



## Time-varying cointegration and the Kalman filter

Burak Alparslan Eroğlu, J. Isaac Miller & Taner Yiğit

**To cite this article:** Burak Alparslan Eroğlu, J. Isaac Miller & Taner Yiğit (2022) Time-varying cointegration and the Kalman filter, *Econometric Reviews*, 41:1, 1-21, DOI: [10.1080/07474938.2020.1861776](https://doi.org/10.1080/07474938.2020.1861776)

**To link to this article:** <https://doi.org/10.1080/07474938.2020.1861776>



Published online: 15 Jan 2021.



Submit your article to this journal [↗](#)



Article views: 348



View related articles [↗](#)



View Crossmark data [↗](#)



# Time-varying cointegration and the Kalman filter

Burak Alparslan Eroğlu<sup>a</sup>, J. Isaac Miller<sup>b</sup>, and Taner Yiğit<sup>c</sup>

<sup>a</sup>Department of Economics, İstanbul Bilgi University, Eyüpsultan, İstanbul, Turkey; <sup>b</sup>Department of Economics, University of Missouri, Columbia, Missouri, USA; <sup>c</sup>Department of Economics, Bilkent University, Bilkent, Ankara, Turkey

## ABSTRACT

We show that time-varying parameter state-space models estimated using the Kalman filter are particularly vulnerable to the problem of spurious regression, because the integrated error is transferred to the estimated state equation. We offer a simple yet effective methodology to reliably recover the instability in cointegrating vectors. In the process, the proposed methodology successfully distinguishes between the cases of no cointegration, fixed cointegration, and time-varying cointegration. We apply these proposed tests to elucidate the relationship between concentrations of greenhouse gases and global temperatures, an important relationship to both climate scientists and economists.

## KEYWORDS

Climate change; Kalman filter; spurious regression; time-varying cointegration

## JEL CLASSIFICATION

C12; C32; C51; Q54

## 1. Introduction

The analysis of cointegration has been a key issue in econometrics since the 1980s, especially for empirical macroeconomists investigating the long-run relations between key economic variables. A crucial underlying assumption in many of these analyses is the constancy of the relation between cointegrated economic variables. However, when one deals with such long spans of time, the expectation of a stable relation between any variable seems to be an assumption that is difficult to satisfy.

That is why a branch of the econometrics literature focused on various forms of changing relationships between cointegrated variables – or cointegration vector instability. While some dealt with smoothly varying cointegrating vectors [Fourier flexible form functions in Park and Hahn (1999); Chebyshev polynomials in Bierens and Martins (2010)], some dealt with structural breaks in the cointegrating vector (Hansen and Johansen, 1999; Seo, 1998), some examined them as functions of covariates (Xiao, 2009), and others allowed for stochastic freedom to the cointegration vector [via bi-integrated processes as in Hansen (1992), Harris et al. (2002), and McCabe et al. (2006)].

Despite its advantage and popularity in analyzing time-varying parameter models in the stationary realm, the state-space modeling and the Kalman Filter (KF) technique seem to have been overlooked in the literature on cointegration vector instability. A possible reason for such an omission compared to the case of cointegration with constant parameters is that error is assumed to be Gaussian (e.g., Hamilton, 1994, pp. 399–400), which is much more restrictive than the case with constant parameters. A possible reason compared to either the stationary case or the time-varying cointegration cases discussed above is the intractability of asymptotic analysis. Asymptotic analyses of the latter, of state-space models with time-varying parameters of Stoffer

and Wall (1991), Hamilton (1994), *inter alia*, rely on an assumption of Ljung and Caines (1980) that rules out unit roots.

In fact, we show using a numerical example and analytically that when a state-space model with time-varying coefficients (which we will call TVSSM) is spurious because the error term contains a unit root, the Kalman filter suppresses the nonstationarity in the error, and it re-emerges in the estimated state variable, making the problem difficult to detect. This may cut either way in applied work: identification of a time-varying relationship may be driven by a spurious regression or a legitimate relationship that varies over time may be dismissed mistakenly as being spurious.

We devise an estimation and testing methodology that utilizes the KF in the context of a TVSSM. Such an approach holds an advantage over existing methods because it enables estimation of cointegrating regressions with coefficients that are constant or stochastically varying. A contribution of our study is to enable the use of the KF in the presence of integrated variables without the risk of a spurious relation. We offer a very simple yet very effective solution with the addition of another state equation controlling the persistence level in the measurement error, extending KF to wider set of applications.

A related study in the literature is that of Chang et al. (2009), in which the authors derive the asymptotics of the ML estimator when estimating a stochastic trend using the KF. However, their study estimates a constant cointegrating vector while we focus on a potentially time-varying one. Our extension is far from trivial, because even though estimation is not much more complicated, time-varying coefficients prohibit conditional variances in the filter from reaching steady states, rendering the asymptotic analysis intractable or nearly so.

Instead of asymptotic analysis, we propose a bootstrap methodology motivated by that of Stoffer and Wall (1991) for TVSSMs, but which is adapted to accommodate unit roots along the lines of Chang and Park (2003). Extensive simulations and an application to the long-run relationship between anthropogenic greenhouse gas concentrations and global temperatures illustrate the utility of our approach to distinguish between a cointegrating regression with fixed coefficient, a time-varying cointegrating relationship, and no cointegration (a spurious relationship).

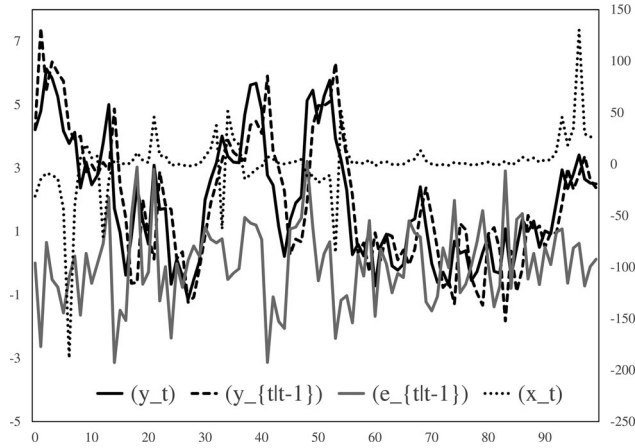
In Section 2, we start our analysis by illustrating how the KF is susceptible to finding a spurious time-varying cointegration relation when applied to unrelated integrated variables and we introduce our modification of the underlying model to make it robust to this case. We summarize our proposed bootstrap-based testing techniques in Section 3. We display finite sample results of the tests in Section 4 and conduct our empirical application in Section 5. Section 6 concludes, and two appendices contain additional tables and selected technical results.

## 2. State-space model with time-varying coefficients

### 2.1. An example of spurious vs. time-varying cointegrating regression

In many studies that use the KF, model specification, especially for the transition equation, has been determined without too much care about the (non)stationarity of the variables in the analysis (Canarellsal et al., 1990; Cooley and Prescott, 1976; Engle and Watson, 1987; Evans, 1991; Kim, 1993, 2006). For instance, Canarellsal et al. (1990) show that using time-varying coefficients in the relationship between the exchange rate and relative prices helps demonstrate that the relationship holds relatively well between the currencies of five developed economies. However, a later comment by Honohan (1993) shows that such a depiction of time-varying coefficients in the presence of integrated variables can be problematic, and more specifically the lack of cointegration may go unnoticed by utilizing time-varying specification for the cointegrating vector.

This problem can be illustrated with a simple simulation exercise. Assume a simple TVSSM setup:



**Figure 1.** Independent I(1) series. Two independent I(1) series,  $(x_t)$  and  $(y_t)$ , the in-sample prediction  $(y_{t|t-1})$  and fitted residuals  $(e_{t|t-1})$  estimated using a simple TVSSM with a sample size of 100.

$$\begin{aligned} y_t &= \alpha + x_t \beta_t + \varepsilon_t \quad t = 1, \dots, N \\ \beta_t &= \mu + T \beta_{t-1} + \eta_t \end{aligned}$$

The scalar series  $(y_t)$  and  $(x_t)$  are both I(1) and  $\beta_t$  is the time-varying coefficient (state variable) that governs the relationship between the two. The error term  $\varepsilon_t$  is implicitly defined and may have serial correlation.<sup>1</sup> When the two observed variables are independent I(1) series, the correct value for  $\beta_t$  should be zero for all  $t$ , and the error term  $\varepsilon_t$  must be I(1). In this simple simulation, we set  $\alpha = 0$  and simulate  $(y_t)$  and  $(x_t)$  to be independent random walks, so that  $\beta_t = 0$  for all  $t$ .

We may write  $\beta_t = (1 - T)^{-1} \mu + (\beta_0 - (1 - T)^{-1} \mu) T^t + \sum_{i=0}^{t-1} T^i \eta_{t-i}$  for  $|T| < 1$ . In order for  $\beta_t = 0$  for all  $t$ , we need  $\mu = 0$ . If  $\beta_0 = 0$  and  $\sigma_\eta = \text{var}(\eta_t)^{1/2} = 0$ ,  $T$  need not be zero. We estimate  $T$ ,  $\sigma_\varepsilon = \text{var}(\varepsilon_t)^{1/2}$ , and  $\sigma_\eta$  by maximum likelihood, with  $\alpha$  restricted to its true value of zero and with  $\mu = \beta_0 = 0$  likewise restricted. Under these restrictions, we may rewrite the model to be estimated as

$$y_t = x_t \left( \sum_{i=1}^{t-1} T^i \eta_{t-i} + \eta_t \right) + \varepsilon_t$$

and we expect that  $\hat{\sigma}_\eta \approx 0$ . If not, we expect  $\hat{T} \approx 0$  to eliminate the first term in parentheses.

Our parameter estimates are  $\hat{\sigma}_\varepsilon = 0.73$ ,  $\hat{T} = 0.92$ , and  $\hat{\sigma}_\eta = 0.10$ , and graphical results are shown in Fig. 1. The value of  $\hat{\sigma}_\varepsilon$  is not important to our example, and the figure suggests that it is estimated reasonably. What is important to note is that  $\hat{\sigma}_\eta$  is not close to zero, but neither is  $\hat{T}$ . Instead,  $\hat{T}$  is close to 1! Instead of estimating  $\beta_t = 0$  for all  $t$ , these parameter values lead  $\beta_t$  to be highly flexible, fitting two unrelated series very closely to each other in a spurious fashion.

The cause of the problem is that the fitted residual is forced to be I(0), as the figure suggests, causing all the nonstationarity to be transferred to the time-varying parameter. This is shown mathematically in Appendix B. The Kalman filter adjusts the state variable (time-varying cointegration vector) heavily using the forecast error, leading to a spurious relation between the two independent and integrated series. Bayesian estimation of the above setup also gives similar results unless the priors for  $\sigma_\eta$  are restricted significantly, indicating that the problem is not specific to the Kalman estimation technique.

<sup>1</sup>We also run the simulations with the serially correlated error terms. However, these simulations do not produce a different results than what we are presenting. To keep the exposition simple and brief, we only demonstrate the results with the serially uncorrelated error terms.

Does allowing the intercept to vary over time help with this problem? Yes and no. For independent I(1) series ( $y_t$ ) and ( $x_t$ ), set we could set  $\beta_t = \beta$  and then  $\alpha_t$  would take on a unit root to estimate  $y_t - x_t\beta$  so that  $(\varepsilon_t)$  would become I(0). The regression model is no longer spurious in the sense of having an I(1) error, but it is just as spurious in the sense of falsely estimating a relationship between two independent I(1) series. No value of  $\beta$  could be identified, because  $\alpha_t$  would adjust to accommodate any value of  $\beta$ . Nor should any value of  $\beta$  be identified by such a model. What is important, however, is the persistence of the time-varying intercept, which identifies whether or not the model is spurious.

