and (2).

(a) From (21) and (22) it follows that

$$\begin{aligned} S_{n1} &= n^{-1} \sum_{i=1}^n \Delta y_i \Delta y'_i \xrightarrow{P} B\Omega_\eta B' + 2\Omega_\epsilon, \\ S_{n2} &= n^{-1} \sum_{i=2}^n (\Delta y_i \Delta y'_{i-1} + \Delta y_{i-1} \Delta y'_i) \xrightarrow{P} -2\Omega_\epsilon, \end{aligned} \tag{26}$$

define $n^{1/2}$ -consistent asymptotically Gaussian estimators for $B\Omega_\eta B'$ and Ω_ε , irrespective of the identification of B .

(b) If B and B_\perp are identified as $B = (I_m, \gamma)'$, $B'_\perp = (\gamma', -I_{p-m})$, and Ω_η is adjusted accordingly, then $\hat{\gamma}_{reg}$ in (24) is n -consistent with asymptotic Mixed Gaussian distribution

$$n(\hat{\gamma}_{reg} - \gamma) = -n(\hat{B} - B)'B_\perp = nB'(\hat{B}_\perp - B_\perp) \tag{27}$$

$$\xrightarrow{D} -\left(\int_0^1 W_\eta W_\eta' du\right)^{-1} \int_0^1 W_\eta(dW_\varepsilon)'B_\perp.$$

(c) If B is identified as $B = (C'_\eta, C'_\eta \gamma)'$, $B'_\perp = (\gamma', -I_{p-m})$, and $\Omega_\eta = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$, then $\hat{B} - B = O_p(n^{-1/2})$, but (27) still holds for $-n(\hat{B} - B)'B_\perp = \hat{C}'_\eta(\hat{\gamma}_{reg} - \gamma)$, so that some linear combinations of \hat{B} are n -consistent.

Note that the parameters $B = (I_m, \gamma)'$, Ω_η , and Ω_ε can be estimated consistently from (24) and (26) by

$$\hat{B} = \begin{pmatrix} I_m \\ \hat{\gamma}'_{reg} \end{pmatrix}, \hat{\Omega}_\varepsilon = -\frac{1}{2}S_{n2}, \text{ and } \hat{\Omega}_\eta = (\hat{B}'\hat{B})^{-1}\hat{B}'(S_{n1} + S_{n2})\hat{B}(\hat{B}'\hat{B})^{-1}. \tag{28}$$

In the simulations of Examples 1 and 2 the initial value is $T_0 = 0$, so the Kalman filter with $T_0 = 0$ is used to calculate the extracted trend $E_t T_t$ using observations and known parameters. Similarly the estimator of the extracted trend $\hat{E}_t T_t$ is calculated using observations and estimated parameters based on Lemma 3. The next section investigates to what extent these estimated trends cointegrate with the extracted trends, and if they cointegrate with each other.

4. Cointegration between Trends and Their Estimators

This section gives the main results in two theorems with proofs in the Appendix. In Theorem 1 it is shown, using the state space model to extract the trends and the estimator from Lemma 3, that $BE_t T_t - \hat{B}\hat{E}_t T_t$ is asymptotically stationary. For the CVAR model it holds that $BT_t^* - \check{B}\check{T}_t^* \xrightarrow{P} 0$, such that this spread is asymptotically stationary. Finally, the estimated trends in the two models are compared, and it is shown that $\hat{B}\hat{E}_t T_t - \check{B}\check{T}_t^*$ is asymptotically stationary. The conclusion is that in terms of cointegration of the trends and their estimators, it does not matter which model is used to extract the trends, as long as the focus is on the identified trends BT_t and BT_t^* .

Theorem 1. Let y_t and T_t be generated by the DGP given in (1) and (2).

(a) If the state space model is used to extract the trends, and Lemma 3 is used for estimation, then $BE_t T_t - \hat{B}\hat{E}_t T_t$ is asymptotically stationary.

(b) If the vector autoregressive model is used to extract the trends and for estimation, then $BT_t^* - \check{B}\check{T}_t^* \xrightarrow{P} 0$.

(c) Under assumptions of (a) and (b), it holds that $\hat{B}\hat{E}_t T_t - \check{B}\check{T}_t^*$ is asymptotically stationary.

In Theorem 2 a necessary and sufficient condition for asymptotic stationarity of $T_t^* - \check{T}_t^*$, $E_t T_t - \hat{E}_t T_t$, and $\hat{E}_t T_t - \check{T}_t^*$ is given.

