Discussion Papers
Department of Economics
University of Copenhagen

No. 17-02

Cointegration between trends and their estimators in state space models and CVAR models

Søren Johansen

Morten Nyboe Tabor

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)

Cointegration between trends and their estimators in state space models and CVAR models

Søren Johansen[*,†]
University of Copenhagen
and CREATES

Morten Nyboe Tabor[‡]
University of Copenhagen

February 28, 2017

Abstract

In a linear state space model, $y_{t+1} = BT_t + \varepsilon_{t+1}$, we investigate if the unobserved trend, T_t, cointegrates with the extracted trend $\mathsf{E}_t T_t$, and with the estimated trend $\hat{\mathsf{E}}_t T_t$, in the sense that the spreads $T_t - \mathsf{E}_t T_t$ and $E_t T_t - \hat{E}_t T_t$ are stationary. We find that this result holds for $BT_t - BE_t T_t$ and $BE_t T_t - \hat{B}\hat{E}_t T_t$. For the trends T_t and $\hat{E}_t T_t$, however, this type cointegration depends on the identification of B and T_t. The same results are found, if the observations, y_t, from the state space model are analysed using a cointegrated vector autoregressive model, where the trend is defined as the common trend. Finally we investigate cointegration between trends and their estimators based on the two models, and find the same results. We illustrate with two examples and confirm the results by a small simulation study

Keywords: Cointegration of trends, State space models, CVAR models

JEL Classification: C32.

1 Introduction and Summary

In connection with a project on long-run causal order for nonstationary processes, see Hoover, Johansen, Juselius and Tabor (2014), we used the state space model (SSM)

$$\begin{aligned} T_{t+1} &= T_t + \eta_{t+1}, \\ y_{t+1} &= BT_t + \varepsilon_{t+1}, \end{aligned} \tag{1}$$

$t = 1, \ldots, n$, to estimate the effect, B, of unobserved independent random walks, T_t, on the observed series, y_t. When analysing some simulation studies, we noticed that the estimator

[*]The first author is grateful to CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The second author 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.

[†]Address for correspondence: Department of Economics University of Copenhagen, Øster Farimagsgade 5 Building 26, 1353 Copenhagen K, Denmark. `E-mail: soren.johansen@econ.ku.dk.`

[‡]Department of Economics, University of Copenhagen. `E-mail: morten.nyboe.tabor@econ.ku.dk`

of the extracted trend, $\hat{E}_t T_t$, sometimes did not cointegrate with the simulated trend, T_t, in the sense that $T_t - \hat{E}_t T_t$ was not stationary. In order to understand this phenomenon we have analysed the state space model and found the solution, namely that we always have cointegration $(1, -1)$ between BT_t and its estimator $\hat{B}\hat{E}_t T_t$, but not in general between T_t and $\hat{E}_t T_t$. The latter result depends on how the parameter B and the trend T_t are identified. We then investigated the same questions using a cointegrated vector autoregressive model (CVAR) to analyse $y_1, t = 1, \ldots, n$, and found the same results. Finally we compared the estimated trends extracted from the SSM and the CVAR models and found the same results.

The results are given in Theorems 1, 2 and 3. For the trends from the CVAR, the trends from the SSM, and finally for a comparison for the estimated trends from the two models.

2 The DGP and the statistical models

The data generating process (DGP) is formulated as

$$\begin{aligned} T^0_{t+1} &= T^0_t + \eta_{t+1}, \\ y_{t+1} &= B^0 T^0_t + \varepsilon_{t+1}, \end{aligned} \tag{2}$$

where $T^0_t \in \mathbb{R}^m$ and $y_t \in \mathbb{R}^p$, $t = 1, \ldots, n$. We assume $m < p$, and that B^0 has full column rank. Moreover η_t and ε_t are mutually independent and i.i.d. Gaussian with mean zero and variances $\Omega^0_\eta = diag(\sigma^0_{11}, \ldots, \sigma^0_{pp}) > 0$ and $\Omega^0_\varepsilon > 0$ respectively. Note that the trend enters with a lag in the observation equation. This is of course just a question of notation, as we can define the trend $S^0_{t+1} = T^0_t$, and the observation equation becomes $y_{t+1} = B^0 S^0_{t+1} + \varepsilon_{t+1}$.

2.1 The statistical models

We analyse the observations $y_1, \ldots, y_n$ from the DGP given in (2) using two different statistical models each containing the DGP.

2.1.1 The first statistical model

We consider the state space model (SSM) defined by (1) for the observed process, y_t, and unobserved process, T_t, with $p \times m$ parameter B of full rank, and where $\Omega_\eta > 0$ and $\Omega_\varepsilon > 0$ are freely varying positive covariance matrices of dimensions $m \times m$ and $p \times p$ respectively. This statistical model can be analysed using the Kalman filter to calculate the likelihood function, and an optimizing algorithm can be used to find the maximum likelihood estimator.

2.1.2 The second model

Next we reformulate the DGP in error correction form for the observation (y_t, T^0_t)

$$\begin{pmatrix} \Delta y_{t+1} \\ \Delta T^0_{t+1} \end{pmatrix} = \begin{pmatrix} -I_p & B^0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ T^0_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ \eta_{t+1} \end{pmatrix}.$$

It is seen that $y_1, \ldots, y_n$ are partial observations from a CVAR for $(y'_t, T^{0\prime}_t)'$, and we use the infinite order CVAR as statistical model

$$\Delta^2 y_t = \alpha\beta' y_{t-1} - \Gamma \Delta y_{t-1} + \sum_{i=0}^{\infty} \Gamma_i \Delta^2 y_{t-i} + \delta_t. \tag{3}$$

Using results of Saikkonen (1992) and Saikkonen and Lütkepohl (1996), parameters α, β, Γ, and residuals δ_t can be estimated consistently by choosing a finite lag length $k_n \to \infty$, such that $k_n^3/n \to 0$.