We leverage the intuition about a time-varying intercept to test for cointegration. This can be accomplished in one of two ways. Either we keep  $(\varepsilon_t)$  stationary and allow  $\alpha$  to vary over time, as just discussed, with the possibility of a unit root, or we keep  $\alpha$  fixed and allow for the possibility of a unit root in  $(\varepsilon_t)$ . In either case, a unit root in the sum of the intercept and error term reveals the presence of a spurious relationship. In our analysis below, we fix the intercept and allow for a unit root in the error. This convention avoids the additional error introduced by allowing the intercept to be stochastic whether or not it contains a unit root.

## 2.2. General state-space model

The aforementioned risk of finding a spurious relationship in using unrelated I(1) series and the Kalman filter necessitates a modification that makes the filter robust to this problem. We suggest replacing the measurement equation error with a new state variable. The new state variable helps us differentiate between an existing cointegration relationship or lack thereof by estimating the degree of persistence in the measurement error. Such a modification allows us not only to represent different types of cointegration in the state-space setup, but it also provides us the ability to test for their existence.

Consider the following TVSSM, a generalization of that above that allows for higher-order serial correlation and a vector of regressors:

$$y_t = \alpha + x_t' \beta_t + w_t \quad t = 1, \dots, N \quad (1)$$

$$\beta_t = \mu + T\beta_{t-1} + \eta_t \quad (2)$$

where  $x_t$  is a  $p \times 1$  vector of regressors, which we assume are not mutually cointegrated. The series  $(\beta_t)$  is comprised of  $p \times 1$  vectors of time-varying coefficients, and  $\alpha$ ,  $\mu$ , and  $T$  are model parameters of dimensions  $1 \times 1$ ,  $p \times 1$ , and  $1 \times 1$ , respectively.

The scalar error term  $w_t$  can be represented as an autoregressive process given by

$$w_t = \theta w_{t-1} + \sum_{i=1}^k \delta_i \Delta w_{t-i} + \varepsilon_t \quad t = 1, \dots, N \quad (3)$$

and we assume that  $\varepsilon_t$  and  $\eta_t$  are jointly i.i.d. random variables with zero mean, zero covariance, and variances given by  $\sigma_\varepsilon^2$  (scalar) and  $\Sigma_\eta$  ( $p \times p$  matrix), respectively, and that  $|\delta_k| < \infty$  for  $k < \infty$ , or, in the limit,  $\sum \delta_k^2 < \infty$  as  $k \rightarrow \infty$ . Further discussion of the model proceeds as if  $k$  is known, but of course we select  $k$  using the Schwarz–Bayesian Information Criterion (BIC) by evaluating models that vary across  $k$ .

We set the initial condition for  $w_t$  to zero in the implementation of the filter below. In the case of  $\theta = 1$ , the initial condition of  $w_t$  gives the mean  $w_t$ , so that  $\alpha$  cannot be identified unless this initial condition is fixed. On the other hand, if  $|\theta| < 1$ , the effect of the initial condition is negligible in a large sample. The allowance for serial correlation in the error term serves the same purpose – to ameliorate omitted variable bias – as specifying the intercept to vary over time, which we do not do.

While the transition equation in (2) helps us identify the type of cointegration, the autoregressive representation of  $w_t$  determines the existence of cointegration between  $y_t$  and the regressors. Allowing a time-varying intercept would accomplish the same thing, but with an additional error term. Specifically if (i)  $\theta = 1$ ,  $w_t$  contains a single unit root and the observable variables in Eq. (1) are not cointegrated. Whereas, if (ii)  $|\theta| < 1$ ,  $\mu \neq 0$ ,  $T = 0$ , and  $\Sigma_\eta = 0$ , these variables are cointegrated with fixed coefficients. Alternatively, if (iii)  $|\theta| < 1$ ,  $T = 1$  and very small  $\Sigma_\eta$ , we have a smoothly varying cointegration relationship which may mimic that of Park and Hahn (1999). Finally, (iv)  $|\theta| < 1$ ,  $T < 1$ , and  $\Sigma_\eta \neq 0$  imply stochastically varying cointegration.

The sufficient condition that  $T = 0$  for cointegration with fixed coefficients is not necessary if we consider the initial condition  $\beta_0$ . As long as  $|T| < 1$ , we may write  $\beta_t = (1 - T)^{-1}\mu + (\beta_0 - (1 - T)^{-1}\mu)T^t + \sum_{i=0}^{t-1} T^i \eta_{t-i}$ , which shows that  $\mu \neq 0$  allows a non-zero mean of  $\beta_t$ . Setting  $\beta_0 = (1 - T)^{-1}\mu$  allows for estimation of a fixed cointegrating coefficients given by  $(1 - T)^{-1}\mu$  when  $\Sigma_\eta = 0$ . However,  $\beta_t = \beta_0 + \mu t + \sum_{i=0}^{t-1} \eta_{t-i}$  when  $T = 1$ . Non-zero  $\mu$  allows drift and thus non-constancy in the cointegrating coefficients. In that case,  $\mu = 0$  imposes constancy and the cointegrating coefficients become  $\beta_0$ .

In case there is a constant cointegrating relationship, we suggest setting  $\beta_0$  to be  $\beta_N$  from a previous run of the filter with  $\beta_0 = (1 - T)^{-1}\mu$ . If the estimated value of  $T$  stays below unity, as we expect it will when  $\theta = 0$  is not imposed, the constant cointegrating relationship will be given by  $(1 - T)^{-1}\mu$ . On the other hand, if the value of  $T$  approaches unity,  $(1 - T)^{-1}\mu$  explodes, so that the constant cointegrating relationship pushes  $\mu$  to zero.

In this paper, we limit our analysis to testing for the cases of no cointegration, fixed cointegration and (general) time-varying cointegration. For these three scenarios, focusing on the parameters  $\theta$  and  $\Sigma_\eta^{1/2}$  suffices to achieve the desired objectives. Further analysis of specific cases of time-varying cointegration is left for future study.

Defining the state variables  $\xi_t$  by

$$\xi_t = \begin{cases} [\beta'_t & w_t]' & \text{if } k = 0 \\ [\beta'_t & w_t & \Delta w_t & \cdots & \Delta w_{t-k+1}]' & \text{if } k > 0 \end{cases}$$

and defining  $z_t = (x'_t, 1)'$ , the measurement and transition equations of the TVSSM are given by

$$y_t = \alpha + \begin{cases} z'_t \xi_t & \text{if } k = 0 \\ [z'_t & 0 & \cdots & 0] \xi_t & \text{if } k > 0 \end{cases} \quad (4)$$

$$\xi_t = A + F\xi_{t-1} + QE_t \quad (5)$$

where

$$E_t = \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_\eta & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}\right)$$

$$A = \begin{cases} [\mu' & 0]' & \text{if } k = 0 \\ [\mu' & 0 & 0 & \cdots & 0]' & \text{if } k > 0 \end{cases} \quad F = \begin{cases} \begin{bmatrix} T * I_{p \times p} & 0 \\ 0 & \theta \end{bmatrix} & \text{if } k = 0 \\ \begin{bmatrix} T * I_{p \times p} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \theta & \delta_1 & \cdots & \delta_{k-1} & \delta_k \\ 0 & \theta - 1 & \delta_1 & \cdots & \delta_{k-1} & \delta_k \\ 0 & 0 & 1 & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & 1 & 0 \end{bmatrix} & \text{if } k > 0 \end{cases}$$

and

$$Q = I_{(p+1) \times (p+1)} \quad \text{if } k = 0 \quad \text{or} \quad Q = \begin{bmatrix} I_{(p+1) \times (p+1)} & & & \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \quad \text{if } k > 0$$

Let the unknown parameter vector be

$$\Psi = \begin{cases} [\alpha, \sigma_e^2, T, \mu', \text{vech}(\Sigma_\eta^{1/2})', \theta]' & \text{if } k = 0 \\ [\alpha, \sigma_e^2, T, \mu', \text{vech}(\Sigma_\eta^{1/2})', \theta, \delta']' & \text{if } k > 0 \end{cases}$$

with  $\delta = (\delta_1, \dots, \delta_k)'$  and  $\text{vech}(\Sigma_\eta^{1/2})$  denoting the half-vectorization of the lower Cholesky decomposition of  $\Sigma_\eta$ , so that we have  $q = 4 + p(3 + p)/2 + k$  unknown parameters. By utilizing maximum likelihood (ML) estimation of the TVSSM defined in Eqs. (4) and (5), we can obtain the estimates for the parameter vector  $\Psi$ . We denote the ML estimates obtained using the KF by  $\hat{\Psi} = (\hat{\alpha}, \hat{\sigma}_e^2, \hat{T}, \hat{\mu}', \text{vech}(\hat{\Sigma}_\eta^{1/2})', \hat{\theta}, \hat{\delta}')'$  with the convention that  $\text{vech}(\hat{\Sigma}_\eta^{1/2})$  denotes the estimate of  $\text{vech}(\Sigma_\eta^{1/2})$ . The likelihood function and implementation of the KF are standard, and we describe them in [Appendix B](#).<sup>2</sup>

As with any state-space model, the Kalman filter is derived under the assumption of Gaussian errors. Departures from Gaussian errors mean that the updating equation no longer provides the optimal forecast of the state given contemporaneous information about the observation. However, the Kalman gain still retains its interpretation as a projection of the state onto the observation.<sup>3</sup>

In the case of a state-space model with constant parameters, the Kalman gain is not a function of data and a steady-state may be obtained, in which case the assumption on Gaussian errors may be relaxed to allow pseudo-ML estimation. In the nonlinear case, we cannot say that the linear projection provides the optimal forecast, and in general, the variances do not reach a steady state, rendering asymptotic analysis intractable, as we explain in [Appendix B](#).

In light of the complication of nonlinearity and possibly non-Gaussianity, we propose a semi-parametric bootstrap procedure for our hypothesis tests. Redrawing from the empirical error distribution both avoids the need for a formal asymptotic analysis of the ML-based estimator and should ameliorate departures from Gaussian errors. As an additional check, we also test for and do not find significant departures from normality in our empirical application.

### 3. Bootstrap design

#### 3.1. General algorithm

Consider the following general null hypothesis with  $r$  restrictions:  $H_0 : R\Psi = m_0$  where  $R$  is a  $r \times q$  restriction matrix and  $m_0$  is a known  $r \times 1$  vector. To test this null hypothesis, we use the following bootstrapping algorithm:

<sup>2</sup>Loosening the assumption that  $T$  is a scalar and allowing it to be estimated as a  $p \times p$  matrix may be accommodated by replacing  $T * I_{p \times p}$  with  $T$  in the definition of  $F$  and  $T$  with  $\text{vec}(T)$ , the vectorization of  $T$ , in the definition of  $\Psi$ . We do not consider this possibility further in this paper.

<sup>3</sup>A general formulation of the Kalman gain is given by  $\text{cov}(\xi_t, y_t | \mathcal{F}_{t-1}, x_t) \text{var}(y_t | \mathcal{F}_{t-1}, x_t)^{-1}$  where  $\mathcal{F}_{t-1}$  is the information set driven by  $y_{t-1}$ . Hamilton (1994) reminds us that  $(\xi_t', y_t')' | \mathcal{F}_{t-1}, x_t$  (pg. 399) is Gaussian and uses the well-known formula for updating a Gaussian forecast (pg. 102) to argue that the Kalman gain weights the new information optimally. Without Gaussian errors, the Kalman gain, expressed as  $\Omega_{21}\Omega_{11}^{-1}$  in the notation of Hamilton (1994) on page 102, still retains its interpretation as a projection, but optimality is no longer assured.