Theorem 2. In the notation of Theorem 1, any of the spreads $T_t^* - \check{T}_t^*$, $E_t T_t - \hat{E}_t T_t$ or $\hat{E}_t T_t - \check{T}_t^*$ is asymptotically stationary if and only if B and the trend are identified such that the corresponding estimator for B satisfies $n^{1/2}(\hat{B} - B) = o_p(1)$ and $n^{1/2}(\check{B} - B) = o_p(1)$.

The missing cointegration between $E_t T_t$ and $\hat{E}_t T_t$, say, can be explained in terms of the identity

$$\hat{B}(E_t T_t - \hat{E}_t T_t) = (\hat{B} - B)E_t T_t + (BE_t T_t - \hat{B}\hat{E}_t T_t).$$

Here the second term, $BE_tT_t - \hat{B}\hat{E}_tT_t$, is asymptotically stationary by Theorem 1(a). But the first term, $(\hat{B} - B)T_t$, is not necessarily asymptotically stationary, because in general, that is, depending on the identification of the trend and B , it holds that $\hat{B} - B = O_p(n^{-1/2})$ and $E_tT_t = O_p(n^{1/2})$, see (16).

The parametrization $B = (I_m, \gamma)'$ ensures n -consistency of \hat{B} , so there is asymptotic stationarity of $T_t^* - \check{T}_t^*$, $E_tT_t - \hat{E}_tT_t$, and $\hat{E}_tT_t - \check{T}_t^*$ in this case. This is not so surprising because

$$BE_tT_t - \hat{B}\hat{E}_tT_t = \begin{pmatrix} E_tT_t - \hat{E}_tT_t \\ \gamma'E_tT_t - \hat{\gamma}'\hat{E}_tT_t \end{pmatrix},$$

is stationary. Another situation where the estimator for B is n -consistent is if $B = (B_1, \dots, B_m)$ satisfies linear restriction of the columns, $R_i'B_i = 0$, or equivalently $B_i = R_{i\perp}\phi_i$ for some ϕ_i , and the condition for identification is satisfied

$$\text{rank}\{R_i'(R_{1\perp}\phi_1, \dots, R_{m\perp}\phi_m)\} = r - 1, \text{ for } i = 1, \dots, m, \tag{29}$$

see Fisher (1966). For a just-identified system, one can still use $\hat{\gamma}_{reg}$, and then solve for the identified parameters. For overidentified systems, the parameters can be estimated by a nonlinear regression of z_{2t} on z_{1t} reflecting the overidentified parametrization. In either case the estimator is n -consistent such that $T_t^* - \check{T}_t^*$, $E_tT_t - \hat{E}_tT_t$, and $\hat{E}_tT_t - \check{T}_t^*$ are asymptotically stationary.

If the identification involves the variance Ω_η , however, the estimator of B is only $n^{1/2}$ -consistent, and hence no cointegration is found between the trend and estimated trend.

The analogy with the results for the CVAR, where β and α need to be identified, is that if β is identified using linear restrictions (29) then $\hat{\beta}$ is n -consistent, whereas if β is identified by restrictions on α then β is $n^{1/2}$ -consistent. An example of the latter is if β is identified as the first m rows of the matrix $\Pi = \alpha\beta'$, corresponding to $\alpha = (I_m, \phi)'$, then $\hat{\beta}$ is $n^{1/2}$ -consistent and asymptotically Gaussian, see (Johansen 2010, Section 4.3).

5. A Small Simulation Study

The two examples introduced in Section 2.1 are analysed by simulation. The equations are given in (5) and (3). Both examples have $p = 3$ and $m = 2$. The parameters B and Ω_η contain $6 + 3$ parameters, but the 3×3 matrix $B\Omega_\eta B'$ is of rank 2 and has only 5 estimable parameters. Thus, 4 restrictions must be imposed to identify the parameters. In both examples the Kalman filter with $T_0 = 0$ is used to extract the trends, and the cointegrating regression in Lemma 3 is used to estimate the parameters.

Example 1 continued. The parameter B is given in (4), and the parameters are just-identified. Now

$$E_tBT_{1t} - \hat{E}_t\hat{B}T_{1t} = \begin{pmatrix} E_tT_{1t} - \hat{E}_tT_{1t} \\ E_tT_{2t} - \hat{E}_tT_{2t} \\ b_{31}E_tT_{1t} + b_{32}E_tT_{2t} - \hat{b}_{31}\hat{E}_tT_{1t} - \hat{b}_{32}\hat{E}_tT_{2t} \end{pmatrix}. \tag{30}$$

As $E_tT_{1t} - \hat{E}_tT_{1t}$ and $E_tT_{2t} - \hat{E}_tT_{2t}$ are the first two rows of $E_tBT_{1t} - \hat{E}_t\hat{B}T_{1t}$ in (30), they are both asymptotically stationary by Theorem 1(a).