Note that the parameters of the SSM model consist of $B, \Omega_\eta, \Omega_\varepsilon$, whereas the parameters of the CVAR are $\alpha, \beta, \Gamma, \Gamma_i, i = 0, \ldots, \infty$,and Ω_δ. Let $B_\perp^0$ be $p \times (p-m)$ of full rank satisfying $B_\perp^{0\prime} B^0 = 0$. We see from (2), that $B_\perp^{0\prime} y_t = B_\perp^{0\prime} \varepsilon_t$ is stationary, such that the cointegrating vectors in the DGP are $\beta^0 = B_\perp^0$. Hence it is natural to assume that parameters in the two models are related by $\beta_\perp = B$.

Moreover, the long-run variance in the SSM, $\lim_{n\to\infty} n^{-1} Var(y_n) = B\Omega_\eta B'$, is the same as the long-run variance in the CVAR, $\lim_{n\to\infty} n^{-1} Var(y_n) = C\Omega C'$, where $C = B(\alpha_\perp' \Gamma B)^{-1} \alpha_\perp'$, which gives another relation between the parameters, see Johansen and Juselius (2014).

2.2 The trends

Given the two models, there are two ways of defining trends. In the SSM generated by (2), the trend T_t is part of the model formulation, and in the CVAR (3) we define the trend as the common trend

$$T_t^* = (\alpha_\perp' \Gamma B)^{-1} \alpha_\perp' \sum_{s=1}^{t} \delta_s,$$

where δ_t is the prediction error for y_t given the infinite past $\{y_s, s < t\}$, see (3).

Thus we have two representations of y_t,

$$SSM : y_t = BT_t + \varepsilon_t - B\eta_t, \tag{4}$$

$$CVAR : y_t = C \sum_{s=1}^{t} \delta_s + u_t = B(\alpha_\perp' \Gamma B)^{-1} \alpha_\perp' \sum_{s=1}^{t} \delta_s + u_t = BT_t^* + u_t, \tag{5}$$

using the Granger Representation Theorem, see Johansen (1996, Theorem 4.2). Finally u_t is an asymptotically stationary process.

2.3 The Kalman filter

Conditional on $y_1, \ldots, y_t$, T_t is Gaussian with mean $\mathsf{E}_t T_t = \mathsf{E}(T_t | y_1, \ldots, y_t)$ and variance $\mathsf{V}_t = Var(T_t | y_1, \ldots, y_t)$. It is well known, see Durbin and Koopman (2012) or Harvey (1989), that $\mathsf{E}_t T_t$ and V_t can be calculated recursively by the Kalman filter starting with $\mathsf{E}_1 T_1 = 0$ and $\mathsf{V}_1 = \Omega_\eta$, using the equations for $t = 1, \ldots, T-1$,

$$\mathsf{E}_{t+1} T_{t+1} = \mathsf{E}_t T_t + \mathsf{K}_t (y_{t+1} - \mathsf{E}_t y_{t+1}) \tag{6}$$

$$\mathsf{V}_{t+1} = \Omega_\eta + \mathsf{V}_t - \mathsf{V}_t B' (B \mathsf{V}_t B' + \Omega_\varepsilon)^{-1} B \mathsf{V}_t,$$

where $\mathsf{K}_t = \mathsf{V}_t B' (B \mathsf{V}_t B' + \Omega_\varepsilon)^{-1}$ is the Kalman gain. Let $\bar{B} = B(B'B)^{-1}$ and define $\Omega_B = \Omega_B = Var(\bar{B}' \varepsilon_t | B_\perp' \varepsilon_t)$.

Lemma 1 *The recursion for* V_t *can be expressed as*

$$\mathsf{V}_{t+1} = \Omega_\eta + \mathsf{V}_t - \mathsf{V}_t (\mathsf{V}_t + \Omega_B)^{-1} \mathsf{V}_t. \tag{7}$$

Solving the eigenvalue problem

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

we find eigenvectors W and eigenvalues $\Lambda = diag(\lambda_1, \dots, \lambda_m)$, such that $W'\Omega_B W = I_m$ and $W'\Omega_\eta W = \Lambda$. Then, if $\mathsf{V}_1 = \Omega_\eta$, we find $W'\mathsf{V}_t W \to W'\mathsf{V}W = \mathsf{\Lambda}_\infty$, where

$$\lambda_{i,\infty} = \frac{1}{2}\{\lambda_i + (\lambda_i^2 + 4\lambda_i)^{1/2}\}. \tag{8}$$

It follows that

$$\mathsf{K}_t B = \mathsf{V}_t B'(B\mathsf{V}_t B' + \Omega_\varepsilon)^{-1} B \to \mathsf{K}B = \mathsf{V}(\mathsf{V}' + \Omega_B)^{-1}, \tag{9}$$

such that $\mathsf{K}B$ has positive eigenvalues less than one.

2.4 Estimation

With the above definitions of trends, there is an identification issue between B and T_t (or B and Ω_η) and between B and T_t^*, because for any $m \times m$ matrix M of full rank, we can use BM^{-1} as parameter and MT_t as trend and $M\Omega_T M'$ as variance, and similarly for T_t^*. In order to estimate B, T, and Ω_η, we need to impose restrictions, and we shall give two examples of identification.

2.4.1 Identification 1.

Because B has rank m, we can permute the rows and assume that $B' = (B_1', B_2')$, where B_1 is $m \times m$ and has full rank. Then we redefine the parameters and trend as

$$\dot{B} = \begin{pmatrix} I_m \\ B_2 B_1^{-1} \end{pmatrix} = \begin{pmatrix} I_m \\ \gamma' \end{pmatrix}, \ \dot{\Omega}_\eta = B_1 \Omega_\eta B_1', \ \dot{T}_t = B_1 T_t. \tag{10}$$

This parametrization is useful because it separates parameters that are n-consistently estimated, γ, from those that are $n^{1/2}$-consistently estimated, $\Omega_\eta, \Omega_\varepsilon$, see Lemma 2. Note that we simply identify the (correlated) trends by defining T_{1t} as the trend in y_{1t}, T_{2t} as the trend in y_{2t}, and so on.