1. Estimate the unrestricted TVSSM by ML estimation of the full set of parameters  $\Psi$ . Denote the estimated parameter vector by  $\hat{\Psi}$  as above.
2. Compute the appropriate test statistic  $\hat{\tau}_R(\Psi)$  associated with the proposed null hypothesis.
3. Estimate the restricted TVSSM by ML estimation with the null restriction imposed,  $R\Psi = m_0$ . Denote the estimated parameter vector by  $\hat{\Psi}_R$  and predicted values of the state vector using restricted parameter estimates in the Kalman recursions in Eq. (B2) of [Appendix B](#) by  $\check{\zeta}_{R,t|t-1} = (\beta'_{R,t|t-1}, w_{R,t|t-1})'$ .<sup>4</sup>
4. Construct the fitted residuals  $\varepsilon_{R,t|t-1}$  from [Eq. \(3\)](#) using restricted parameter estimates and the series  $(w_{R,t|t-1})$ .
5. Re-sample the demeaned innovations,  $\varepsilon_{R,t|t-1} - \bar{\varepsilon}_R$ , with replacement to get  $\varepsilon_t^{(b)}$ .
6. Construct  $w_t^{(b)}$  using restricted parameter estimates and  $\varepsilon_t^{(b)}$  in [Eq. \(3\)](#) and  $y_t^{(b)}$  using restricted parameter estimates,  $\beta_{R,t|t-1}$ , and  $w_t^{(b)}$  in (1).
7. Estimate the unrestricted TVSSM by ML estimation with the observations  $x_t$  and  $y_t^{(b)}$ .
8. Compute the test statistic as  $\hat{\tau}_R^{(b)}(\Psi)$  from the ML estimates in step 7.
9. Repeat steps 5–8 for  $b = 1, \dots, B$  to obtain an empirical distribution for the test statistic  $\hat{\tau}_R(\Psi)$ . The bootstrap  $p$ -value of the test statistic can be obtained from this empirical distribution.

Two important caveats must be made. First, in contrast to the bootstrap procedure of Stoffer and Wall (1991), our bootstrap draws from estimated innovations of the TVSSM itself rather than from the innovations of the KF used to estimate it. More closely following their procedure would entail constructing the scalar Kalman innovation given by

$$\begin{aligned} \varepsilon_{R,t|t-1} &= \sigma_{R,t|t-1}^{-1/2} (y_t - y_{t|t-1}) \\ &= \sigma_{R,t|t-1}^{-1/2} ((\alpha - \hat{\alpha}_R) + x'_t(\beta_t - \beta_{R,t|t-1}) + (w_t - w_{R,t|t-1})) \end{aligned}$$

in step 4 and then constructing  $\check{\zeta}_t^{(b)}$  and  $y_t^{(b)}$  from

$$\begin{bmatrix} \check{\zeta}_{R,t+1|t} \\ y_t \end{bmatrix} = \begin{bmatrix} \hat{A}_R \\ \hat{\alpha}_R \end{bmatrix} + \begin{bmatrix} \hat{F}_R & 0 \\ \check{z}'_t & 0 \end{bmatrix} \begin{bmatrix} \check{\zeta}_{R,t|t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{F}_R \Omega_{R,t|t-1} z_t \sigma_{R,t|t-1}^{-1} \\ 1 \end{bmatrix} \sigma_{R,t|t-1}^{1/2} \varepsilon_t^{(b)}$$

in step 6, where  $\Omega_{R,t|t-1}$  and  $\sigma_{R,t|t-1}$  are the conditional variances of  $(\check{\zeta}_t - \check{\zeta}_{R,t|t-1})$  and  $(y_t - y_{R,t|t-1})$ . Nonstationarity undermines a key assumption made by those authors and extensive simulations by the present authors suggest that size distortion is considerably worse when bootstrapping the Kalman innovation rather than the model innovation.

Second, our bootstrap draws only from  $(w_t - w_{R,t|t-1})$  but not from  $(\beta_t - \beta_{R,t|t-1})$ . In other words,  $\beta_{R,t|t-1}$  is held constant in repeated draws, which implies that  $\mu$ ,  $T$ , and  $\Sigma_\eta$  are held constant at the null value under both the null and alternative – even when the test involves a hypothesis about  $\Sigma_\eta$ ! This seems counter-intuitive, but we explain it in more detail below.

### 3.2. Testing for cointegration

In order to test for cointegration, we focus on the parameter  $\theta$  in the TVSSM. As mentioned earlier,  $\theta=1$  indicates the lack of a cointegrating relationship between the observable variables. Consequently, our null hypothesis is  $H_0 : \theta = 1$  against the alternative  $H_a : \theta < 1$ . To conduct

<sup>4</sup>Throughout the paper, we use the notation  $Z_R$  as  $Z$  is an object obtained by imposing the restriction depicted in the null  $H_0 : R\Psi = m_0$ . Moreover,  $Z_{t|s}$  is standard notation in the Kalman filtering literature for the expectation of  $Z_t$  conditional on information available at time  $s$ .



this hypothesis test, we can use a standard  $t$ -test with a  $t$ -statistic given by  $\hat{\tau}_R(\Psi) = t_\theta = (\hat{\theta} - 1)/se(\hat{\theta})$  where  $se(\hat{\theta})$  is the standard deviation<sup>5</sup> of the ML estimator of  $\theta$ .

We obtain critical values for this test using the bootstrap routine above. We apply each step in the algorithm with the following modifications:

3\* Same as step 3, but impose the restriction  $\theta = 1$  by setting  $m_0 = 1$  and  $R$  equal to a  $1 \times q$  row vector of zeros with a one in the column corresponding to  $\theta$  in the vector  $\Psi$ .

8\* Compute the bootstrap  $t$ -statistic as  $t_\theta^{(b)} = (\hat{\theta}^{(b)} - 1)/se(\hat{\theta}^{(b)})$ .

The rest of the steps are the same as in the general algorithm described above. Because  $\theta = 1$ , steps 4 and 6 are accomplished using

$$\Delta w_{R,t|t-1} = \sum_{i=1}^k \delta_{Ri} \Delta w_{R,t-i|t-1-i} + \varepsilon_{R,t|t-1},$$

in a manner similar to the sieve bootstrap of Chang and Park (2003).

Now let  $\nu$  be the significance level. After obtaining a bootstrap empirical distribution, we can calculate the critical value  $CV_\nu(t_\theta^{(b)})$  as the  $\nu$ -th quantile of this distribution. Therefore, we reject the null hypothesis if  $t_\theta < CV_\nu(t_\theta^{(b)})$ .

### 3.3. Testing for parameter constancy

We now describe the procedure for testing the constancy of a single coefficient, for example,  $\beta_{1t}$ . A joint test for multiple elements of  $\beta_t$  is a special case of the Wald test discussed below. We can impose the constancy of this coefficient by setting  $\sigma_1^2 = 0$ . Consequently, we test the null hypothesis  $H_0 : \sigma_1^2 = 0$  against the alternative  $H_a : \sigma_1^2 > 0$ . For this case, we utilize a right-sided  $t$ -test with a  $t$ -statistic given by  $\hat{\tau}_R(\Psi) = t_{\sigma_1^2} = \hat{\sigma}_1^2/se(\hat{\sigma}_1^2)$ .

We use the algorithm above with the following modifications for this test:

3\*\* Same as step 3, but impose the restriction  $\sigma_1^2 = 0$ , setting  $m_0 = 0$  and  $R$  equal to a  $1 \times q$  row vector of zeros with a one in the column corresponding to  $\sigma_1^2$  in the vector  $\Psi$ .

8\*\* Compute the bootstrap  $t$ -statistic as:  $t_{\sigma_1^2}^{(b)} = (\hat{\sigma}_1^{(b)2})/se((\hat{\sigma}_1^{(b)2})$ .

We calculate the critical value  $CV_{(1-\nu)}(t_{\sigma_1^2}^{(b)})$  with  $\nu$  significance level from the bootstrap empirical distribution. Because the test is right-sided, the rejection rule becomes  $t_{\sigma_1^2} > CV_{(1-\nu)}(t_{\sigma_1^2}^{(b)})$ .

What seems unusual in this procedure is that we hold  $\sigma_1^2$  (more generally  $\Sigma_\eta$ ) constant under both the null and alternative, as we note above. How can such a test have power against a false null? The reason lies in the autoregressive coefficient matrix in the Kalman recursions in Eq. (B3) of Appendix B. Both  $(\beta_t - \beta_{R,t|t-1})$  and  $(w_t - w_{R,t|t-1})$  depend on *both*  $\eta_t$  and  $\varepsilon_t$ , which means that *both* series of fitted residuals have discriminatory power to distinguish between  $\sigma_1^2 = 0$  and  $\sigma_1^2 > 0$  in the state-space model. So, even though  $\beta_t^{(b)} = \beta_{R,t|t-1}$  is fixed in repeated bootstrap samples, variations in  $w_t^{(b)} - w_{R,t|t-1}$  are enough to discriminate against a false null.

### 3.4. Joint test

Now we focus on a joint hypothesis testing and consider the joint null:  $H_0 : \theta = 1, \text{vech}(\Sigma_\eta^{1/2}) = 0$ . Since we have multiple restrictions under this hypothesis, we use the Wald test, and we construct the test statistic as follows:

$$W_{\theta, \Sigma_\eta} = (\text{vech}(\hat{\Sigma}_\eta^{1/2})', \hat{\theta} - 1) \left[ \text{var}(\text{vech}(\hat{\Sigma}_\eta^{1/2})', \hat{\theta}) \right]^{-1} (\text{vech}(\hat{\Sigma}_\eta^{1/2})', \hat{\theta} - 1)'$$

<sup>5</sup>The variance of the parameter estimators can be obtained from the inverse of the Hessian matrix of the likelihood evaluated at the estimated parameter values.

where  $\text{var}(\text{vech}(\hat{\Sigma}_\eta^{1/2})', \hat{\theta})'$  can be computed from Hessian matrix of the log-likelihood evaluated at the ML estimates. The bootstrap algorithm for this test is the same as the general procedure with the following modifications:

3\*\*\* Same as step 3, but impose the restrictions  $\theta = 1$  and  $\text{vech}(\Sigma_\eta^{1/2}) = 0$ , setting  $m_0 = (0, \dots, 0, 1)'$  and  $R$  equal to a  $(p+1) \times q$  row vector of zeros with a  $p \times p$  identity matrix in the first  $p$  rows and in the  $p$  columns corresponding to  $\Sigma_\eta$  and a one in the last row and column corresponding to  $\theta$ .

8\*\*\* Compute the bootstrap Wald statistic as:

$$W_{\theta, \Sigma_\eta}^{(b)} = (\text{vech}((\hat{\Sigma}_\eta^{(b)})^{1/2})', \hat{\theta}^{(b)} - 1) \left[ \text{var}(\text{vech}((\hat{\Sigma}_\eta^{(b)})^{1/2})', \hat{\theta}^{(b)})' \right]^{-1} (\text{vech}((\hat{\Sigma}_\eta^{(b)})^{1/2})', \hat{\theta}^{(b)} - 1)'$$

Notice that the Wald test is a right-sided test in which we reject the null for large values of the test statistic. Given the significance level  $\nu$ , we reject the null hypothesis if  $W_{\theta, \Sigma_\eta} > CV_{(1-\nu)}(W_{\theta, \Sigma_\eta}^{(b)})$ , where  $CV_{(1-\nu)}(W_{\theta, \Sigma_\eta}^{(b)})$  is the  $(1-\nu)$  quantile of the empirical bootstrap distribution found from the above steps.