To illustrate the results, data are simulated with $n = 100$ observations starting with $T_0 = 0$ and parameter values $b_{31} = b_{32} = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = 0$, such that

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}, \Omega_\eta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{31}$$

The parameters are estimated by (28) and the estimates become $\hat{b}_{31} = 0.48$, $\hat{b}_{32} = 0.41$, $\hat{\sigma}_1^2 = 0.93$, $\hat{\sigma}_{12} = 0.26$, and $\hat{\sigma}_2^2 = 1.63$. The extracted and estimated trends are plotted in Figure 1. Panels a

and *b* show plots of $(E_t T_{1t}, \hat{E}_t T_{1t})$ and $(E_t T_{2t}, \hat{E}_t T_{2t})$, respectively, and it is seen that they co-move. In panels *c* and *d* the differences $\hat{E}_t T_{1t} - E_t T_{1t}$ and $\hat{E}_t T_{2t} - E_t T_{2t}$ both appear to be stationary in this parametrization of the model. ■

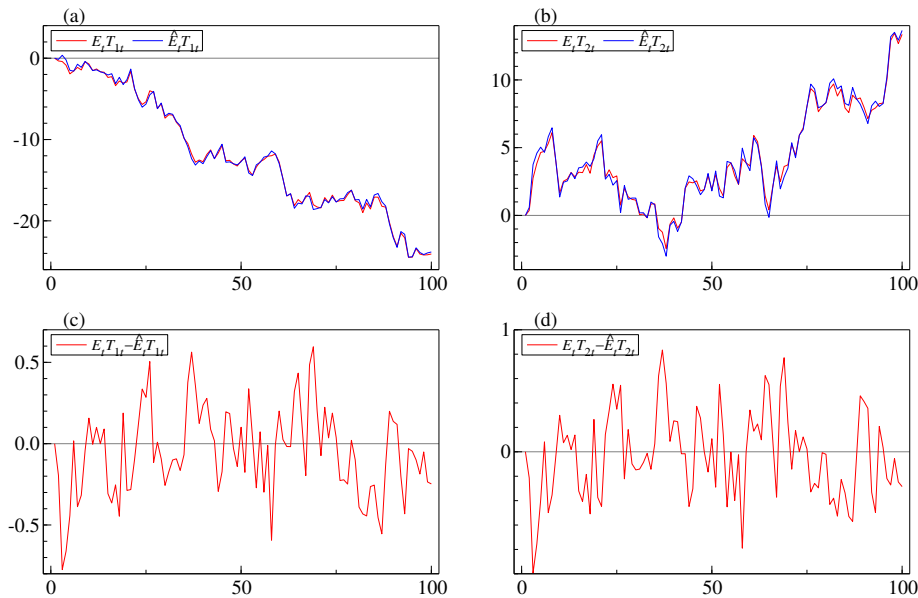


Figure 1. The figure shows the extracted and estimated trends for the simulated data in Example 1 with the identification in (19). Panels *a* and *b* show plots of $E_t T_{1t}$ and $\hat{E}_t T_{1t}$, and $E_t T_{2t}$ and $\hat{E}_t T_{2t}$, respectively. Note that in both cases, the processes seem to co-move. In panels *c* and *d*, $E_t T_{1t} - \hat{E}_t T_{1t}$ and $E_t T_{2t} - \hat{E}_t T_{2t}$ are plotted and appear stationary, because they are both recovered from $BE_t T_t - \hat{B}\hat{E}_t T_t$ as the first two coordinates, see (19).

Example 2 continued. The parameter *B* in this example is given in (6) such that

$$E_t B T_t - \hat{E}_t \hat{B} T_t = \begin{pmatrix} E_t T_{1t} - \hat{E}_t T_{1t} \\ b_{21} E_t T_{1t} + E_t T_{2t} - \hat{b}_{21} \hat{E}_t T_{1t} - \hat{E}_t T_{2t} \\ b_{31} E_t T_{1t} + b_{32} E_t T_{2t} - \hat{b}_{31} \hat{E}_t T_{1t} - \hat{b}_{32} \hat{E}_t T_{2t} \end{pmatrix}. \tag{32}$$