2.4.2 Identification 2.

We also want to consider the normalization where $\dot{\Omega}_\eta$ is diagonal, and define a Cholesky decomposition $\dot{\Omega}_\eta = C_\eta diag(\sigma_{11}, \dots, \sigma_{mm}) C_\eta'$, and the new parameters and the trend

$$\ddot{B} = \begin{pmatrix} C_\eta \\ \gamma' C_\eta \end{pmatrix}, \ \ddot{\Omega}_\eta = diag(\sigma_{11}, \dots, \sigma_{mm}), \ \ddot{T}_t = C_\eta^{-1} \dot{T}_t, \tag{11}$$

such that $\ddot{B}\ddot{T}_t = \dot{B}\dot{T}_t = BT_t$ and $\ddot{B}\ddot{\Omega}_\eta \ddot{B}' = \dot{B}\dot{\Omega}_\eta \dot{B}' = B\Omega_\eta B'$.

This parametrization is also useful because it defines independent trends and how they load into the observations. An example is to define T_{1t} 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}, etc.

2.4.3 A simple estimator

We can estimate the state space model using the Kalman filter to calculate the likelihood function for given parameters and then maximize the likelihood function using a general optimizing algorithm. But we can also find very simple (but not efficient) estimators, which are easier to analyse.

Irrespective of the identification, we find the relations

$$Var(\Delta y_t) = B\Omega_\eta B' + 2\Omega_\varepsilon, \tag{12}$$

$$Cov(\Delta y_t, \Delta y_{t+1}) = -\Omega_\varepsilon. \tag{13}$$

In the identified parametrization (10), where $B = (I_m, \gamma)'$, we take $B_\perp = (\gamma', -I_{p-m})'$, define $z_{1t} = (y_{1t}, \dots, y_{mt})'$ and $z_{2t} = (y_{m+1,t}, \dots, y_{pt})$ with a similar decomposition of $\varepsilon_t = (\xi'_{1t}, \xi'_{2t})'$. Then the equation for $y_{t+1} = (z_{1,t+1}, z_{2,t+1})'$ becomes

$$z_{1,t+1} = T_t + \xi_{1,t+1}, \tag{14}$$

$$z_{2,t+1} = \gamma' T_t + \xi_{2,t+1} = \gamma' z_{1,t+1} - \gamma'\xi_{1,t+1} + \xi_{2,t+1} = \gamma' z_{1,t+1} - \varepsilon'_t B_\perp \tag{15}$$

Lemma 2 *Assume B is identified as in (10), that is $B' = (I_m, \gamma)$, $B'_\perp = (\gamma, -I_{p-m})$, and Ω_η is adjusted accordingly. Using (12) and (13), we can find $n^{1/2}$-consistent estimators, which are asymptotically Gaussian for the variance matrices Ω_η and Ω_ε.*

From equation (15), we find by regressing $z_{2,t+1}$ on $z_{1,t+1}$, that the estimator $\hat{\gamma}_{reg}$ is n-consistent with asymptotic Mixed Gaussian distribution

$$n(\hat{\gamma}_{reg} - \gamma) = nB'(\hat{B}_\perp - B_\perp) \xrightarrow{\text{D}} (\int_0^1 W_\eta W'_\eta du)^{-1} \int_0^1 W_\eta (dW_\varepsilon)' B_\perp. \tag{16}$$

If B is identified as in (11), that is $B = (C'_\eta, C'_\eta \gamma)'$, and $\Omega_\eta = diag(\sigma_{11}, \dots, \sigma_{pp})$, then $\hat{B} - B = O_P(n^{-1/2})$ but (16) still holds for $nB'(\hat{B}_\perp - B^0_\perp)$.

Using the Kalman filter, we can calculate the extracted trend $\mathsf{E}_t(T_t)$ based on observations and known parameters, and the estimator of the extracted trend $\hat{T}_t = \hat{\mathsf{E}}_t T_t$, based on observations and estimated parameters.

Using the CVAR we find maximum likelihood estimators of the common trend $\breve{T}^*_t = (\breve{\gamma}'_\perp \breve{\Gamma} \breve{B})^{-1} \breve{\gamma}'_\perp \sum_{s=1}^t \breve{\delta}_s$. Note that we use $\hat{B}$ to indicate an estimator in SSM, and $\breve{B}$ to indicate an estimator in CVAR.

Based on these trends, we can ask whether the trends cointegrate $(1, -1)$ with the process y_t, whether they cointegrate with the estimated trends, and whether the estimated trends from the two models cointegrate similarly with each other.

3 Cointegration between trends and their estimators

This section gives the main results in three theorems with proofs in the Appendix. In Theorem 1 we show that in the CVAR model, the estimated trend $\breve{B}\breve{T}^*_t$ cointegrates with the trend BT^*_t, and hence with y_t, by showing that $y_t - BT^*_t$ and $BT^*_t - \breve{B}\breve{T}^*_t$ are asymptotically stationary. However, whether $T^*_t - \breve{T}^*_t$ is asymptotically stationary depends on the identification of B, T_t, and Ω_η. Theorem 2 gives the same results for the state space model, and finally in Theorem 3, we compare the estimated trends in the two models and show that $\hat{B}\hat{T}_t - \breve{B}\breve{T}^*_t$ is asymptotically stationary, but the same does not in general hold for $\hat{T}_t - \breve{T}^*_t$.

The conclusion is that in terms of cointegration of the trends, it does not matter which model we use, as long as we focus on identified trends BT_t and BT^*_t.