The Wald test comes with a caveat. As noted above,  $(w_t - w_{R,t|t-1})$  contains information about both  $\theta$  and  $\Sigma_\eta$ . Yet  $(w_t - w_{R,t|t-1})$  is a scalar series. As a result, we expect a rank degeneracy in the Hessian matrix which could lead to a degenerate limiting distribution of the Wald statistic under the null. We discard iterations of the bootstrap for which the Hessian fails to invert. Simulation results (not shown) suggest that failure to invert is more problematic as  $\Sigma_\eta$  increases and as the sample size increases, which is consistent with such an asymptotic degeneracy.

#### 4. Finite sample performance

We conduct Monte Carlo simulations to check the finite sample performance of the proposed test and bootstrapping routine. In our simulations, we use  $M=10,000$  Monte Carlo replications and two different sample sizes,  $N=100, 500$ . In practice, we recommend setting  $B=999$ , e.g., and implementing a standard bootstrap, so that the test is replicated  $B+1=1,000$  times, as we do in our empirical application. Extending this to the Monte Carlo simulations would entail  $M(B+1)=10,000,000$  replications, which would be prohibitively expensive. Fortunately, using the fast double bootstrap of Davidson and MacKinnon (2007) for the simulations allows for  $B=1$ , so that the tests are evaluated only  $M(B+1)=20,000$  times (See also Trokić, 2019).<sup>6</sup>

We present results with  $N=100$  and relegate those with  $N=500$  to the appendix. We build our simulation exercise using the data-generating process in Eqs. (1)–(3) with  $\alpha=0$ ,  $T=0.7$ ,  $\mu=1/(1-T)=10/3$ , and  $\sigma_\varepsilon^2=1$ , where  $x_t$  is a scalar random walk with unit conditional variance. Unless  $k=0$ , we set  $\delta_i=0.5^i$  for all  $i=1, \dots, k$ .<sup>7</sup>

This setup allows us to consider the scenarios with fixed coefficient cointegration, time-varying cointegration with stochastically varying coefficient, and no cointegration. Specifically, we vary  $\theta$  along  $[0.8, 1]$ , the highest value indicating no cointegration while lower values indicate a long-run relationship between  $y_t$  and  $x_t$ . Lowering the value of  $\theta$  shows power of the test of the null  $H_0: \theta=1$ . Noting that  $\Sigma_\eta = \sigma_\eta^2$  is a scalar, we vary  $\sigma_\eta$  along  $[0, 1]$ . This parameter controls the strength of time variation in the cointegrating vector. The lowest value 0 indicates a fixed cointegrating coefficient while positive values point toward time variation.

<sup>6</sup>We replicated a subset of our results using the more computationally intensive bootstrap (not shown). The rejection rates were not systematically different, but the computational time was measured in weeks for a subset of the results rather than hours for the whole set of results: high cost and no clear benefit.

<sup>7</sup>We pick these values such that the nonstationarity of  $w_t$  stems only from the value of  $\theta_0$ .

**Table 1.** Rejection frequencies with  $N=100$  and  $k=0$ .  $t$ -test of  $H_0 : \theta = 1$  (top panel)  $t$ -test of  $H_0 : \sigma_1 = 0$  (middle panel), and Wald test of  $H_0 : \sigma_1 = 0, \theta = 1$  (bottom panel). Nominal size is 5%.

$\theta_0$	$\sigma_{0,\eta}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.046	0.044	0.054	0.049	0.052	0.054	0.054	0.050	0.052	0.054	0.047
0.99	0.109	0.104	0.124	0.106	0.117	0.111	0.089	0.096	0.087	0.085	0.065
0.95	0.519	0.501	0.491	0.472	0.422	0.416	0.379	0.345	0.316	0.260	0.113
0.90	0.820	0.808	0.763	0.742	0.694	0.646	0.589	0.554	0.517	0.383	0.190
0.80	0.979	0.970	0.941	0.894	0.845	0.793	0.743	0.684	0.637	0.501	0.204
$\theta_0$	$\sigma_{0,\eta}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.047	0.209	0.49	0.685	0.795	0.856	0.921	0.931	0.958	0.971	0.995
0.99	0.048	0.208	0.488	0.71	0.787	0.862	0.907	0.932	0.955	0.975	0.996
0.95	0.052	0.198	0.488	0.673	0.784	0.846	0.899	0.932	0.944	0.969	0.995
0.90	0.046	0.194	0.457	0.642	0.783	0.834	0.895	0.925	0.945	0.964	0.995
0.80	0.049	0.176	0.440	0.619	0.746	0.825	0.874	0.912	0.936	0.965	0.993
$\theta_0$	$\sigma_{0,\eta}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.054	0.102	0.352	0.557	0.722	0.803	0.867	0.905	0.924	0.946	0.964
0.99	0.025	0.047	0.162	0.347	0.551	0.662	0.744	0.818	0.872	0.911	0.942
0.95	0.026	0.032	0.025	0.055	0.139	0.251	0.356	0.514	0.591	0.710	0.877
0.90	0.100	0.123	0.085	0.061	0.071	0.092	0.144	0.185	0.325	0.445	0.721
0.80	0.366	0.396	0.374	0.271	0.182	0.129	0.115	0.122	0.154	0.218	0.513

#### 4.1. Rejection rates: size and power

We compute the rejection rates of three different null hypotheses. The first one is  $H_0 : \theta = 1$  which distinguishes between the existence of a long run cointegrating relationship or lack thereof. The next one is  $H_0 : \sigma_\eta^2 = 0$  which analyses fixed versus time-varying nature of the relation between the regressand and regressor. Both rejection rates are calculated using  $t$ -statistics and the bootstrap distributions described in the previous section. Finally, we have the joint hypothesis test  $H_0 : \theta = 1, \sigma_\eta^2 = 0$  conducted with a Wald statistic and its bootstrapped critical value. All tests are evaluated using a nominal size of 5%.

Table 1 (top panel) displays the results of the single  $t$ -test of the null  $H_0 : \theta = 1$  with  $N=100$  and  $k=0$ . Although we treat  $k=0$  as being known in this table, tables in the appendix that focus on serial correlation reflect selection of  $k$  by BIC, as do our results in the empirical application. The first row where  $\theta = 1$  displays the size of the test, while the remaining rows show power. Size and power calculations are performed for different values of  $\sigma_\eta$ , which are displayed in different columns of the table. Examination of the results demonstrates that our test for cointegration approximates the nominal size well for every value of  $\sigma_\eta$ , even in a relatively small sample. Power is satisfactory when  $\sigma_\eta$  is small. Higher values of  $\sigma_\eta$  decrease the power of the test, because more variation in  $\beta_t$  makes it harder to detect a weak cointegrating relationship – i.e., with a near unit root in the error term.

Table 1 (middle panel) focuses on the parameter  $\sigma_\eta$  and shows results of a single  $t$ -test of the null  $H_0 : \sigma_\eta^2 = 0$  with  $N=100$  and  $k=0$  (known). The first column represents the size, and the columns to the right show power for  $\sigma_\eta > 0$ . Again, the nominal size is satisfactorily approximated and the test is powerful against a wide range of alternative values of  $\sigma_\eta$ . Variation in  $\theta$  does not appear to have very much impact on size or power. Moreover, the power of the test increases very quickly as we move away from the null to the higher values of  $\sigma_\eta$ . Evidently, a time-varying relationship is easily and robustly detected regardless of whether or not  $y_t$  and  $x_t$  are cointegrated, as long as a lack of cointegration is appropriately modeled by estimating  $\theta$ .

The power and size performance of the joint test is illustrated in Table 1 (bottom panel) with  $N=100$  and  $k=0$  (known). For this test, size is defined by the rejection rate when  $\theta = 1$  and  $\sigma_\eta^2 = 0$  in the top left corner of the table. A rate of 0.054 captures the nominal size of 0.050 well.

We expect power to increase as we move away from the null, and this is indeed the case for the alternative  $\theta = 1, \sigma_\eta^2 > 0$  (top row). However, for  $\theta < 1$ , there appears to be a non-monotonicity: power does not necessarily increase for a fixed value of  $\theta$  as we increase  $\sigma_\eta^2$  away from the null, and vice versa. We interpret this result to mean that it is hard to distinguish a spurious relationship from a time-varying cointegrating (but near-spurious) relationship in a relatively small sample. As may be expected, the results for  $N=500$  across all three tests are qualitatively similar to but more powerful than those just discussed (Table A.1 in Appendix A).

Adding serial correlation to the model by increasing  $k$  in (3) does not affect the results very much, as we show in additional tables in the Appendix A. Simulations of the three tests are re-run with  $N=100, 500$  for three different serial dependence scenarios, namely no serial correlation, AR(1), and AR(2). Specifically, Tables A.2 and A.3 show sample sizes of  $N=100, 500$  with orders of serial correlation selected by BIC and then fixed to the BIC value for the bootstraps replications. The results when  $k$  is unknown are largely similar to those with known  $k$ , except power appears to be weaker for higher values of  $k$ .

Looking at Table A.2, we first test the individual null hypotheses,  $\theta = 1$  and  $\sigma_\eta^2 = 0$ , and then the joint test of  $\theta = 1, \sigma_\eta^2 = 0$ . The top line in the top panel for  $H_0 : \theta = 1$  shows us the size of the test. We notice very slight size distortions with higher parameter variability,  $\sigma_{0,\eta}$ . We also note that power of the test is quite strong for constant parameter ( $\sigma_{0,\eta} = 0$ ) and gets stronger with higher lag lengths. Yet, as parameter variability increases, the power performance of the test to declines. Interestingly, the power decline gets weaker with higher lag orders/stronger serial dependence.

The results for  $H_0 : \sigma_\eta^2 = 0$  also show that the test is correctly sized with very minor distortions as the cointegration relation gets stronger ( $\theta_0$  falls) and lag length increases. Power performance of this test is also good as gains are very fast and are affected very little from the complexity of serial dependence. Declines in the value of  $\theta_0$  lead to some power loss, but it is quite negligible. The joint test results are very similar to the ones described with no serial correlation. There again is a potential confusion between parameter variability and spuriousness of the cointegration relation.

The joint test has a difficult time separating high parameter variability from nonstationarity of ( $w_t$ ) (spurious relation) as we note that the combination of both violations tends to reduce power performance. The larger sample results in Table A.3 show that power performance increases dramatically with sample size in all scenarios with serial correlation.

#### 4.2. Decision tables

We supplement the above results with decision tables to aid in specification testing. Tables A.4 and A.5 display the results for our  $t$ -tests for sample sizes 100 and 500, respectively. The first panels in Tables A.4 and A.5 represent the percentages of times that we fail to reject either of the null hypotheses using the  $t$ -tests but ignoring the Wald test, choosing a fixed coefficient but spurious relationship. The top left corner shows the percentage of times we correctly identify this case. For both sample sizes, the procedure does so about 91% of the time. In general, when one uses two size- $\alpha$   $t$ -tests on a joint hypothesis, one can expect size between  $\alpha$  and  $2\alpha$ , meaning that the size distortion is up to  $2\alpha - \alpha = \alpha$ . In this case, the nominal size is 5% and the distortion is  $100\% - 91\% - 5\% = 4\% \leq 5\%$ . On the other hand, false positives happen up to 85% of the time in the case of very local alternatives with  $N=100$ , but no more than 6% of the time with  $N=500$ .

The second panel shows the percentage of times we reject  $H_0 : \theta = 1$  yet fail to reject  $H_0 : \sigma_\eta^2 = 0$ , indicating a fixed cointegrating vector. Elements of the leftmost column other than the top left corner show the percentage of times this case is correctly identified. As expected, local alternatives may make detection difficult for  $N=100$ , but not so for  $N=500$ . False positives are

no higher than 5%, the size of the respective tests, in the latter case. The third panel shows the opposite case, a spurious relationship with a time-varying parameter, with the percentage of correct decisions across the top row but excluding the top left corner. The results and intuition are symmetrically analogous to those of the previous case.