By the results in Theorem 1(a), all three rows are asymptotically stationary, in particular $E_t T_{1t} - \hat{E}_t T_{1t}$. Moreover, the second row of (32), $(b_{21} E_t T_{1t} - \hat{b}_{21} \hat{E}_t T_{1t}) + (E_t T_{2t} - \hat{E}_t T_{2t})$, is asymptotically stationary. Thus, asymptotic stationarity of $E_t T_{2t} - \hat{E}_t T_{2t}$ requires asymptotic stationarity of the term

$$b_{21} E_t T_{1t} - \hat{b}_{21} \hat{E}_t T_{1t} = (b_{21} - \hat{b}_{21}) E_t T_{1t} + \hat{b}_{21} (E_t T_{1t} - \hat{E}_t T_{1t}). \tag{33}$$

Here, the second term, $\hat{b}_{21} (E_t T_{1t} - \hat{E}_t T_{1t})$, is asymptotically stationary because $E_t T_{1t} - \hat{E}_t T_{1t}$ is. However, the first term, $(b_{21} - \hat{b}_{21}) E_t T_{1t}$, is not asymptotically stationary because \hat{b}_{21} is $n^{1/2}$ -consistent. In this case $n^{1/2}(b_{21} - \hat{b}_{21}) \xrightarrow{D} Z$, which has a Gaussian distribution, and $n^{-1/2} E_{[nu]} T_{1[nu]} \xrightarrow{D} W_{\eta_1}(u)$, where W_{η_1} is the Brownian motion generated by the sum of η_{1t} . It follows that their product

$$(b_{21} - \hat{b}_{21}) E_{[nu]} T_{1[nu]} = \{n^{1/2}(b_{21} - \hat{b}_{21})\} \{n^{-1/2} E_{[nu]} T_{1[nu]}\}$$

converges in distribution to the product of *Z* and $W_{\eta_1}(u)$, $n \rightarrow \infty$, and this limit is nonstationary. It follows that $E_t T_{2t} - \hat{E}_t T_{2t}$ is not asymptotically stationary for the identification in this example. This argument is a special case of the proof of Theorem 2.

To illustrate the results, data are simulated from the model with $n = 100$ observations starting with $T_0 = 0$ and parameter values $b_{21} = 0.0$, $b_{31} = b_{32} = 0.5$, and $\sigma_1^2 = \sigma_2^2 = 1$, which is identical to (31).

The model is written in the form (19) with a transformed B and Ω_η , as

$$B^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ b_{31} - b_{32}b_{21} & b_{32} \end{pmatrix}, \quad \Omega_\eta^\dagger = \begin{pmatrix} \sigma_1^2 & b_{21}\sigma_1^2 \\ b_{21}\sigma_1^2 & \sigma_2^2 + b_{21}^2\sigma_1^2 \end{pmatrix}.$$

The parameters are estimated as in Example 1 and we find $\hat{b}_{31} - \hat{b}_{32}\hat{b}_{21} = 0.48$, $\hat{b}_{32} = 0.41$, $\hat{\sigma}_1^2 = 0.93$, $\hat{b}_{21}\hat{\sigma}_{12} = 0.26$, and $\hat{\sigma}_2^2 + \hat{b}_{21}^2\hat{\sigma}_1^2 = 1.63$, which are solved for $\hat{b}_{21} = 0.28$, $\hat{b}_{31} = 0.59$, $\hat{b}_{32} = 0.41$, $\hat{\sigma}_1^2 = 0.93$, and $\hat{\sigma}_2^2 = 1.56$. The extracted and estimated trends are plotted in Figure 2. The panels *a* and *b* show plots of $(E_t T_{1t}, \hat{E}_t T_{1t})$ and $(E_t T_{2t}, \hat{E}_t T_{2t})$, respectively. It is seen that $E_t T_{1t}$ and $\hat{E}_t T_{1t}$ co-move, whereas $E_t T_{2t}$ and $\hat{E}_t T_{2t}$ do not co-move. In panels *c* and *d*, the differences $E_t T_{1t} - \hat{E}_t T_{1t}$ and $E_t T_{2t} - \hat{E}_t T_{2t}$ are plotted. Note that the first looks stationary, whereas the second is clearly nonstationary. When comparing with the plot of $E_t T_{1t}$ in panel *a*, it appears that the process $\hat{E}_t T_{1t}$ can explain the nonstationarity of $E_t T_{2t} - \hat{E}_t T_{2t}$. This is consistent with Equation (33) with $b_{21} = 0$ and $\hat{b}_{21} = 0.28$. In panel *d*, $E_t T_{2t} - \hat{E}_t T_{2t} - 0.28\hat{E}_t T_{1t}$ is plotted and it is indeed stationary. ■