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

$$\hat{B}(T_t - \hat{T}_t) = (\hat{B} - B)T_t + (BT_t - \hat{B}\hat{T}_t),$$

where $BT_t - \hat{B}\hat{T}_t$ is asymptotically stationary by Theorem 2 (b), but $(\hat{B} - B)T_t$ is not necessarily asymptotically stationary, because in general $\hat{B} - B = O_P(n^{-1/2})$ and $T_t = O_P(n^{1/2})$. If, however, B is identified as in (10), then we have asymptotic stationarity for $T_t - \hat{T}_t = (I_m, 0_{m\times(p-m)})(BT_t - \hat{B}\hat{T}_t)$. Thus, identification and estimation of parameters can explain the lack of cointegration between estimated and true trends, see Examples 1 and 2 for an illustration.

Theorem 1 *Let y_t and T_t be generated by the DGP given in (2). If we use the CVAR (3) for inference, and define the trend $T_t^* = (\alpha_\perp' \Gamma B)^{-1}\alpha_\perp' \sum_{s=1}^t \delta_s$, with estimator $\breve{T}_t^*$, then*
(a) $y_t - BT_t^*$ is asymptotically stationary,
(b) $BT_t^* - \breve{B}\breve{T}_t^*$ is asymptotically stationary,
(c) $T_t^* - \breve{T}_t^*$ is not necessarily asymptotically stationary.

If we choose the parametrization (10), then $(I_m, 0_{m\times(p-m)})(BT_t^* - \breve{B}\breve{T}_t^*) = T_t^* - \breve{T}_t^*$, such that also $T_t^* - \breve{T}_t^*$ is asymptotically stationary, but for the parametrization (11), the result does not hold.

In the state space model we can prove similar results

Theorem 2 *Let y_t and T_t be generated by the DGP given in (2). If we use the state space model defined by (2) for inference, see Lemma 1, then*
(a) $T_t - \mathsf{E}_t T_t$ is asymptotically stationary,
(b) $\mathsf{E}_t BT_t - \hat{\mathsf{E}}_t \hat{B} T_t$ is asymptotically stationary,
(c) $\mathsf{E}_t T_t - \hat{\mathsf{E}}_t T_t$ is not in general asymptotically stationary.

Finally we compare trends estimated in the two models and prove the main result that trends derived from the SSM and CVAR models cointegrate $(1, -1)$, as long as we consider $\hat{B}\hat{T}_t$ and $\breve{B}^*\breve{T}_t^*$. For $\hat{T}_t$ and $\breve{T}_t^*$ we get the result again, that asymptotic stationarity depends on how B and the trends are identified.

Theorem 3 *Let y_t and T_t be generated by the DGP given in (2). We estimate the trend $T_t^* = (\alpha_\perp' \Gamma B)^{-1}\alpha_\perp' \sum_{s=1}^t \delta_s$ by analysing CVAR (3), and estimate $E_t T_t$ using SSM (2). If $\hat{B}$ be the estimator for B derived from SSM, whereas $\breve{B}$ is derived from CVAR, then*
(a) $T_t - T_t^*$ is asymptotically stationary,
(b) $\hat{B}\hat{T}_t - \breve{B}\breve{T}_t^*$ is asymptotically stationary,
(c) $\hat{T}_t - \breve{T}_t^*$ is not necessarily asymptotically stationary.

4 Two examples

We give two examples where $p = 3$ 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, we need to impose 4 restrictions to identify the parameters.

Example 1. We first consider the model with the identification given in (11),

$$B = \begin{pmatrix} 1 & 0 \\ a_{21} & 1 \\ a_{31} & a_{32} \end{pmatrix}, \ \Omega_\eta = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}, \tag{17}$$

such that

$$\mathsf{E}_t BT_t - \hat{\mathsf{E}}_t \hat{B} T_t = \begin{pmatrix} \mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t} \\ a_{21}\mathsf{E}_t T_{1t} + \mathsf{E}_t T_{2t} - \hat{a}_{21}\hat{\mathsf{E}}_t T_{1t} - \hat{\mathsf{E}}_t T_{2t} \\ a_{31}\mathsf{E}_t T_{1t} + a_{32}\mathsf{E}_t T_{2t} - \hat{a}_{31}\hat{\mathsf{E}}_t T_{1t} - \hat{a}_{32}\hat{\mathsf{E}}_t T_{2t} \end{pmatrix}. \tag{18}$$

By the results in Theorem 2 (b), we see that $\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t}$ is asymptotically stationary, because it can be recovered as the first row of $\mathsf{E}_t BT_t - \hat{\mathsf{E}}_t \hat{B} T_t$ in (18). Moreover the second row of (18) $(a_{21}\mathsf{E}_t T_{1t} - \hat{a}_{21}\hat{\mathsf{E}}_t T_{1t}) + (\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t})$ is asymptotically stationary. Thus to prove that $\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t}$ is asymptotically stationary, it is enough to show stationarity of

$$a_{21}\mathsf{E}_t T_{1t} - \hat{a}_{21}\hat{\mathsf{E}}_t T_{1t} = (a_{21} - \hat{a}_{21})\mathsf{E}_t T_{1t} - \hat{a}_{21}(\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t}). \tag{19}$$

Here $\hat{a}_{21}(\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t})$ is asymptotically stationary because $\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t}$ is, but the first term is not, because $\hat{a}_{21}$ is $n^{1/2}$-consistent, and in this case $n^{1/2}(a_{21} - \hat{a}_{21}) \xrightarrow{D} Z$ Gaussian, such that

$$(a_{21} - \hat{a}_{21})\mathsf{E}_t T_{1t}|_{t=[nu]} = n^{1/2}(a_{21} - \hat{a}_{21})n^{-1/2}\mathsf{E}_t T_{1t}|_{t=[nu]} \xrightarrow{D} Z W_{\eta_1}(u),$$

which is nonstationary. This argument is a special case of the proof (22).