The final panel shows the probability of rejecting of both nulls, implying a time-varying cointegrating relationship. The correct decision is made in all elements outside of the top row and leftmost column. With a sample size of  $N=500$ , the correct decision is made 81.7–99.5% of the time. Interestingly, after some point the results deteriorate as both parameters move away from their respective nulls. We believe that this non-monotonicity is due to higher time variation in  $(\beta_t)$  being confused with nonstationarity in  $(w_t)$ , but we leave a formal investigation of the non-monotonicity for future research.

## 5. Empirical application

To illustrate the usefulness of our testing approach, we consider testing for a relationship between concentrations of well-mixed greenhouse gases (GHGs), of which  $\text{CO}_2$  (carbon dioxide) is the leading component, and global mean temperature anomalies (GMTs). There is a lack of consensus in the literature as to the statistical nature of the individual series: stochastically trending along the lines of Kaufmann et al. (2013) and similar papers vs. broken deterministically trending along the lines of Estrada et al. (2013) and similar papers. In spite of this disagreement, there seems to be a strong consensus that the relationship is linear in the sense that the two series share a common (stochastic or broken deterministic) trend. See also Leduc et al. (2016) for evidence of the linearity in this relationship based on output from highly nonlinear global climate models. If the trends are stochastic, then we expect these series to be cointegrated.

On the other hand, there are reasons why this relationship may have changed over time. Over a relatively short time-horizon, the period of roughly 1998–2013 is known as the “hiatus” in global warming, in the sense that the previously warming trend seemed to subside, while GHG concentrations continued to increase. A wide range of explanations for this phenomenon have been put forth, and the interested reader is referred to Schmidt et al. (2014), Pretis et al. (2015), Medhaug et al. (2017), and Miller and Nam (2018), for example.

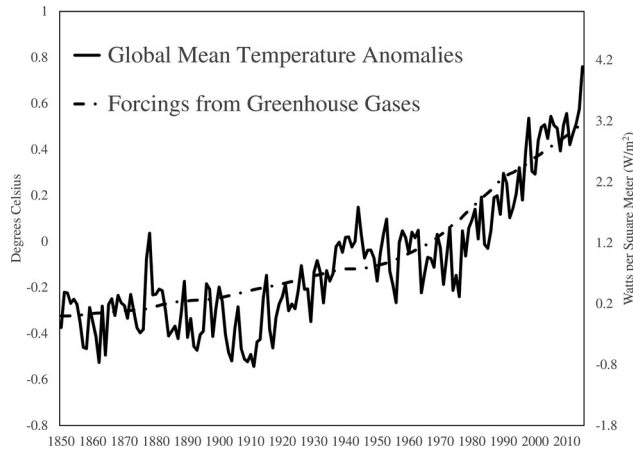
At a longer time horizon, a mid-20th Century change in sulfur dioxide emissions, which have a cooling rather than warming effect, may be responsible for introducing a second stochastic trend in the temperature series in the Northern Hemisphere (Chang et al., 2020). If so, the cointegrating relationship is no longer stable.

At an even longer paleoclimatic time scale, beyond the scope of our data, the range of  $\text{CO}_2$  concentrations over the 800,000 years preceding 1950 was 127 parts per million (ppm), while that of temperature was  $8^\circ\text{C}$  (Miller, 2019). Over the most recent 200 years, the ranges have been 120 ppm (about the same) and  $<1^\circ\text{C}$  (much less), which suggests the possibility of nonlinearity over this long time span.

To test the relationship over a time span long enough to encompass the (second) industrial revolution, we employ annual data spanning 1850–2015. Specifically, we use the HadCRUT4 global temperature anomalies measured relative to 1961–1990 (Morice et al., 2012)<sup>8</sup> and forcings data from Hansen et al. (2017).<sup>9</sup> We define GHGs as those authors do, adding  $\text{CO}_2$ , methane, nitrous oxide, ozone, and CFCs, the latter of which is their catchall for any GHGs not included in the first four groups. Forcings from tropospheric aerosols (anthropogenic), stratospheric

<sup>8</sup>Ensemble median of HadCRUT.4.5.0.0 (annual unsmoothed globally averaged) downloaded from [www.metoffice.gov.uk/hadobs](http://www.metoffice.gov.uk/hadobs) on July 18, 2017.

<sup>9</sup>Downloaded from [www.columbia.edu/~mhs119/Burden](http://www.columbia.edu/~mhs119/Burden) on May 15, 2017.



**Figure 2.** GMTs and GHGs. Global mean temperature anomalies in degrees Celsius ( $^{\circ}\text{C}$ ) and total forcings from greenhouse gases in Watts per square meter ( $\text{W}/\text{m}^2$ ).

**Table 2.** ML estimates and test statistics. Estimates and test statistics of the model in (1), (2), and (3) with GMT and GHG.

Hypothesis	Statistic	C.V.	$p$ -value
$\theta = 1$	-5.948	-2.726	<0.001
$\sigma_{\eta} = 0$	0.283	1.236	0.497
$\theta = 1, \sigma_{\eta} = 0$	3.211	2.124	0.034
Estimates	$\alpha$	$\rho$	$T$
	-0.379	-0.103	0.402
	$\mu$	$\sigma_{\eta}$	$\theta$
	0.170	0.000	0.586

aerosols (volcanic), land use albedo, and solar activity are not GHGs and therefore not included. Figure 2 shows the two series.

Letting  $y_t$  denote GMT and  $x_t$  denote GHG, we estimate the model in (1), (2), and (3) with  $p = 1$  and with  $k = 0$  chosen by the Schwarz–Bayesian information criterion up to a maximum lag length of 12. ML estimates and test statistics are given in Table 2. We estimate  $\theta = 0.586$  and  $\sigma_{\eta} = 0.00002$  with  $T < 1$ . Without running any test, these values strongly suggest a stable cointegrating relationship. Test results show  $t$ -statistics of  $t_{\theta} = -5.948$  and  $t_{\sigma_{\eta}^2} = 0.283$  against respective bootstrapped critical values of  $-2.726$  and  $1.236$ , using  $B = 999$  bootstrap replications. The respective  $p$ -values are  $< 0.001$  and  $0.497$ , providing solid statistical evidence in favor of this conclusion. Rejection of the joint null by a Wald test statistic of  $W_{\theta, \sigma_{\eta}} = 3.211$  against a bootstrapped critical value of  $2.124$  ( $p$ -value of  $0.034$ ) provides further support.

Although we know from the analysis of Chang et al. (2009) that Gaussian errors are not required for asymptotic inference in the presence of a stable cointegrating relationship, we run a Shapiro–Wilk test of the fitted residuals ( $\hat{\varepsilon}_t$ ) in order to more carefully justify the assumptions under which we reached the conclusion about a stable relationship. The Shapiro–Wilk test statistic has a maximum value of 1, which it attains at the null hypothesis that the distribution tested is Gaussian. We obtain a test statistic that exceeds 0.99. We do not know the exact distribution of the fitted residuals under the Gaussian null for precisely the same reasons that we introduce a bootstrap for our proposed tests. Nevertheless, a  $p$ -value of 0.62 for the Shapiro–Wilk test, calculated as if the fitted residuals were the series of interest rather than the underlying error, suggests that 0.99 is indeed close to 1 in a statistical sense. Hence, our statistical evidence cannot detect any departure from normality, which further supports our main result.

The stable cointegrating result is largely consistent with the literature that examines the relationship on this time scale. Both the mid-20th Century cooling and early 21st Century hiatus are

consistent with missing covariates – whether natural variability, natural forcings, or anthropogenic forcings – that cause short- or medium-run deviations from a stable relationship between CO<sub>2</sub> concentrations and temperatures.

These shorter-run deviations need not imply a departure from linearity in the long-run relationship between GHGs and GMT. So, if we were to include such covariates, with the possible exception of climate forcing from tropospheric aerosols, we would not expect any substantive change in the results. Or, put into a different perspective, just as short-run deviations from the law of one price or rational expectations in economics do not disprove those laws, short-run deviations in the present context do not disprove the relationship between these series implied by the Arrhenius law.

The data on tropospheric aerosols are more persistent than other covariates we might include, but the measurement is notoriously uncertain. If such forcings were I(1) and the series were needed to cointegrate GHGs and GMT, we would be estimating a spurious model without the series and  $t_\theta$  would detect the problem. The fact that  $t_\theta$  does not suggests that tropospheric aerosols are not part of the cointegrating relationship between GHGs and GMT: either forcings from tropospheric aerosols are not as persistent as these two series, or else the effect on GMT is not as long-lasting as that of GHGs.

## 6. Conclusion

We address a hole in the literature by developing a bootstrap-based test to distinguish time-varying cointegration from both fixed-coefficient cointegration and spurious regression using a straightforward time-varying state-space representation estimated using maximum likelihood applied to the Kalman filter. This is likely a familiar framework for econometric analysis to many readers, yet time-varying cointegration in this context has not been thoroughly investigated to the best of the authors' knowledge, and our testing procedure is novel.

Our test employs the insight that the intercept plus error term of a spurious relationship will contain a unit root. Although a spurious regression can be “solved” statistically by allowing a time-varying intercept, the relationship of interest is no less spurious in this case. We embed the error term into the state equation of the TVSSM and allow it to take on a unit root if the relationship is spurious. Our testing approach is based on whether or not it does so. This approach is slightly cleaner than but nearly equivalent to that of allowing the intercept to vary over time.

Although a bootstrapped  $t$ -test or Wald test in and of itself is nothing new, the degeneracy that occurs from the Kalman recursion makes our extension far from trivial, and our proposed bootstrap  $t$ -tests overcome both this degeneracy and the nonstationarity that undermines a key assumption of the bootstrap test of Stoffer and Wall (1991).

Moreover, as we show both numerically and analytically, this task is complicated by an inherent difficulty in distinguishing time-varying parameters and a spurious regression using the Kalman filter. In the case of the latter, the filter tends to squeeze the nonstationarity out of the error term, resulting in a (nonstationary) time-varying relationship when no relationship exists.

We apply our testing procedure to the long-run relationship between carbon dioxide concentrations and temperature anomalies, a relationship important to economists both from the point of view of its anthropogenic source and the potential economic damage from climate change – whether natural or man-made. We find this relationship to be stable – neither time-varying nor spurious – which largely agrees with the extant literature.



## Appendix A. Additional tables

**Table A.1.** Rejection frequencies with  $N=500$  and  $k=0$ .  $t$ -test of  $H_0 : \theta = 1$  (top panel)  $t$ -test of  $H_0 : \sigma_1 = 0$  (middle panel), and Wald test of  $H_0 : \sigma_1 = 0, \theta = 1$  (bottom panel). Nominal size is 5%.

$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.048	0.047	0.046	0.043	0.040	0.045	0.042	0.041	0.041	0.038	0.046
0.99	0.999	0.999	0.998	0.995	0.992	0.982	0.975	0.967	0.955	0.939	0.832
0.95	1.000	1.000	1.000	0.999	0.994	0.989	0.982	0.974	0.961	0.943	0.843
0.90	1.000	1.000	1.000	0.998	0.995	0.988	0.978	0.968	0.962	0.931	0.841
0.80	1.000	1.000	1.000	0.997	0.992	0.983	0.977	0.966	0.953	0.925	0.817
$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.042	0.936	0.995	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.99	0.042	0.936	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.95	0.043	0.933	0.993	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.90	0.046	0.929	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.80	0.039	0.906	0.993	0.998	1.000	0.999	1.000	1.000	1.000	1.000	1.000
$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.0491	0.7108	0.9746	0.9983	0.9981	0.9976	0.9989	0.9975	0.9978	0.9985	0.9947
0.99	0.0148	0.2008	0.8352	0.9721	0.991	0.9911	0.9885	0.9886	0.9909	0.9894	0.9885
0.95	0.324	0.4897	0.3812	0.8287	0.8874	0.8781	0.9154	0.9151	0.9291	0.9409	0.9596
0.90	0.6822	0.8766	0.8031	0.7712	0.8845	0.9215	0.9137	0.8936	0.8723	0.8996	0.929
0.80	0.9205	0.9827	0.9857	0.9805	0.971	0.9658	0.956	0.9545	0.9432	0.8997	0.8689

**Table A.2.** Rejection frequencies with  $N=100$  and  $k=0, 1, 2$ , chosen by BIC in testing.  $t$ -test of  $H_0 : \theta = 1$  (top panel)  $t$ -test of  $H_0 : \sigma_1 = 0$  (middle panel), and Wald test of  $H_0 : \sigma_1 = 0, \theta = 1$  (bottom panel). Nominal size is 5%.