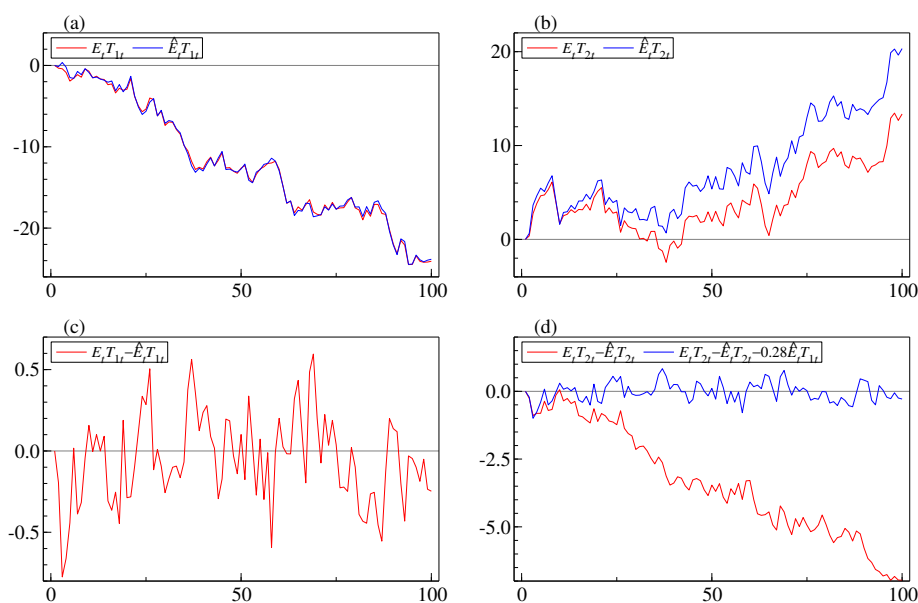


Figure 2. The figure shows the extracted and estimated trends for the simulated data in Example 2 with the identification in (20). Panels *a* and *b* show plots of $E_t T_{1t}$ and $\hat{E}_t T_{1t}$, and $E_t T_{2t}$ and $\hat{E}_t T_{2t}$, respectively. Note that $E_t T_{1t}$ and $\hat{E}_t T_{1t}$ seem to co-move, whereas $E_t T_{2t}$ and $\hat{E}_t T_{2t}$ do not. In panel *c*, $E_t T_{1t} - \hat{E}_t T_{1t}$ is plotted and appears stationary, but in panel *d* the spread $E_t T_{2t} - \hat{E}_t T_{2t}$ is nonstationary, whereas $E_t T_{2t} - \hat{E}_t T_{2t} - 0.28\hat{E}_t T_{1t}$ is stationary.

6. Conclusions

The paper analyses a sample of n observations from a common trend model, where the state is an unobserved multivariate random walk and the observation is a linear combination of the lagged state variable and a noise term. For such a model, the trends and their coefficients in the observation equation need to be identified before they can be estimated separately. The model leads naturally to cointegration between observations, trends, and the extracted trends. Using simulations it was discovered, that the extracted trends do not necessarily cointegrate with their estimators. This problem

is investigated, and it is found to be related to the identification of the trends and their coefficients in the observation equation. It is shown in Theorem 1, that provided only the linear combinations of the trends from the observation equation are considered, there is always cointegration between extracted trends and their estimators. If the trends and their coefficients are defined by identifying restrictions, the same result holds if and only if the estimated identified coefficients in the observation equation are consistent at a rate faster than $n^{1/2}$. For the causality study mentioned in the introduction, where the components of the unobserved trend are assumed independent, the result has the following implication: For the individual extracted trends to cointegrate with their estimators, overidentifying restrictions must be imposed on the trend’s causal impact coefficients on the observations, such that the estimators of these become super-consistent.

Acknowledgments: S.J. is grateful to CREATES—Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. M.N.T. is grateful to the Carlsberg Foundation (grant reference 2013_01_0972). We have benefitted from discussions with Siem Jan Koopman and Eric Hillebrand on state space models and thankfully acknowledge the insightful comments from two anonymous referees.

