



DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY



Center for Research in Econometric Analysis of Time Series

Global Hemispheric Temperatures and Co-Shifting: A Vector Shifting-Mean Autoregressive Analysis

Matthew T. Holt and Timo Teräsvirta

CREATES Research Paper 2017-5

Global Hemispheric Temperatures and Co-Shifting: A Vector Shifting-Mean Autoregressive Analysis*

Timo Teräsvirta[†]

CREATES, *Department of Economics and Business Economics*
Aarhus University
C.A.S.E., *Humboldt-Universität zu Berlin*

Matthew T. Holt[‡]

Department of Economics, Finance & Legal Studies
University of Alabama

January 24, 2017

Abstract

This paper examines local changes in annual temperature data for the northern and southern hemispheres (1850–2014) by using a multivariate generalisation of the shifting-mean autoregressive model of González and Teräsvirta (2008). Univariate models are first fitted to each series by using the QuickShift methodology. Full information maximum likelihood estimates of a bivariate system of temperature equations are then obtained and asymptotic properties of the corresponding estimators considered. The system is then used to perform formal tests of co-movements, called co-shifting, in the series. The results show evidence of co-shifting in the two series. Forecasting this pair of series is considered as well.

Keywords: Co-breaking; Hemispheric temperatures; Vector nonlinear model; Testing linearity; Structural change

JEL Classification Codes: C22; C32; C52; C53; Q54

***Acknowledgment:** This research has been supported by a travel grant from the College of Commerce and Business Administration, University of Alabama and by the Danish National Research Foundation Grant No. DNRF 78. Part of the work for this paper was carried out while the first author was visiting CREATES, Aarhus University, and the work continued during the second author's visit to the Department of Economics, Finance, and Legal Studies, University of Alabama. Kind hospitality at both institutions is gratefully acknowledged. Material from this paper has been presented at the Conference on 'Innovations in Time Series Econometrics' in honour of Helmut Lütkepohl's scientific contributions, Berlin, September 2016, the conference 'Econometric Models of Climate Change', Aarhus University, October 2016, and in seminars at Aarhus University, Mississippi State University, Brigham Young University, and the University of Wisconsin, Madison. We thank participants at these occasions, David Hendry in particular, for their comments. Errors and shortcomings in this work remain our own responsibility.

[†]CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Telephone: +45 8716 5563. E-mail: tterasvirta@econ.au.dk

[‡]Professor and Dwight Harrigan Endowed Faculty Fellow in Natural Resources Economics, Department of Economics, Finance & Legal Studies, University of Alabama, Box 870224, 248 Alston Hall, Tuscaloosa, AL 35487-0224, USA. Telephone: 205-348-8980. Fax: 205-348-0590. E-mail: mtholt@cba.ua.edu.

1 Introduction

In recent years there has been considerable debate in the climate change literature, interestingly enough, not about whether global warming can be detected in available time series data but rather the proper way to characterize this phenomenon in the modelling process. The essence of the debate is this: do global and hemispheric temperature data follow a unit root (difference stationary) process wherein shocks to the world's climate can be expected to have a permanent effect, perhaps with a shared stochastic trend (cointegration) between northern and southern hemispheric observations? Or do the observed data fluctuate around a deterministic, perhaps non-linear trend wherein the mean (trend) of the series might exhibit occasional breaks or shifts? This recent debate is perhaps best characterized by five recent publications in the journal *Climatic Change*. These are: Gay-Garcia, Estrada and Sánchez (2009), Kaufmann, Kauppi and Stock (2010a), Estrada, Gay and Sánchez (2010), Mills (2010), and Kaufmann, Kauppi, Mann and Stock (2013). Arguing in favor of trend stationarity with occasional but infrequent mean breaks, Gay-Garcia et al. (2009) build on prior work by Perron (1989), Perron (1990), Leybourne, Newbold and Vougas (1998), Harvey and Mills (2002), and others who have developed tests for unit roots against the alternative of trend stationarity with occasional mean breaks (shifts). Related work that has concluded that temperature data are best characterized by unit roots and, possibly, stochastic trends include Harvey and Mills (2001), Kaufmann and Stern (2002), Liu and Rodríguez (2005), Kaufmann, Kauppi and Stock (2006a,b), Johansen (2010), Breusch and Vahid (2011). Relevant studies that have assumed that temperature series are stationary but that they follow a deterministic and possibly breaking trend include Harvey and Mills (2001), Seidel and Lanzante (2004), Gil-Alana (2008a,b), Ivanov and Evtimov (2010), and Breusch and Vahid (2011).¹