To illustrate theresults, we simulate data from the model with $n = 100$ observations starting with $T_1 = 0$, and parameter values $a_{21} = 0.0$, $a_{31} = a_{32} = 0.5$, $\sigma_{11} = \sigma_{22} = 1$. We estimate the parameters by Gaussian maximum likelihood using SsfPack 3.0, see Koopman, Shepard, and Doornik (2008), where we fix $y_1 = 0$ and assume a diffuse prior for the initial value T_1. The results are summarized in Figure 1. The panels a and b show plots of $(\mathsf{E}_t T_{1t}, \hat{\mathsf{E}}_t T_{1t})$ and $(\mathsf{E}_t T_{2t}, \hat{\mathsf{E}}_t T_{2t})$ respectively, and we see that they co-move in each panel. In panels c and d we show the differences $\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t}$ and $\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t}$. We note that the first looks stationary, whereas the second is clearly nonstationary. When comparing with the plot of $\mathsf{E}_t T_{1t}$ in panel a, it appears that the process $\hat{\mathsf{E}}_t T_{1t}$ can explain the nonstationarity of $\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t}$ consistently with equation (19) and Table 1. Finally we analysed the four variables $(\mathsf{E}_t T_{1t}, \hat{\mathsf{E}}_t T_{1t}, \mathsf{E}_t T_{2t}, \hat{\mathsf{E}}_t T_{2t})$ using a CVAR and we found two cointegrating relations given in Table 1.■

	$\mathsf{E}_t T_{1t}$	$\hat{\mathsf{E}}_t T_{1t}$	$\mathsf{E}_t T_{2t}$	$\hat{\mathsf{E}}_t T_{2t}$
β_1'	-1	1	0	0
β_2'	0	$-\underset{(0.003)}{0.26}$	-1	1

Table 1: The result of a cointegration analysis of $\mathsf{E}_t T_{1t}$, $\hat{\mathsf{E}}_t T_{1t}$, $\mathsf{E}_t T_{2t}$, and $\hat{\mathsf{E}}_t T_{2t}$ using model (16). We found two cointegrating relations given in the Table. Note that $\hat{\mathsf{E}}_t T_{1t}$ has a significant coefficient in the second cointegrating relation. See also Figure 1 panel d, where $\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t}$ is plotted, and it is seen that $\hat{\mathsf{E}}_t T_{2t}$ is need to get a stationary relation.

Example 2. If instead we choose the parametrization (10)

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a_{31} & a_{32} \end{pmatrix}, \ \Omega_\eta = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \tag{20}$$

we still have just identified parameters, but now

$$\mathsf{E}_t BT_{1t} - \hat{\mathsf{E}}_t \hat{B} T_{1t} = \begin{pmatrix} \mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t} \\ \mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t} \\ a_{31}\mathsf{E}_t T_{1t} + a_{32}\mathsf{E}_t T_{2t} - \hat{a}_{31}\hat{\mathsf{E}}_t T_{1t} - \hat{a}_{32}\hat{\mathsf{E}}_t T_{2t} \end{pmatrix}. \tag{21}$$

Thus, both $\mathsf{E}_t T_{1t} - \hat{\mathsf{E}}_t T_{1t}$ and $\mathsf{E}_t T_{2t} - \hat{\mathsf{E}}_t T_{2t}$ are asymptotically stationary, as the first two rows of $\mathsf{E}_t BT_{1t} - \hat{\mathsf{E}}_t \hat{B} T_{1t}$, in (21), by Theorem 2.

In the simulation of this example, we use the values $a_{21} = 0.0$, $a_{31} = a_{32} = 0.5$ and $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = 0$, such that in fact the DGP for this example is the same as the DGP for Example 1, namely

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

In fact we analysed the same data as in Example 1. We plot the data in Figure 2. The panels a and b show plots of $(\mathsf{E}_t T_{1t}, \hat{\mathsf{E}}_t T_{1t})$ and $(\mathsf{E}_t T_{2t}, \hat{\mathsf{E}}_t T_{2t})$ respectively, and we see that they co-move. In panels c and d we show the differences $\hat{\mathsf{E}}_t T_{1t} - \mathsf{E}_t T_{1t}$ and $\hat{\mathsf{E}}_t T_{2t} - \mathsf{E}_t T_{2t}$, which appear to be stationary in this parametrization of the model. A cointegration analysis of $(\mathsf{E}_t T_{1t}, \hat{\mathsf{E}}_t T_{1t}, \mathsf{E}_t T_{2t}, \hat{\mathsf{E}}_t T_{2t})$ shows that there two cointegrating relations, see Table 2. ■

	$\mathsf{E}_t T_{1t}$	$\hat{\mathsf{E}}_t T_{1t}$	$\mathsf{E}_t T_{2t}$	$\hat{\mathsf{E}}_t T_{2t}$
β_1'	-1	1	0	0
β_2'	0	0	-1	1

Table 2: The result of a cointegration analysis of $\mathsf{E}_t T_{1t}$, $\hat{\mathsf{E}}_t T_{1t}$, $\mathsf{E}_t T_{2t}$, and $\hat{\mathsf{E}}_t T_{2t}$ using model (19). We found the two cointegrating relations given here.

5 Conclusion

We have analysed data generated by a multivariate homogenous linear state space model

$$\begin{aligned} y_{t+1} &= B^0 T_t^0 + \varepsilon_{t+1}, \\ T_{t+1}^0 &= T_t^0 + \eta_{t+1}, \end{aligned}$$

where the state variable, T_t^0, is an unobserved multivariate random walk. We have used a state space model and a CVAR model to estimate the trend in the data. In both cases we find that the estimated trend cointegrates $(1, -1)$ with the trend in the model, provided we only consider BT_t and its estimator. The same result does not hold in general for the trend T_t and its estimator. The reason is that if the parameters are identified using only restrictions on Ω_η, the estimator of B is not n-consistent, and recovering the trend $\hat{T}_t$ from the $\hat{B}\hat{T}_t$ causes some trends not to cointegrate with estimated trends.