Author Contributions: S.J. has contributed most of the mathematical derivations. M.N.T. has performed the simulations and posed the problem to be solved.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Lemma 1 *Proof of (a):* Let $N = (\bar{B}, B_{\perp})$, $\bar{B} = B(B'B)^{-1}$, such that

$$\begin{aligned} K_t' B &= V_t B' [B V_t B' + \Omega_{\varepsilon}]^{-1} B = V_t B' N [(N' B V_t B' N + N' \Omega_{\varepsilon} N)]^{-1} N' B \\ &= V_t \begin{pmatrix} I_m \\ 0 \end{pmatrix}' \begin{pmatrix} V_t + \bar{B}' \Omega_{\varepsilon} \bar{B} & \bar{B}' \Omega_{\varepsilon} B_{\perp} \\ B_{\perp}' \Omega_{\varepsilon} \bar{B} & B_{\perp}' \Omega_{\varepsilon} B_{\perp} \end{pmatrix}^{-1} \begin{pmatrix} I_m \\ 0 \end{pmatrix} = V_t (V_t + \Omega_B)^{-1}, \end{aligned}$$

where

$$\Omega_B = \bar{B}' [\Omega_{\varepsilon} - \Omega_{\varepsilon} B_{\perp} (B_{\perp}' \Omega_{\varepsilon} B_{\perp})^{-1} B_{\perp}' \Omega_{\varepsilon}] \bar{B} = \text{Var}(\bar{B}' \varepsilon_t | B_{\perp}' \varepsilon_t).$$

This proves (13) and (14).

Proof of (b): If the recursion starts with $V_1 = \Omega_{\eta}$, then V_t can be diagonalized by W for all t and the limit satisfies $W' V W = \text{diag}(\tau_1, \dots, \tau_m)$, where

$$\tau_i = \lambda_i + \tau_i - \frac{\tau_i^2}{1 + \tau_i}.$$

This has solution given in (15).

Proof of (c): The first result follows by summation from the recursion for $E_t T_t$ in (10), and the second from $y_{t+1} = v_{t+1} + B E_t T_t$. ■

Proof of Lemma 2 The polynomial $\Phi(z) = I_p - z(I_p - BK')$ describes (17) as

$$(1 - L)y_t = \Phi(L)v_t.$$

Note that $\Phi(1) = BK'$ is singular and $d\Phi(z)/dz|_{z=1} = BK' - I_p = BVB'(BVB' + \Omega_{\varepsilon})^{-1} - I_p$, satisfies $B_{\perp}' (BK' - I_p) K_{\perp} = B_{\perp}' \Omega_{\varepsilon} B_{\perp}$ is nonsingular, where $K_{\perp} = (BVB' + \Omega_{\varepsilon}) B_{\perp}$. This means that the Granger Representation Theorem (Johansen 1996, Theorem 4.5) can be applied and gives the expansion (18) for $\alpha = -K_{\perp} (B_{\perp}' K_{\perp})^{-1}$ and $\beta = B_{\perp}'$. ■

Proof of Lemma 3 *Proof of (a):* Consider first the product moments (21) and (22). The result (26) follows from the Law of Large Numbers and the asymptotic Gaussian distribution of $\hat{\Omega}_{\varepsilon} = -\frac{1}{2} S_{n2}$ and $\hat{\Omega}_{\eta} = \hat{B}' (S_{1n} + S_{n2}) \hat{B}$ follows from the Central Limit Theorem.

Proof of (b): It follows from (23), (24), and (25) that the least squares estimator $\hat{\gamma}_{reg}$ satisfies (27). Let $B_{\perp} = (\gamma', -I_{p-m})'$, then

$$-(\hat{B} - B)'B_{\perp} = \hat{\gamma}_{reg} - \gamma = B'(\hat{B}_{\perp} - B_{\perp}).$$

Proof of (c): Note that for the other parametrization, (20), where $B = (C'_{\eta}, C'_{\eta}\gamma)'$, it holds that $B_{\perp} = (\gamma', -I_{p-m})'$, such that for both parametrizations (27) holds. The estimator of B , in the parametrization (20), is $\hat{B}' = (\hat{C}'_{\eta}, \hat{C}'_{\eta}\hat{\gamma})$, where \hat{C}_{η} is derived from the $n^{1/2}$ -consistent estimator of Ω_{η} , such that for this parametrization, estimation of B is not n -consistent, but only $n^{1/2}$ -consistent and $\hat{B} - B = O_P(n^{-1/2})$. ■

Proof of Theorem 1. *Proof of (a):* Let $w_t = BT_t - \hat{B}\hat{E}_tT_t$, then

$$BE_tT_t - \hat{B}\hat{E}_tT_t = B(E_tT_t - T_t) + (BT_t - \hat{B}\hat{E}_tT_t) = B(E_tT_t - T_t) + w_t. \tag{A1}$$