	$lags = 0$						$lags = 1$					$lags = 2$				
$H_0 : \theta = 1$	$\theta_0$	0	0.1	0.2	0.3	0.4	$\sigma_{0,\eta_1}$					0	0.1	0.2	0.3	0.4
	1	0.053	0.053	0.057	0.054	0.047	0.045	0.049	0.062	0.052	0.054	0.047	0.048	0.046	0.045	0.046
	0.99	0.075	0.090	0.095	0.089	0.090	0.108	0.110	0.111	0.109	0.103	0.177	0.157	0.148	0.142	0.136
	0.95	0.221	0.225	0.209	0.188	0.161	0.528	0.501	0.451	0.370	0.291	0.792	0.750	0.716	0.600	0.472
	0.9	0.499	0.463	0.379	0.309	0.234	0.843	0.807	0.707	0.606	0.477	0.942	0.924	0.884	0.798	0.715
	0.8	0.998	0.953	0.853	0.761	0.591	0.989	0.960	0.863	0.744	0.674	0.993	0.983	0.935	0.854	0.756
	$lags = 0$						$lags = 1$					$lags = 2$				
$H_0 : \sigma_{0,\eta_1} = 0$	$\theta_0$	0	0.1	0.2	0.3	0.4	$\sigma_{0,\eta_1}$					0	0.1	0.2	0.3	0.4
	1	0.050	0.488	0.822	0.928	0.971	0.048	0.519	0.826	0.930	0.961	0.045	0.666	0.880	0.950	0.971
	0.99	0.055	0.471	0.806	0.911	0.967	0.042	0.531	0.821	0.918	0.960	0.048	0.673	0.884	0.947	0.975
	0.95	0.052	0.454	0.779	0.907	0.965	0.047	0.509	0.825	0.920	0.958	0.042	0.649	0.885	0.947	0.971
	0.9	0.048	0.428	0.762	0.904	0.962	0.048	0.515	0.800	0.906	0.951	0.048	0.649	0.867	0.936	0.970
	0.8	0.051	0.420	0.760	0.900	0.956	0.048	0.489	0.778	0.901	0.948	0.049	0.623	0.848	0.926	0.972
	$lags = 0$						$lags = 1$					$lags = 2$				
$H_0 : \sigma_{0,\eta_1} = 0, \theta = 1$	$\theta_0$	0	0.1	0.2	0.3	0.4	$\sigma_{0,\eta_1}$					0	0.1	0.2	0.3	0.4
	1	0.049	0.240	0.537	0.682	0.733	0.052	0.358	0.745	0.881	0.921	0.048	0.578	0.928	0.974	0.983
	0.99	0.043	0.146	0.417	0.575	0.662	0.024	0.174	0.568	0.801	0.869	0.037	0.323	0.824	0.934	0.966
	0.95	0.062	0.130	0.214	0.347	0.451	0.035	0.041	0.186	0.449	0.659	0.013	0.028	0.372	0.750	0.853
	0.9	0.150	0.201	0.200	0.253	0.306	0.114	0.118	0.093	0.262	0.432	0.029	0.027	0.155	0.476	0.727
	0.8	0.321	0.401	0.338	0.326	0.326	0.365	0.388	0.249	0.187	0.269	0.288	0.275	0.148	0.275	0.439



**Table A.3.** Rejection frequencies with  $N=500$  and  $k=0, 1, 2$ , chosen by BIC in testing.  $t$ -test of  $H_0 : \theta = 1$  (top panel)  $t$ -test of  $H_0 : \sigma_1 = 0$  (middle panel), and Wald test of  $H_0 : \sigma_1 = 0, \theta = 1$  (bottom panel). Nominal size is 5%.

		$lags = 0$					$lags = 1$					$lags = 2$				
		$\sigma_0, \eta_1$														
$H_0 : \theta = 1$	$\theta_0$	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4
	1	0.052	0.063	0.070	0.073	0.075	0.049	0.048	0.050	0.055	0.051	0.051	0.037	0.036	0.036	0.035
	0.99	0.293	0.341	0.351	0.331	0.302	0.693	0.680	0.610	0.606	0.561	0.985	0.966	0.934	0.884	0.859
	0.95	0.972	0.961	0.899	0.818	0.743	0.999	0.994	0.965	0.918	0.860	1.000	0.999	0.998	0.973	0.946
	0.9	1.000	0.983	0.927	0.873	0.847	1.000	0.995	0.971	0.928	0.881	1.000	1.000	0.990	0.965	0.941
	0.8	1.000	0.988	0.948	0.916	0.906	1.000	0.997	0.969	0.931	0.900	1.000	0.999	0.987	0.965	0.953
		$lags = 0$					$lags = 1$					$lags = 2$				
		$\sigma_0, \eta_1$														
$H_0 : \sigma_0, \eta_1 = 0$	$\theta_0$	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4
	1	0.048	0.989	1.000	1.000	1.000	0.046	0.993	1.000	1.000	1.000	0.063	0.997	1.000	1.000	1.000
	0.99	0.048	0.991	1.000	1.000	1.000	0.049	0.991	1.000	1.000	1.000	0.062	0.996	1.000	1.000	1.000
	0.95	0.047	0.988	1.000	1.000	1.000	0.050	0.990	0.999	1.000	1.000	0.067	0.995	1.000	1.000	1.000
	0.9	0.054	0.985	1.000	1.000	1.000	0.045	0.988	0.999	1.000	1.000	0.065	0.995	0.999	1.000	1.000
	0.8	0.055	0.981	1.000	1.000	1.000	0.053	0.987	0.999	1.000	1.000	0.065	0.992	1.000	1.000	1.000
		$lags = 0$					$lags = 1$					$lags = 2$				
		$\sigma_0, \eta_1$														
$H_0 : \sigma_0, \eta_1 = 0, \theta = 1$	$\theta_0$	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4	0	0.1	0.2	0.3	0.4
	1	0.051	0.935	0.988	0.990	0.986	0.059	0.998	1.000	0.999	0.999	0.068	1.000	1.000	1.000	1.000
	0.99	0.014	0.753	0.957	0.967	0.964	0.002	0.922	0.995	0.994	0.994	0.000	0.963	0.997	0.999	0.998
	0.95	0.182	0.457	0.797	0.842	0.866	0.140	0.296	0.893	0.927	0.940	0.020	0.124	0.968	0.978	0.972
	0.9	0.486	0.814	0.856	0.858	0.839	0.465	0.730	0.878	0.888	0.915	0.215	0.407	0.919	0.955	0.958
	0.8	0.795	0.979	0.986	0.957	0.928	0.792	0.968	0.962	0.971	0.939	0.697	0.945	0.966	0.975	0.961

**Table A.4.** Decision tables with  $N=100$  and  $k=0$ : Percentage of times reaching a decision. Do not reject either  $\theta=1$  or  $\sigma_1^2=0$  (fixed coefficient spurious, top panel); reject  $\theta=1$  but do not reject  $\sigma_1^2=0$  (fixed coefficient cointegrating, second panel), do not reject  $\theta=1$  but reject  $\sigma_1^2=0$  (time-varying spurious, third panel) and reject both  $\theta=1$  and  $\sigma_1^2=0$  (time-varying cointegrating, bottom panel).

		$\sigma_0, \eta_1$									
$\theta_0$	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.909	0.758	0.485	0.301	0.197	0.137	0.076	0.066	0.041	0.028	0.005
0.99	0.848	0.711	0.451	0.263	0.188	0.124	0.085	0.063	0.041	0.022	0.004
0.95	0.451	0.396	0.255	0.167	0.114	0.084	0.056	0.041	0.034	0.020	0.005
0.90	0.168	0.149	0.115	0.077	0.053	0.045	0.030	0.022	0.018	0.015	0.004
0.80	0.018	0.019	0.017	0.016	0.013	0.012	0.012	0.010	0.008	0.009	0.005
		$\sigma_0, \eta_1$									
$\theta_0$	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.045	0.033	0.025	0.014	0.008	0.007	0.004	0.003	0.002	0.001	0.000
0.99	0.105	0.081	0.061	0.027	0.026	0.014	0.008	0.005	0.004	0.003	0.000
0.95	0.496	0.405	0.257	0.159	0.102	0.070	0.045	0.027	0.022	0.011	0.001
0.90	0.786	0.657	0.427	0.281	0.164	0.121	0.075	0.053	0.037	0.021	0.002
0.8	0.932	0.806	0.543	0.366	0.241	0.163	0.115	0.079	0.056	0.026	0.002
		$\sigma_0, \eta_1$									
$\theta_0$	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.045	0.198	0.461	0.650	0.751	0.810	0.870	0.884	0.907	0.918	0.948
0.99	0.044	0.185	0.425	0.632	0.696	0.764	0.826	0.841	0.872	0.893	0.931
0.95	0.030	0.103	0.254	0.361	0.464	0.500	0.565	0.614	0.650	0.720	0.883
0.90	0.012	0.044	0.122	0.181	0.253	0.309	0.381	0.425	0.465	0.603	0.806
0.80	0.003	0.011	0.042	0.091	0.142	0.196	0.245	0.307	0.354	0.490	0.791
		$\sigma_0, \eta_1$									
$\theta_0$	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.002	0.011	0.029	0.035	0.044	0.046	0.050	0.047	0.051	0.053	0.047
0.99	0.004	0.023	0.063	0.078	0.091	0.097	0.082	0.091	0.083	0.082	0.065
0.95	0.023	0.095	0.234	0.313	0.321	0.346	0.334	0.318	0.294	0.249	0.112
0.90	0.035	0.151	0.335	0.461	0.530	0.525	0.514	0.501	0.480	0.362	0.188
0.80	0.046	0.164	0.398	0.528	0.605	0.629	0.628	0.605	0.582	0.475	0.202

**Table A.5.** Decision tables with  $N=500$  and  $k=0$ : Percentage of times reaching a decision. Do not reject either  $\theta=1$  or  $\sigma_1^2=0$  (fixed coefficient spurious, top panel); reject  $\theta=1$  but do not reject  $\sigma_1^2=0$  (fixed coefficient cointegrating, second panel), do not reject  $\theta=1$  but reject  $\sigma_1^2=0$  (time-varying spurious, third panel) and reject both  $\theta=1$  and  $\sigma_1^2=0$  (time-varying cointegrating, bottom panel).