Considering unit roots and co-integration amounts to taking a global look at the nonstationary series under study. In the case of hemispheric temperature series it is quite natural to assume that both are nonstationary and possess a unit root. It is also natural to expect that the unit root in them is common in the sense there exists a linear combination of the series that eliminates it. One might even expect the coefficients of this combination to be 1 and -1 (or some nonzero multiples of them). Many papers already mentioned find this to be the case.

In this paper we shall nevertheless model these temperature series differently. Instead

¹Harvey and Mills (2001) report results for both stochastic trend models as well as models with deterministic trends that change (shift) in a potentially smooth manner. As well, Breusch and Vahid (2011) also explore the possibility of both stochastic and deterministic trends for available temperature data, with the latter being allowed to break at least once.

of global common features we are interested in local similarities between them. The idea is to try and find episodes or periods during which the time series have ‘behaved similarly.’ The starting–point of this approach is to assume that the series can be described by a number of deterministic shifts and stationary movements around them. Similarity is taken to mean that some of the shifts are common, that is, they can be eliminated by a linear combination of the two series. As a by–product, the series may be forecast by extrapolating the deterministic component, although such forecasts are not particularly reliable when the forecasting horizon increases.

Our analysis employs variants of the smooth transition autoregressive model in which the transition variable is time, see, for example, Lin and Teräsvirta (1994) and González and Teräsvirta (2008). In fact, we generalise the latter model, that is, the Shifting–Mean Autoregressive Model (SM–AR) to the multivariate case, which results in the Vector Shifting–Mean Autoregressive (VSM–AR) model. In terms of a modelling framework the present paper is most closely related to that of Harvey and Mills (2001). Even so, these authors did not consider the possibility of co–shifting between the northern and southern hemisphere in their analysis.

The plan of the paper is as follows. In Section 2 we present the univariate SM–AR model and in Section 3 its multivariate counterpart, the VSM–AR model. Section 4 is devoted to modelling and a discussion of the concept of co–shifting. The hemispheric temperature series are introduced in Section 5. Univariate SM–AR results using these series are considered in Section 6, whereas Section 7 contains the multivariate VSM–AR results. Section 8 concludes and proofs appear in Appendix A.

2 Shifting–Mean Autoregressive Model

We begin with a brief review of the Shifting Mean Autoregressive (SM–AR) by González and Teräsvirta (2008). This model, which is an autoregressive model with a time–varying intercept is defined as follows:

$$y_t = \delta(t) + \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t, \quad (1)$$

where $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ and where $\delta(t)$ is an intercept term that possibly varies with time. The time–varying intercept is a linear combination of logistic functions:

$$\delta(t) = \delta_0 + \sum_{i=1}^q \delta_i g(t/T; \gamma_i, c_i), \quad (2)$$

where

$$g(t/T; \gamma_i, c_i) = (1 + \exp\{-\gamma_i(t/T - c_i)\})^{-1}, \quad \gamma_i > 0, \quad (3)$$

and T is the number of observations. It follows that the value of $\delta(t/T)$ changes (possibly nonmonotonically) from δ_0 to $\delta_0 + \sum_{j=1}^q \delta_j$ as a function of t . The definition (3) implies that $g(t/T; \gamma_j, c_j)$ and thus $\delta(t/T)$ is continuous and infinitely many times differentiable in γ_j and c_j . Rescaled time in the argument of $g(t/T; \gamma_j, c_j)$ leaves the relative location of transitions intact as $T \rightarrow \infty$. Assume that the roots of the polynomial $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$ lie outside the unit circle. Then the expectation of y_t at t/T equals