Here $B(E_tT_t - T_t)$ is stationary, so that it is enough to show that w_t is asymptotically stationary. From the definition of T_{t+1} and the Kalman filter recursion (10) calculated for $T_0 = 0$ and for the estimated parameters, it holds that

$$\begin{aligned} BT_{t+1} &= BT_t + B\eta_{t+1}, \\ \hat{B}\hat{E}_{t+1}T_{t+1} &= \hat{B}\hat{E}_tT_t - \hat{B}\hat{K}'_t(y_{t+1} - \hat{B}\hat{E}_tT_t). \end{aligned}$$

Subtracting the expressions gives

$$\begin{aligned} BT_{t+1} - \hat{B}\hat{E}_{t+1}T_{t+1} &= BT_t + B\eta_{t+1} - \hat{B}\hat{E}_tT_t - \hat{B}\hat{K}'_t(y_{t+1} - \hat{B}\hat{E}_tT_t) \\ &= BT_t - \hat{B}\hat{E}_tT_t - \hat{B}\hat{K}'_t(BT_t + \varepsilon_{t+1} - \hat{B}\hat{E}_tT_t) + B\eta_{t+1}, \end{aligned}$$

which gives the recursion

$$w_{t+1} = (I_p - \hat{B}\hat{K}'_t)w_t - \hat{B}\hat{K}'_t\varepsilon_{t+1} + B\eta_{t+1}. \tag{A2}$$

Note that $(I_p - \hat{B}\hat{K}'_t)$ is not a contraction, because $p - m$ eigenvalues are one. Hence it is first proved that $\hat{B}'_{\perp}w_t$ is small and then a contraction is found for $\hat{B}'w_t$. From the definition of w_t , it follows from (27), that

$$\hat{B}'_{\perp}w_t = \hat{B}'_{\perp}BT_t = \hat{B}'_{\perp}(B - \hat{B})T_t = -(B_{\perp} - \hat{B}_{\perp})'\hat{B}T_t = O_P(n^{-1})O_P(n^{1/2}) = O_P(n^{-1/2}).$$

Next define $\hat{\hat{B}} = \hat{B}(\hat{B}'\hat{B})^{-1}$ and $\hat{\hat{B}}_{\perp} = \hat{B}_{\perp}(\hat{B}'_{\perp}\hat{B}_{\perp})^{-1}$, such that $I_p = \hat{\hat{B}}\hat{\hat{B}}' + \hat{\hat{B}}_{\perp}\hat{\hat{B}}'_{\perp}$. From (A2) it follows by multiplying by $\hat{\hat{B}}'$ and using $\hat{\hat{B}}B = \hat{\hat{B}}'(B - \hat{B}) + I_m = I_m + O_P(n^{-1/2})$, that

$$\begin{aligned} \hat{\hat{B}}'w_{t+1} &= (\hat{\hat{B}} - \hat{K}_t)'w_t - \hat{K}'_t\varepsilon_{t+1} + \hat{\hat{B}}'B\eta_{t+1} \\ &= (\hat{\hat{B}} - \hat{K}_t)'(\hat{\hat{B}}\hat{\hat{B}}' + \hat{\hat{B}}_{\perp}\hat{\hat{B}}'_{\perp})w_t - \hat{K}'_t\varepsilon_{t+1} + \eta_{t+1} + \hat{\hat{B}}'(B - \hat{B})\eta_{t+1} \\ &= (I_m - \hat{K}'_t\hat{\hat{B}})\hat{\hat{B}}'w_t - \hat{K}'_t\varepsilon_{t+1} + \eta_{t+1} + O_P(n^{-1/2}), \end{aligned}$$

because $\hat{\hat{B}}'(B - \hat{B})\eta_{t+1} = O_P(n^{-1/2})$ and $(\hat{\hat{B}} - \hat{K}_t)'\hat{\hat{B}}_{\perp}\hat{\hat{B}}'_{\perp}w_t = -\hat{K}'_t\hat{B}_{\perp}\hat{\hat{B}}'_{\perp}w_t = O_P(n^{-1/2})$.

From (14) it is seen that $I_m - \hat{K}'_t\hat{B} \xrightarrow{P} \Omega_B(V + \Omega_B)^{-1}$ and $(I_m - K'B)^n \rightarrow 0, n \rightarrow \infty$. This shows that $\hat{\hat{B}}'w_t$ and hence w_t is asymptotically a stationary AR(1) process.