6 Appendix

Proof of Lemma 1. *Proof of (7):* We define $M=(\bar{B}, B_\perp)$, for $\bar{B} = B(B'B)^{-1}$, and find

$$\mathsf{K}_t B = \mathsf{V}_t B' M[(M'B\mathsf{V}_t B'M + M'\Omega_\varepsilon M)]^{-1} M'B$$
$$= \mathsf{V}_t \begin{pmatrix} I_m \\ 0 \end{pmatrix}' \begin{pmatrix} \mathsf{V}_t + \bar{B}'\Omega_\varepsilon \bar{B} & \bar{B}'\Omega_\varepsilon B_\perp \\ B_\perp'\Omega_\varepsilon \bar{B} & B_\perp'\Omega_\varepsilon B_\perp \end{pmatrix}^{-1} \begin{pmatrix} I_m \\ 0 \end{pmatrix} = \mathsf{V}_t(\mathsf{V}_t + \Omega_B)^{-1},$$

where

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

Proof of (8): If the recursion starts with $\mathsf{V}_1 = \Omega_\eta$, then all V_t can be diagonalized by W, such that the recursion for $\lambda_{i,t}$, the eigenvalue of V_t, becomes

$$\lambda_{i,t+1} = \lambda_i + \lambda_{i,t} - \frac{\lambda_{i,t}^2}{1 + \lambda_{i,t}}.$$

It is seen that $\lambda_{i,t}$ is increasing in t and the limit is $\lambda_{i,\infty} = \{\lambda_i + (\lambda_i^2 + 4\lambda_i)^{1/2}\}/2$. See for instance Chan, Goodwin and Sin (1984) for more general results. ■

Proof of Lemma 2. Consider first the product moments and their limits (12) and (13),

$$S_{1n} = n^{-1} \sum_{t=1}^{n} \Delta y_t \Delta y_t' \xrightarrow{\text{P}} B\Omega_\eta B' + 2\Omega_\varepsilon,$$
$$S_{2n} = n^{-1} \sum_{t=1}^{n} \Delta y_t \Delta y_{t+1}' \xrightarrow{\text{P}} -\Omega_\varepsilon.$$

Irrespective of the identfication, $B\Omega_\eta B'$ and Ω_ε can be estimated $n^{1/2}$-consistently with Gaussian limit distribution using the Central Limit Theorem.

To prove (16), we find from (14) and (15), that the least squares estimator $\hat{\gamma}_{reg}$ satisfies

$$n(\hat{\gamma}_{reg} - \gamma) = -(n^{-2} \sum_{t=1}^{n-1} z_{1,t+1} z_{1,t+1}')^{-1} n^{-1} \sum_{t=1}^{n-1} z_{1,t+1}\varepsilon_{t+1}' B_\perp$$
$$\xrightarrow{\text{D}} -(\int_0^1 W_\eta W_\eta' du)^{-1} \int_0^1 W_\eta (dW_\varepsilon)' B_\perp.$$

Here W_η is Brownian motion generated from η_t, adjusted to the identification of $B = (I_m, \gamma)'$, and W_ε is Brownian motion generated from ε_t. We choose $B_\perp = (\gamma^{0\prime}, -I_{p-m})'$ and find the relations

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

Note that for the other parametrization (11), where $B = (C_\eta', C_\eta'\gamma)'$ and we can choose the same $B_\perp = (\gamma^{0\prime}, -I_{p-m})'$, such that for both parametrizations we have (16). The estimator of B, however, changes in the parametrization (11), and we find

$$\hat{B} = \begin{pmatrix} \hat{C}_\eta \\ \hat{\gamma}'\hat{C}_\eta \end{pmatrix},$$

where $\hat{C}_\eta$ is derived from the $n^{1/2}$-consistent estimator of Ω_η, such that for this parametrization, we do not get n-consistent estimation of B, but only that $\hat{B} - B = O_P(n^{-1/2})$. ■

Proof of Theorem 1. *Proof of (a):* This follows from (5).

Proof of (b): The Granger representation (5) holds for parameters and residuals, but because y_t is also a solution to the equations with estimated parameters and estimated residuals, we have the same representation in terms of these. This implies that

$$BT_t^* - \breve{B}\breve{T}_t^* = C\sum_{s=1}^{t}\delta_s - \breve{C}\sum_{s=1}^{t}\breve{\delta}_s = y_t - u_t - (y_t - \breve{u}_t) = \breve{u}_t - u_t,$$

which is an asymptotically stationary process.

Proof of (c): We next find from

$$\breve{B}(T_t^* - \breve{T}_t^*) = (BT_t^* - \breve{B}\breve{T}_t^*) + (\breve{B} - B)T_t^*,$$

that the first term is stationary by (b), and the last term $(\breve{B} - B)T_t^* = O_P(1)$ is not stationary because for the parametrization (11) we have only $n^{1/2}$-consistent estimators, such that

$$n^{1/2}(\breve{B} - B) = O_P(1) \text{ and } n^{-1/2}T_{[nu]}^* \xrightarrow{\text{D}} W_\delta(u), \tag{22}$$

where W_δ is Brownian motion generated by δ_t. Thus, in general $\breve{B}(T_t^* - \breve{T}_t^*)$ and therefore $T_t^* - \breve{T}_t^*$ is not asymptotically stationary. ■

Proof of Theorem 2. *Proof of* (a)*:* We define the deviation between T_t and the extracted trend, $v_t = T_t - \mathsf{E}_t T_t$, and find

$$y_{t+1} - \mathsf{E}_t y_{t+1} = BT_t + \varepsilon_{t+1} - \mathsf{E}_t BT_t = Bv_t + \varepsilon_{t+1}, \tag{23}$$

$$T_{t+1} - \mathsf{E}_t T_t = T_t - \mathsf{E}_t T_t + \eta_{t+1} = v_t + \eta_{t+1}. \tag{24}$$