$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.912	0.062	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.99	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.95	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.046	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.99	0.957	0.064	0.004	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.95	0.957	0.067	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.90	0.954	0.071	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80	0.961	0.094	0.007	0.002	0.000	0.001	0.000	0.000	0.000	0.000	0.000
$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.041	0.892	0.948	0.956	0.960	0.955	0.958	0.959	0.959	0.962	0.954
0.99	0.000	0.001	0.002	0.005	0.008	0.018	0.025	0.033	0.045	0.061	0.168
0.95	0.000	0.000	0.000	0.001	0.006	0.011	0.018	0.026	0.039	0.058	0.157
0.90	0.000	0.000	0.000	0.002	0.005	0.013	0.022	0.032	0.038	0.069	0.159
0.80	0.000	0.000	0.000	0.003	0.008	0.017	0.023	0.034	0.047	0.075	0.183
$\theta_0$	$\sigma_{0,\eta_1}$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	1.00
1	0.002	0.044	0.046	0.043	0.040	0.045	0.042	0.041	0.041	0.038	0.046
0.99	0.042	0.935	0.994	0.995	0.991	0.982	0.975	0.967	0.955	0.939	0.832
0.95	0.043	0.933	0.993	0.998	0.994	0.989	0.982	0.974	0.961	0.943	0.843
0.90	0.046	0.929	0.994	0.997	0.995	0.987	0.978	0.968	0.962	0.931	0.841
0.80	0.039	0.906	0.992	0.995	0.991	0.983	0.976	0.966	0.953	0.925	0.817

## Appendix B. State-space model and the Kalman filter

We outline the Kalman filtering procedure for our state space model, both to explain the procedure itself and to utilize it for further analysis below. For simplicity, assume that  $p=1$  and  $k=0$ . The state equation may be rewritten as

$$\begin{aligned}\xi_t &= \sum_{i=0}^{t-1} F^i A + \sum_{i=0}^{t-1} F^i E_{t-i} + F^t \xi_0 \\ \begin{bmatrix} \beta_t \\ w_t \end{bmatrix} &= \begin{bmatrix} \mu \sum_{i=0}^{t-1} T^i \\ 0 \end{bmatrix} + \begin{bmatrix} \sum_{i=0}^{t-1} T^i \eta_{t-i} \\ \sum_{i=0}^{t-1} \theta^i \varepsilon_{t-i} \end{bmatrix} + \begin{bmatrix} T^t \beta_0 \\ \theta^t w_0 \end{bmatrix}\end{aligned}$$

We define the filtration  $(\mathcal{F}_t)$  such that  $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \dots, x_{t+1}, x_t, x_{t-1}, \dots)$ . Specifically,  $x_t$  is  $\mathcal{F}_{t-1}$ -measurable, so that  $E[x_t | \mathcal{F}_s] = x_t$  for any  $s \geq t-1$ . As is standard in the Kalman filtering literature, we use the subscript  $t|s$  to denote the conditional expectation or conditional variance. Specifically,  $\xi_{t|s}$  and  $\Omega_{t|s}$  are the conditional expectation and variance of  $\xi_t$  given information at time  $s$ , while  $y_{t|s}$  and  $\sigma_{t|s}$  are the conditional expectation and variance of  $y_t$  given information at time  $s$ .

The prediction equations of the Kalman filter are as follows:

- $\xi_{t|t-1} = A + F\xi_{t-1|t-1}$
- $y_{t|t-1} = \alpha + z_t' \xi_{t|t-1}$
- $\Omega_{t|t-1} = F\Omega_{t-1|t-1}F' + V$ , where

$$V = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

- $\sigma_{t|t-1} = z_t' \Omega_{t|t-1} z_t = \omega_{t|t-1}^{11} x_t^2 + 2\omega_{t|t-1}^{21} x_t + \omega_{t|t-1}^{22}$ , where

$$\Omega_{t|t-1} = \begin{bmatrix} \omega_{t|t-1}^{11} & \omega_{t|t-1}^{21} \\ \omega_{t|t-1}^{21} & \omega_{t|t-1}^{22} \end{bmatrix}$$

The updating equations of Kalman filter are as follows:

- $\xi_{t|t} = \xi_{t|t-1} + \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} (y_t - y_{t|t-1})$
- $\Omega_{t|t} = \Omega_{t|t-1} - \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t' \Omega_{t|t-1}$

The latter equation may be simplified considerably. By straightforward matrix algebra,

$$\begin{aligned} \Omega_{t|t} &= \Omega_{t|t-1} \left( I - \sigma_{t|t-1}^{-1} \begin{bmatrix} x_t^2 & x_t \\ x_t & 1 \end{bmatrix} \begin{bmatrix} \omega_{t|t-1}^{11} & \omega_{t|t-1}^{21} \\ \omega_{t|t-1}^{21} & \omega_{t|t-1}^{22} \end{bmatrix} \right) \\ &= \Omega_{t|t-1} \left( I - \sigma_{t|t-1}^{-1} \begin{bmatrix} \omega_{t|t-1}^{11} x_t^2 + \omega_{t|t-1}^{21} x_t & \omega_{t|t-1}^{21} x_t^2 + \omega_{t|t-1}^{22} x_t \\ \omega_{t|t-1}^{11} x_t + \omega_{t|t-1}^{21} & \omega_{t|t-1}^{21} x_t + \omega_{t|t-1}^{22} \end{bmatrix} \right) \\ &= \sigma_{t|t-1}^{-1} \Omega_{t|t-1} \begin{bmatrix} (\omega_{t|t-1}^{21} x_t + \omega_{t|t-1}^{22}) & -(\omega_{t|t-1}^{21} x_t^2 + \omega_{t|t-1}^{22} x_t) \\ -(\omega_{t|t-1}^{11} x_t + \omega_{t|t-1}^{21}) & (\omega_{t|t-1}^{11} x_t^2 + \omega_{t|t-1}^{21} x_t) \end{bmatrix} \end{aligned}$$

Multiplying through by  $\Omega_{t|t-1}$  simplifies elegantly to

$$\Omega_{t|t} = \sigma_{t|t-1}^{-1} \det(\Omega_{t|t-1}) \begin{bmatrix} 1 & -x_t \\ -x_t & x_t^2 \end{bmatrix}$$

after some algebra.

### B.1 Kalman dynamics of $\Omega_{t+1|t}$ , $\xi_{t+1|t}$ and $\xi_{t+1} - \xi_{t+1|t}$

Now, combining the prediction and updating equations for the conditional variance yields

$$\Omega_{t+1|t} = (z_t' \Omega_{t|t-1} z_t)^{-1} \det(\Omega_{t|t-1}) \begin{bmatrix} T^2 & -T\theta x_t \\ -T\theta x_t & \theta^2 x_t^2 \end{bmatrix} + V \quad (\text{B.1})$$

so that the evolution of  $\Omega_{t|t-1}$  is described by a complicated difference equation. This equation is a function of the data  $x_t$  in such a way that a steady state cannot be guaranteed, in contrast to the difference equation of Chang et al. (2009) for the case in which the coefficients cannot vary over time. Doing so for the conditional mean yields

$$\xi_{t+1|t} = A + F(\xi_{t|t-1} + \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} (y_t - \alpha - z_t' \xi_{t|t-1})) \quad (\text{B.2})$$

similarly. Note that Eqs. (B1) and (B2) may be used in place of the prediction and updating equations in implementing the Kalman filter.

Subtracting (B.2) from the state equation of the TVSSM yields

$$\begin{aligned} \xi_{t+1} - \xi_{t+1|t} &= F\xi_t + E_{t+1} - F(\xi_{t|t-1} + \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t' (\xi_t - \xi_{t|t-1})) \\ &= \Xi_{t|t-1} (\xi_t - \xi_{t|t-1}) + E_{t+1} \end{aligned} \quad (\text{B.3})$$

where  $\Xi_{t|t-1} = F(I - \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t')$ . Eq. (B3) reveals that the evolution of  $\xi_{t+1} - \xi_{t+1|t}$ , the difference between the state and its conditional mean, follows a vector autoregression of order 1 with a time-varying coefficient matrix  $F(I - \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t')$ . Write

$$(I - \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t') = \Omega_{t|t-1}^{1/2} (I - \Omega_{t|t-1}^{1/2'} z_t' \Omega_{t|t-1} z_t)^{-1} z_t' \Omega_{t|t-1}^{1/2} \Omega_{t|t-1}^{-1/2}$$

to see that this matrix is idempotent of rank 1, so that it has eigenvalues given by 0 and 1. After some algebra, we obtain

$$\begin{aligned} \Xi_{t|t-1} &= F(I - \Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1} z_t') = \begin{bmatrix} T(1 - \gamma_t) & -\frac{T}{x_t} \gamma_t \\ -\theta x_t(1 - \gamma_t) & \theta \gamma_t \end{bmatrix} \text{ where} \\ \gamma_t &= (z_t' \Omega_{t|t-1} z_t)^{-1} (\omega_{t|t-1}^{11} x_t + \omega_{t|t-1}^{21}) x_t = K_{1,t|t-1} x_t \end{aligned}$$

with  $K_{1,t|t-1}$  being the first element of the  $2 \times 1$  vector of the Kalman gain  $\Omega_{t|t-1} z_t \sigma_{t|t-1}^{-1}$ .

Supposing that  $T\theta > 0$  and  $\omega_{t|t-1}^{21} x_t > 0$ , so that  $0 < \gamma_t < 1$ . The eigenvalues of  $\Xi_{t|t-1}$  are 0 and  $\theta \gamma_t + T(1 - \gamma_t)$ . Supposing also that  $T < 1$  or  $\theta < 1$ , the eigenvalues of  $\Xi_{t|t-1}$  are strictly less than 1. Because we are testing  $\theta = 1$ , we restrict  $0 \leq T < 1$  to ensure that they are. The unit bound on the eigenvalues means that the difference between the state and its conditional mean is not divergent, or that the conditional mean calculated by the filter is a reasonable approximation to the unobserved state. This is also the case in the simpler model considered by

Chang et al. (2009), in which  $\theta=0$ ,  $T=1$ , and  $\beta_t = \beta$ .

## B.2 Effect of a spurious regression on $\beta_{t+1|t}$ : $T \simeq 1$

Suppose that  $\theta=0$  is fixed in estimation, but that there is a unit root in  $(w_t)$ , so that  $\theta=1$  in the data generating process. Suppose further that  $T=0$  in the data generating process, but that it is estimated. The regression is spurious. Similarly to Eq. (B3)

$$\begin{aligned}\xi_{t+1} - \xi_{t+1|t} &= \left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} T & 0 \\ 0 & 0 \end{bmatrix} \right) \xi_t + \Xi_{t|t-1}(\xi_t - \xi_{t|t-1}) + E_{t+1} \\ &= (-T\beta_t, w_t)' + \Xi_{t|t-1}(\xi_t - \xi_{t|t-1}) + E_{t+1}\end{aligned}$$

where  $\Xi_{t|t-1}$  is defined as above but with  $\theta=0$ . The last two terms may be written as

$$\begin{bmatrix} T(1 - \gamma_t)(\beta_t - \beta_{t|t-1}) - TK_{1,t|t-1}(w_t - w_{t|t-1}) + \eta_{t+1} \\ \varepsilon_{t+1} \end{bmatrix}$$

when  $\theta=0$ . The second element of the vector  $(\xi_{t+1} - \xi_{t+1|t})$  is

$$(w_{t+1} - w_{t+1|t}) = w_t + \varepsilon_{t+1} \text{ or } w_{t+1|t} = \Delta w_{t+1} - \varepsilon_{t+1},$$

which means we incorrectly estimate  $w_t$  to be  $I(0)$ , so that the difference between  $w_t$  and its estimate is  $I(1)$ . The first element of the vector  $(\xi_{t+1} - \xi_{t+1|t})$  may be rewritten as

$$(\beta_{t+1} - \beta_{t+1|t}) - T(1 - \gamma_t)(\beta_t - \beta_{t|t-1}) = -T\beta_t - TK_{1,t|t-1}(w_t + \varepsilon_t) + \eta_{t+1}$$

Suppose for the moment that  $\gamma_t$  and  $x_t$  are approximately constant so that  $K_{1,t|t-1}$  is also. In that case, the right-hand side is  $I(1)$ , so that the left-hand side must also be  $I(1)$ .  $\beta_t$  is  $I(0)$ , so the only remaining series on the left-hand side  $-\beta_{t|t-1}$ , the series that estimates  $\beta_t$  - must be  $I(1)$ , pushing the estimate of the autoregressive parameter  $T$  toward unity.