$$\mathbf{E}y_t = \phi^{-1}(1)\delta(t/T).$$

The specification given by (1)–(3) provides considerable flexibility in modelling shifting means in time series data, depending on the number of shifts, q , and the values taken by the parameters. For example, large values of γ_i imply that the underlying shift, whose mid-point occurs at time c_i , becomes rather abrupt. Alternatively, for small values of γ_i (and assuming for the moment that $q = 1$), the shift from δ_0 to $\delta_0 + \delta_1$ is smooth, and will take more time to complete.

3 Vector Shifting–Mean Autoregressive Model

In order to consider the two temperature series jointly we generalise the SM–AR model into a vector model as follows. Let

$$\mathbf{y}_t = \boldsymbol{\delta}(t/T) + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad (4)$$

where $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})'$ is a $k \times 1$ stochastic vector, $\boldsymbol{\varepsilon}_t \sim \text{iid}(\mathbf{0}, \boldsymbol{\Omega})$, where $\mathbf{E}\boldsymbol{\varepsilon}_t = \mathbf{0}$, $\boldsymbol{\Omega}$ is a $k \times k$ positive definite covariance matrix, $\boldsymbol{\Phi}_j$, $j = 1, \dots, p$, are $k \times k$ parameter matrices, and T is the number of observations. Furthermore, $\boldsymbol{\delta}(t/T) = (\delta_1(t/T), \dots, \delta_k(t/T))'$ is a $k \times 1$ time-varying intercept vector comparable to (2), where

$$\delta_j(t/T) = \delta_{j0} + \sum_{i=1}^{q_j} \delta_{ji} g(t/T; \gamma_{ji}, c_{ji}). \quad (5)$$

It follows that the time-varying intercept vector in (4) equals

$$\boldsymbol{\delta}(t/T) = \boldsymbol{\delta}_0 + \sum_{i=1}^q \mathbf{G}_i(t/T) \boldsymbol{\delta}_i, \quad (6)$$

where $\boldsymbol{\delta}_i$ is a $k \times 1$ parameter vector, $i = 0, 1, \dots, q$. Furthermore, $\mathbf{G}_i(t/T)$, are diagonal matrices defined as:

$$\mathbf{G}_i(t/T) = \text{diag}(g(t/T; \gamma_{i1}, c_{i1}), \dots, g(t/T; \gamma_{ik}, c_{ik})), \quad (7)$$

where, as before,

$$g(t/T; \gamma_{ji}, c_{ji}) = (1 + \exp\{-\gamma_{ji}(t/T - c_{ji})\})^{-1}, \quad \gamma_{ji} > 0, \quad (8)$$

for $j = 1, \dots, k$, and $i = 1, \dots, q$. Equations (4)–(8) define the Vector Shifting–Mean Autoregressive Model.

We make the following assumptions:

Assumption A1. *The roots of $|\mathbf{I} - \sum_{j=1}^p \boldsymbol{\Phi}_j z^j| = 0$ lie outside the unit circle.*

Assumption A2. *In the transition function $g(t/T; \gamma_i, c_i)$, $\gamma_i > 0$, $i = 1, \dots, q$; $c_1 < \dots < c_q$. This implies $g(t/T; \gamma_i, c_i) \neq g(t/T; \gamma_j, c_j)$ for $i \neq j$.*

If this assumption is relaxed such that $c_j = c_{j+1}$ for some j , then $g(t/T; \gamma_j, c_j) \neq g(t/T; \gamma_{j+1}, c_{j+1})$ requires $\gamma_j \neq \gamma_{j+1}$.