From the Kalman filter equations (6), we find

$$v_{t+1} = T_{t+1} - \mathsf{E}_{t+1}T_{t+1} = T_{t+1} - \mathsf{E}_t T_t - \mathsf{K}_t(y_{t+1} - \mathsf{E}_t y_{t+1}), \tag{25}$$

such that, using (23) and (24),

$$v_{t+1} = v_t + \eta_{t+1} - \mathsf{K}_t(Bv_t + \varepsilon_{t+1}) = (I_m - \mathsf{K}_t B)v_t - \mathsf{K}_t\varepsilon_{t+1} + \eta_{t+1}.$$

From (9) we have that $I_m - \mathsf{K}_t B \to \Omega_B(V + \Omega_B)^{-1}$, which has positive eigenvalues less than 1, and is hence stable. Moreover ε_{t+1} is i.i.d. and the AR coefficient $I_p - \mathsf{K}_t B$ is stable for large t, so we find that (25) determines an asymptotically stationary process for $T_t - \mathsf{E}_t T_t$.

Proof of (b)*:* We define $u_t = BT_t - \hat{\mathsf{E}}_t\hat{B}T_t$, and find

$$\mathsf{E}_t BT_t - \hat{\mathsf{E}}_t\hat{B}T_t = (\mathsf{E}_t BT_t - BT_t) + (BT_t - \hat{\mathsf{E}}_t\hat{B}T_t) = B(\mathsf{E}_t T_t - T_t) + u_t.$$

The first term is stationary by *(a)*, such that it is enough to show that u_t is asymptotically stationary. From (23), (24), and (25) we find, using estimated values of parameters,

$$y_{t+1} - \hat{\mathsf{E}}_t y_{t+1} = BT_t + \varepsilon_{t+1} - \hat{\mathsf{E}}_t\hat{B}T_t = u_t + \varepsilon_{t+1},$$

and the recursion

$$u_{t+1} = (I_p - \hat{B}\hat{\mathsf{K}}_t)u_t - \hat{B}\hat{\mathsf{K}}_t\varepsilon_{t+1} + B\eta_{t+1}. \tag{26}$$

From the definition of u_t, we find from (16), that

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

Next we multiply (26) by $\overline{\hat{B}}' = (\hat{B}'\hat{B})^{-1}\hat{B}'$ and use $\overline{\hat{B}}'B = \overline{\hat{B}}'(B-\hat{B}) + I_m = I_m + O_P(n^{-1/2})$, to find

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

because both $\overline{\hat{B}}'(B - \hat{B})\eta_{t+1}$ and $(I_m - \hat{\mathsf{K}}_t)\hat{B}_\perp'\overline{\hat{B}}_\perp'u_t$ are $O_P(n^{-1/2})$. From (9) we see that $I_m - \hat{\mathsf{K}}_t\hat{B} \overset{\mathrm{P}}{\to} \Omega_B(V + \Omega_B)^{-1}$ is stable for large n and t, which shows that $\overline{\hat{B}}'u_{t+1}$ and hence u_t is asymptotically stationary.

Proof of (c)*:* We apply previous results and find from the identity

$$\hat{B}(T_t - \hat{E}_t T_t) = (\hat{B} - B)T_t + (BT_t - \hat{B}\hat{E}_t T_t),$$

that $BT_t - \hat{B}\hat{E}_t T_t$ is asymptotically stationary by (b). For the second term we find $n^{1/2}(\hat{B} - B) = O_P(1)$ and $n^{-1/2}T_{[nu]} \overset{\mathrm{D}}{\to} W_\eta$, see (22), such that $(\hat{B} - B)T_t$ is not necessarily asymptotically stationary. Hence $T_t - \hat{E}_t T_t$ is not in general a stationary process. ∎

Proof of Theorem 3. *Proof of* (a)*:* This follows from (4) and (5).

Proof of (b)*:* We compare each estimator with the corresponding trend and find

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

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

Proof of (c)*:* We find similarly

$$\hat{T}_t - \breve{T}_t^* = (\hat{T}_t - T_t) + (T_t - T_t^*) + (T_t^* - \breve{T}_t^*), \tag{27}$$

where $T_t - T_t^*$ is asymptotically stationary by (a). For the first term, we decompose as

$$\hat{B}(T_t - \hat{T}_t) = (\hat{B} - B)T_t + (BT_t - \hat{B}\hat{T}_t),$$

where $(\hat{B} - B)T_t$ is asymptotically nonstationary, because $\hat{B}$ is in general only $n^{1/2}-$consistent. The same argument applies to the last term in (27), which shows that in general we do not get asymptotic stationarity. ∎

7 References

Chan, S.W., Goodwin, G.C. Sin, K.W.S. (1984). Convergence properties of the Ricatti difference equation in optimal filtering of nonstabilizable systems. *IEEE Transaction on Automatic Control*, Vol AC-29, 2, 110–118.

Durbin, J. and Koopman, S.J. (2012). *Time Series Analysis by State Space Methods,* 2nd ed. Oxford University Press, Oxford.

Harvey, A. (1989). *Forecasting. Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Hoover, K. D., Johansen, S., Juselius, K., and Tabor, M.N. (2014). Long-run Causal Order A Progress Report.

Johansen, S., (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, 2nd ed.* Oxford University Press, Oxford.

Johansen, S., and Juselius, K. (2014). An asymptotic invariance property of the common trends under linear transformations of the data. *Journal of Econometrics* 178, 310–315.

Koopman, S. J., Shepard, N., and Doornik, J. A. (2008). *Statistical Algorithms for Models in State Space Form: SsfPack 3.0.* Timberlake Consultants Press.