B.3 On the complication of asymptotic analysis

Let  $\varphi = (T, \theta)'$  and define the conditional log-likelihood concentrated on this subvector of parameters to be

$$\ell_{t|t-1}(\varphi) = -\frac{1}{2} \ln \sigma_{t|t-1}(\varphi) - \frac{1}{2\sigma_{t|t-1}(\varphi)} (y_t - y_{t|t-1}(\varphi))^2$$

up to an irrelevant constant. The score may be written as

$$\begin{aligned}\frac{\partial \ell_{t|t-1}(\varphi)}{\partial \varphi} &= \left[ \frac{y_t - y_{t|t-1}}{\sigma_{t|t-1}} \right] \frac{\partial y_{t|t-1}}{\partial \varphi} + \frac{1}{2} \left[ \left( \frac{y_t - y_{t|t-1}}{\sigma_{t|t-1}} \right)^2 - \frac{1}{\sigma_{t|t-1}} \right] \frac{\partial \sigma_{t|t-1}}{\partial \varphi} \\ &= \frac{\partial z_t' \xi_{t|t-1}}{\partial \varphi} \frac{z_t'}{z_t' \Omega_{t|t-1} z_t} [\xi_t - \xi_{t|t-1}] \\ &\quad + \frac{1}{2} \frac{\partial z_t' \Omega_{t|t-1} z_t}{\partial \varphi} \left[ \frac{z_t' ((\xi_t - \xi_{t|t-1})(\xi_t - \xi_{t|t-1})' - \Omega_{t|t-1}) z_t}{(z_t' \Omega_{t|t-1} z_t)^2} \right]\end{aligned}\tag{B.4}$$

where the argument  $\varphi$  of  $\xi_{t|t-1}$  and  $\Omega_{t|t-1}$  is suppressed. Maximum likelihood estimation also requires the Hessian, but we do not need it to make our point.

Suppose that  $z_t = (1, x_t)'$  for all  $t$ , so that there is no variation in  $z_t$ , reducing to the case of a measurement equation that is linear in the state similar to but not exactly that considered by Chang et al. (2009). In this case,  $\Omega_{t|t-1}$  could reach a steady state value  $\Omega$ , so that the score becomes

$$\frac{\partial z_t' \xi_{t|t-1}}{\partial \varphi} \frac{z_t'}{z_t' \Omega z_t} [\xi_t - \xi_{t|t-1}] + \frac{1}{2} \frac{\partial z_t' \Omega z_t}{\partial \varphi} \left[ \frac{z_t' ((\xi_t - \xi_{t|t-1})(\xi_t - \xi_{t|t-1})' - \Omega) z_t}{(z_t' \Omega z_t)^2} \right]$$

The non-zero eigenvalue  $\theta\gamma_t + T(1 - \gamma_t)$  of  $\Xi_{t|t-1}$  would be time-invariant and inside the unit circle, so that  $(\xi_t - \xi_{t|t-1})$  would have a stationary – albeit rank degenerate – representation with conditional variance given by  $\Omega$ . As a result, both terms of the score vector would lend themselves to asymptotic analysis similar to that of Chang et al. (2009) with  $\sqrt{N}$  rates from the stationarity of the square-bracketed terms, very loosely speaking. Naturally, the Hessian would also need to be considered to complete the analysis.

Because  $(x_t)$  is a stochastic series, the variances  $\Omega_{t|t-1}$  and  $\sigma_{t|t-1}$  cannot settle into a steady state, so the square-bracketed terms of (B.4) are not stationary. As long as  $|\gamma_t| < 1$  as discussed above, we can expect the difference equation in (B.3) to have a solution, but that solution will be a moving average with time-varying coefficients that are (bounded) functions of unit root processes.

There are two avenues for considering asymptotic analysis of objects like these: indirect modeling using (possibly) near-epoch dependence along the lines of Davidson (1994) and De Jong (1997) or direct modeling of nonlinear transformations of unit root processes along the lines of Park and Phillips (1999, 2001), Chang et al. (2001), De Jong (2004), and De Jong and Wang (2005). Both asymptotic approaches are substantially complicated by the high degree of nonlinearity in the parameters and by infinite order of the moving average in the inverse of  $(\xi_t - \xi_{t|t-1})$ . Because  $|\gamma_t| < 1$  but a full asymptotic analysis may be intractable or very difficult, the bootstrap provides a practically useful solution.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Bierens, H. J., Martins, L. F. (2010). Time-varying cointegration. *Econometric Theory* 26(5):1453–1490. doi:10.1017/S0266466609990648
- Canarellsal, G., Pollard, S. K., Lai, K. S. (1990). Cointegration between exchange rates and relative prices: another view. *European Economic Review* 34(7):1303–1322. doi:10.1016/0014-2921(90)90002-G
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., Park, S. (2020). Evaluating trends in time series of distributions: a spatial fingerprint of human effects on climate. *Journal of Econometrics* 214(1):274–294. doi:10.1016/j.jeconom.2019.05.014
- Chang, Y., Miller, J. I., Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications. *Journal of Econometrics* 150(2):231–247. doi:10.1016/j.jeconom.2008.12.007
- Chang, Y., Park, J. Y. (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* 24(4):379–400. doi:10.1111/1467-9892.00312
- Chang, Y., Park, J. Y., Phillips, P. C. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal* 4(1):1–36. doi:10.1111/1368-423X.00054
- Cooley, T. F., Prescott, E. C. (1976). Estimation in the presence of stochastic parameter variation. *Econometrica: Journal of the Econometric Society* 44(1):167–184. doi:10.2307/1911389
- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford: OUP.
- Davidson, R., MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis* 51(7):3259–3281. doi:10.1016/j.csda.2006.04.001
- De Jong, R., Wang, C.-H. (2005). Further results on the asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 21(02):413–430. doi:10.1017/S026646660505022X
- De Jong, R. M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13(3):353–367. doi:10.1017/S0266466600005843
- De Jong, R. M. (2004). Addendum to “asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 20(03):627–635. doi:10.1017/S0266466604203085
- Engle, R. F., Watson, M. W. (1987). The kalman filter model: Applications to forecasting and rational expectations. In: *Advances in Econometrics: Firth World Congress, I*. Cambridge: Cambridge University Press.
- Estrada, F., Perron, P., Martínez-López, B. (2013). Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience* 6(12):1050–1055. doi:10.1038/ngeo1999
- Evans, M. (1991). Discovering the link between inflation rates and inflation uncertainty. *Journal of Money, Credit and Banking* 23(2):169–184. doi:10.2307/1992775
- Hamilton, J. D. (1994). *Time Series Econometrics*. Princeton: Princeton U. Press.
- Hansen, B. E. (1992). Heteroskedastic cointegration. *Journal of Econometrics* 54(1-3):139–158. doi:10.1016/0304-4076(92)90103-X
- Hansen, H., Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR-models. *The Econometrics Journal* 2(2):306–333. doi:10.1111/1368-423X.00035
- Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K., Beerling, D. J., Cao, J., Marcott, S., Masson-Delmotte, V., Prather, M. J., Rohling, E. J., Shakun, J., Smith, P., Laci, A., Russell, G., Ruedy, R. (2017). Young people’s burden: requirement of negative CO<sub>2</sub> emissions. *Earth System Dynamics* 8(3):577–616. doi:10.5194/esd-8-577-2017
- Harris, D., McCabe, B., Leybourne, S. (2002). Stochastic cointegration: estimation and inference. *Journal of Econometrics* 111(2):363–384. doi:10.1016/S0304-4076(02)00111-2
- Honohan, P. (1993). Cointegration and time-varying parameters: a comment. *European Economic Review* 37(6):1279–1281. doi:10.1016/0014-2921(93)90136-X
- Kaufmann, R. K., Kauppi, H., Mann, M. L., Stock, J. H. (2013). Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Climatic Change* 118(3-4):729–743. doi:10.1007/s10584-012-0683-2

- Kim, C.-J. (1993). Unobserved-component time series models with markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business & Economic Statistics* 11(3):341–349.
- Kim, C.-J. (2006). Time-varying parameter models with endogenous regressors. *Economics Letters* 91(1):21–26. doi:[10.1016/j.econlet.2005.10.007](https://doi.org/10.1016/j.econlet.2005.10.007)
- Leduc, M., Matthews, H. D., de Elía, R. (2016). Regional estimates of the transient climate response to cumulative CO<sub>2</sub> emissions. *Nature Climate Change* 6(5):474–478. doi:[10.1038/nclimate2913](https://doi.org/10.1038/nclimate2913)
- Ljung, L., Caines, P. E. (1980). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics* 3(1-4):29–46.
- McCabe, B., Leybourne, S., Harris, D. (2006). A residual-based test for stochastic cointegration. *Econometric Theory* 22(03):429–456. doi:[10.1017/S026646660606021X](https://doi.org/10.1017/S026646660606021X)
- Medhaug, I., Stolpe, M. B., Fischer, E. M., Knutti, R. (2017). Reconciling controversies about the ‘global warming hiatus’. *Nature* 545(7652):41–47. doi:[10.1038/nature22315](https://doi.org/10.1038/nature22315)
- Miller, J. I. (2019). Testing cointegrating relationships using irregular and non-contemporaneous series with an application to paleoclimate data. *Journal of Time Series Analysis* 40(6):936–950. doi:[10.1111/jtsa.12469](https://doi.org/10.1111/jtsa.12469)
- Miller, J. I., Nam, K. (2018). Dating hiatuses: a statistical model of the recent slowdown in global warming – and the next one. *Earth System Dynamics Discussion* 11(4):1123–1132. doi:[10.5194/esd-2018-81](https://doi.org/10.5194/esd-2018-81)
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres* 117(D8):n/a–n/a. doi:[10.1029/2011JD017187](https://doi.org/10.1029/2011JD017187)
- Park, J. Y., Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory* 15(5):664–703. doi:[10.1017/S0266466699155026](https://doi.org/10.1017/S0266466699155026)
- Park, J. Y., Phillips, P. C. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15(03):269–298. doi:[10.1017/S0266466699153015](https://doi.org/10.1017/S0266466699153015)
- Park, J. Y., Phillips, P. C. (2001). Nonlinear regressions with integrated time series. *Econometrica* 69(1):117–161. doi:[10.1111/1468-0262.00180](https://doi.org/10.1111/1468-0262.00180)
- Pretis, F., Mann, M. L., Kaufmann, R. K. (2015). Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change* 131(4):705–718. doi:[10.1007/s10584-015-1391-5](https://doi.org/10.1007/s10584-015-1391-5)
- Schmidt, G. A., Shindell, D. T., Tsigaridis, K. (2014). Reconciling warming trends. *Nature Geoscience* 7(3):158–160. doi:[10.1038/ngeo2105](https://doi.org/10.1038/ngeo2105)
- Seo, B. (1998). Tests for structural change in cointegrated systems. *Econometric Theory* 14(2):222–259. doi:[10.1017/S0266466698142044](https://doi.org/10.1017/S0266466698142044)
- Stoffer, D. S., Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the American Statistical Association* 86(416):1024–1033. doi:[10.1080/01621459.1991.10475148](https://doi.org/10.1080/01621459.1991.10475148)
- Trokić, M. (2019). Wavelet energy ratio unit root tests. *Econometric Reviews* 38(1):69–94. doi:[10.1080/07474938.2016.1222232](https://doi.org/10.1080/07474938.2016.1222232)
- Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics* 152(2):81–92. doi:[10.1016/j.jeconom.2009.01.008](https://doi.org/10.1016/j.jeconom.2009.01.008)