Assumption A3. *Parameter space Θ^* is compact, the true parameter $\boldsymbol{\theta}_0^*$ is an interior point of Θ^* .*

Assumption A4. *The density is positive (bounded away from zero) for all $\boldsymbol{\theta}^* \in \Theta^*$.*

Assumption A5. *The errors $\boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is positive definite.*

From these assumptions it follows that while the VSM–AR model is nonstationary, $\{\mathbf{y}_t - \boldsymbol{\delta}(t/T)\}$ is a stationary and ergodic sequence. Let $\boldsymbol{\Phi}(z) = \mathbf{I} - \sum_{j=1}^p \boldsymbol{\Phi}_j z^j$, assuming that A1 holds. Its inverse exists and equals $\boldsymbol{\Phi}^{-1}(z) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i z^i = \boldsymbol{\Psi}(z)$. Set $\boldsymbol{\Psi}(1) = \boldsymbol{\Psi}$. Then, analogously to the univariate case,

$$\mathbf{E}\mathbf{y}_t = \boldsymbol{\Psi}\boldsymbol{\delta}(t/T). \quad (9)$$

According to A3, the elements of $\boldsymbol{\delta}(t/T)$ are bounded. It follows that the shifting mean (9) is bounded as well, and \mathbf{y}_t is bounded in probability. In this respect the VSM–AR model differs from nonstationary VARs with stochastic as well as broken linear trends, see Kaufmann, Kauppi and Stock (2006a,b, 2010b) and others. It is also different from the smooth transition trend model of Harvey and Mills (2002).

4 Modelling with the VSM–AR

4.1 Specification of the Model

In practice, the number of lags in (4) and the number of transitions (logistic functions) in $\delta_j(t/T)$ are unknown and have to be determined from the data. It may be noted there is a potential identification problem due to the construction of (6). If $\gamma_{ij} = 0$, δ_{ij} and c_{ij} are unidentified nuisance parameters. This is a common problem in many nonlinear time series models; see for example Teräsvirta, Tjøstheim and Granger (2010, Chapter 5) for discussion. As a result, specification of the number of transitions q_j proceeds from specific to general.

In this work we first determine the number of transitions and the lag length p thereafter. The number of transitions for the j th equation, q_j , is essentially determined using the specification technique suggested by White (2006), called **QuickNet**. Following González, Hubrich and Teräsvirta (2009), we call our variant **QuickShift**. By using **QuickNet** or **QuickShift**, a nonlinear specification and estimation problem is converted into a linear model selection and estimation problem. For the j th equation, this problem is one of choosing a subset of transition functions from the set

$$S_j = \left\{ (1 + \exp \{ -\gamma_{ji}(t/T - c_{ji}) \})^{-1}, i = 1, \dots, M \right\}, \quad (10)$$

where M is large. The set is obtained by constructing a grid of points (γ_{ji}, c_{ji}) and computing the values of the corresponding logistic function. It is clear that the quality of the estimates depends on the size of S_j . For this reason, in an application like the present one with 162 observations, the number of elements in S_j is likely to exceed the number of observations. The transitions are chosen from S_j as follows. First regress y_{jt} on an intercept (i.e., mean center y_{jt}). Next, choose the element of S_j that is most strongly correlated with the mean centered y_{jt} (has the strongest contribution to the explanation of y_{jt}) and add it to the regression. Compute the residuals from this regression. Choose the element of S_j that has the largest (absolute) correlation with this residual and add it in turn to the regression. Continue using the same rule until a model selection criterion tells one to stop. White (2006) used BIC by Rissanen (1978) and Schwarz (1978).

In **QuickShift** we instead test the hypothesis that the latest added regressor in the new regression has coefficient zero. If the original significance level equals α_0 , for reasons of parsimony the significance level of the k th test equals $\alpha_{k-1} = \tau^{k-1}\alpha_0$, where $0 < \tau < 1$. Testing and adding transitions is continued until the first non-rejection. Since the dynamics, the lags of \mathbf{y}_t , see (4), are ignored, the Newey–West standard deviation estimates are applied in the tests. Having found q_j , the maximum lag, p_j , may be determined using the same technique. In this paper, however, we

choose the same lag length for the whole system; see Section 6.