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

Saikkonen, P., Lütkepohl, H. (1996). Infinite order cointegrated vector autoregressive processes. Estimation and Inference. *Econometric Theory* 12, 814–844.

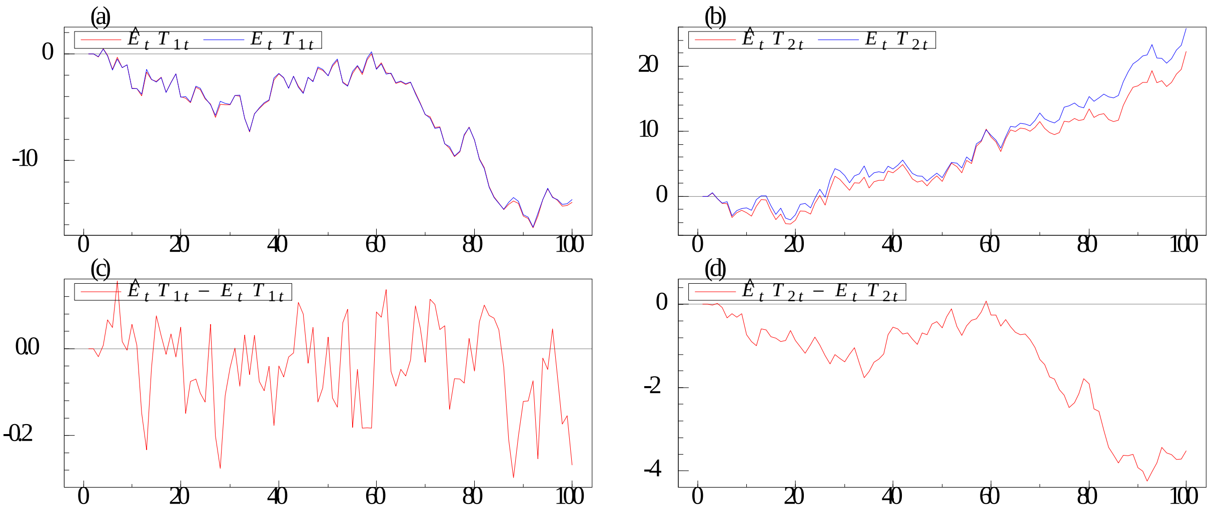

Figure 1: The result of a simulation of model (17) are plotted. Panels a and b show plots of $\hat{E}_t T_{1t}$ and $E_t T_{1t}$, and $\hat{E}_t T_{2t}$ and $E_t T_{2t}$ respectively. We note that in both cases, the processes seem to co-move. In panel c we have plotted $\hat{E}_t T_{1t} - E_t T_{1t}$ which appears stationary, but in panel d we note that the spread $\hat{E}_t T_{2t} - E_t T_{2t}$ is nonstationary. This is accordance with the finding in Table 1, where it is seen that $\hat{E}_t T_{1t}$ is needed to find cointegration, see (18).

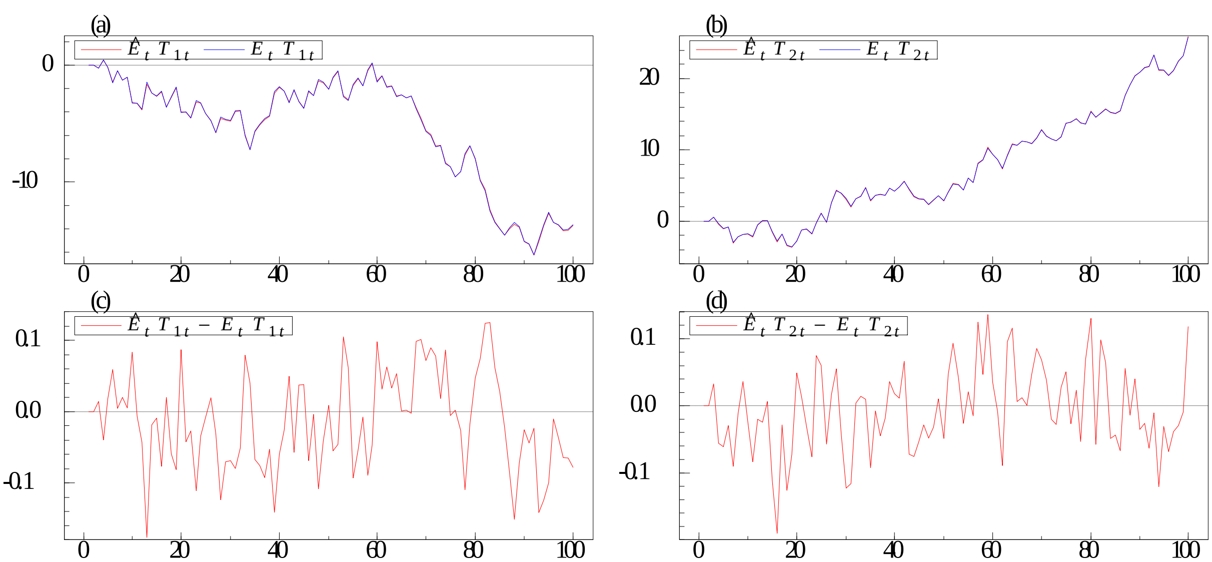

Figure 2: The results of a simulation of model (20) are plotted. Panels a and b show plots of $\hat{E}_t T_{1t}$ and $E_t T_{1t}$, and $\hat{E}_t T_{2t}$ and $E_t T_{2t}$ respectively. We note that in both cases, the processes seem to co-move. In panels c and d, we have plotted $\hat{E}_t T_{1t} - E_t T_{1t}$ and $\hat{E}_t T_{2t} - E_t T_{2t}$, which now appear stationary, because they are both recovered from $\hat{E}_t \hat{A} T_t - E_t A T_t$ as the first two coordinates see (19).