Proof of (b): The CVAR (18) is expressed as $\Pi(L)y_t = v_t$, and the parameters are estimated using maximum likelihood with lag length $k_n \rightarrow \infty$ and $k_n^3/n \rightarrow 0$. This gives estimators $(\check{\alpha}, \check{\beta}, \check{\Gamma}, \check{C}, \check{\Sigma})$ and residuals \check{v}_t . The representation of y_t in terms of v_t is given by

$$y_t = C \sum_{i=1}^t v_i + \sum_{i=0}^{\infty} C_i v_{t-i} + A,$$

where $\beta' A = 0$. This relation also holds for the estimated parameters and residuals, and subtracting one finds

$$BT_t^* \sum_{i=1}^t v_i - \check{B}\check{T}_t^* \sum_{i=1}^t \check{v}_i = \sum_{i=0}^{\infty} \check{C}_i \check{v}_{t-i} - \sum_{i=0}^{\infty} C_i v_{t-i} - A + \check{A}.$$

It is seen that the right hand side is $o_p(1)$ and hence asymptotically stationary.

Proof of (c): Each estimated trend is compared with the corresponding trend which gives

$$\hat{B}\hat{E}_t T_t - \check{B}\check{T}_t^* = (\hat{B}\hat{E}_t T_t - BT_t) + (BT_t - BT_t^*) + (BT_t^* - \check{B}\check{T}_t^*).$$

Here the first term is asymptotically stationary using Theorem 2(a), the middle term is asymptotically stationary, and the last is $o_p(1)$ by Theorem 1(b). ■

Proof of Theorem 2. The proof is the same for all the spreads, so consider $E_t T_t - \hat{E}_t T_t$, and the identity

$$\hat{B}(BE_t T_t - \hat{B}\hat{E}_t T_t) = \hat{B}'(B - \hat{B})E_t T_t + \hat{B}'\hat{B}(E_t T_t - \hat{E}_t T_t).$$

The left hand side is asymptotically stationary by Theorem 1(a) and therefore $E_t T_t - \hat{E}_t T_t$ is asymptotically stationary if and only

$$\hat{B}'(B - \hat{B})E_t T_t = [n^{1/2}\hat{B}'(B - \hat{B})][n^{-1/2}E_t T_t],$$

is asymptotically stationary. Here the second factor converges to a nonstationary process,

$$n^{-1/2}E_{[nu]} T_{[nu]} = n^{-1/2}E_0 T_0 + n^{-1/2} \sum_{j=2}^{[nu]} K'_{j-1} v_j \xrightarrow{D} W_v(u),$$

see (16), so for the term $[n^{1/2}\hat{B}'(B - \hat{B})][n^{-1/2}E_t T_t]$ to be asymptotically stationary it is necessary and sufficient that $n^{1/2}\hat{B}'(B - \hat{B}) \xrightarrow{P} 0$. ■

References

- Chan, Siew Wah, Graham Clifford Goodwin, and Kwai Sang Sin. 1984. Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transaction on Automatic Control* 29: 110–18.
- Durbin, Jim, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*, 2nd ed. Oxford: Oxford University Press.
- Fisher Frank, M. 1966. *The Identification Problem in Econometrics*. New York: McGraw-Hill.
- Harvey, Andrew. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, Andrew C. 2006. Forecasting with Unobserved Components Time Series Models. In *Handbook of Economic Forecasting*. Edited by G. Elliot, C. Granger and A. Timmermann. Amsterdam: North Holland, pp. 327–412.
- Harvey, Andrew C., and Siem Jan Koopman. 1997. Multivariate structural time series models. In *System Dynamics in Economics and Financial Models*. Edited by C. Heij, J. M. Schumacher, B. Hanzon and C. Praagman. New York: John Wiley and Sons.

- Hoover, Kevin D., Søren Johansen, Katarina Juselius, and Morten Nyboe Tabor. 2014. Long-run Causal Order: A Progress Report. Unpublished manuscript.
- Johansen, Søren. 1996. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed. Oxford: Oxford University Press.
- Johansen, Søren. 2010. Some identification problems in the cointegrated vector autoregressive model. *Journal of Econometrics* 158: 262–73.
- Johansen, Søren, and Katarina Juselius. 2014. An asymptotic invariance property of the common trends under linear transformations of the data. *Journal of Econometrics* 178: 310–15.
- Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*, 2nd ed. Cambridge: Cambridge University Press.
- Saikkonen, Pentti. 1992. Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8: 1–27.
- Saikkonen, Pentti, and Helmut Lütkepohl. 1996. Infinite order cointegrated vector autoregressive processes. Estimation and Inference. *Econometric Theory* 12: 814–44.
- Spirites, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*, 2nd ed. Cambridge: MIT Press.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).