4.2 Estimation of Parameters

The accuracy of the linear estimates obtained by using the grid may already be sufficient for many practical purposes if the grid is sufficiently dense. Alternatively, one can use the parameter values thus obtained as starting-values for nonlinear maximum likelihood (ML) estimation. For inference, we have to consider asymptotic properties of our ML estimators. In order to discuss them, note that the parameters of the model are estimated equation by equation. This holds for a linear VAR model, see e.g., Lütkepohl (2005, pp. 70–72) but is also true for the VSM–AR model in which nonlinearity is restricted to the deterministic intercept. Define $\Phi_j = [\phi'_{j1}, \dots, \phi'_{jk}]'$, where ϕ_{jm} is the m th row of Φ_j . Setting $\mathbf{y}_{t-m} = (\mathbf{y}'_{1,t-m}, \dots, \mathbf{y}'_{k,t-m})'$, the i th equation of (4) can be written as

$$y_{jt} = \delta_j(t/T) + \sum_{m=1}^{p_j} \phi'_{jm} \mathbf{y}_{t-m} + \varepsilon_{jt}, \quad (11)$$

$j = 1, \dots, k$. Let θ_0 be the true vector of parameters in (11). Asymptotic properties of the maximum likelihood estimator $\hat{\theta}$ of θ_0 are studied in Appendix A.

Since an analytic solution to the problem of maximising the log-likelihood function of (11) does not exist, we have to choose a suitable algorithm for the purpose; see, for example, Teräsvirta, Tjøstheim and Granger (2010, Chapter 12). Here we apply the Broyden, Fletcher, Goldfarb and Shanno (BFGS) quasi-Newton method. QuickShift is used to provide the initial values for the algorithm. Following Goodwin, Holt and Prestemon (2011), for numerical reasons we define $\gamma = \exp(\eta_i)$ and estimate η_i .²

4.3 Co-shifting

As mentioned in the Introduction, local similarities in the two temperature series are our main concern in this study. They are defined through shifts (if any) in the intercept vector of the VSM–AR model. We are interested in knowing whether several of more of the shifts in the equations of our model are common, that is, are shared by both equations. Putting it more formally, the shifting mean is a feature, and if there is a linear combination of the elements of \mathbf{y}_t such that the feature is eliminated, it is

²This parametrization for γ_i automatically ensures that $\gamma_i > 0$ holds, which implies that the identification condition $\gamma_i > 0$ holds. Moreover, it facilitates a grid search wherein equal spacings between (large) γ_i values are not optimal.

a common feature. To illustrate this in our framework, consider a simple bivariate VSM-AR model with two shifts in the intercept:

$$\mathbf{y}_t = \boldsymbol{\delta}_0 + \mathbf{G}_1(t/T)\boldsymbol{\delta}_1 + \mathbf{G}_2(t/T)\boldsymbol{\delta}_2 + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad (12)$$

where

$$\mathbf{G}_i(t/T) = \text{diag}(g(t/T; \gamma_{i1}, c_{i1}), g(t/T; \gamma_{i2}, c_{i2})),$$

and $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2})'$, $i = 1, 2$. The shifting intercept is a common feature if there is a two-dimensional vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\mathbf{y}_t$ is linear, that is, the two shifts in (12) have been eliminated. A necessary condition for this to happen is that $g(t/T; \gamma_{ij}, c_{ij}) = g(t/T; \gamma_i, c_i)$, $j = 1, 2$, or, in matrix notation:

$$\mathbf{G}_i(t/T) = g(t/T; \gamma_i, c_i)\mathbf{I}_2, \quad i = 1, 2, \quad (13)$$

where \mathbf{I}_2 is a (2×2) identity matrix. This implies that the slope and location parameters in the two transition functions in both equations are the same. If the necessary condition (13) is completed by the condition $\boldsymbol{\delta}_2 = \lambda\boldsymbol{\delta}_1$, $\lambda \neq 0$, there exists a vector $\boldsymbol{\alpha} \neq \mathbf{0}$ that eliminates the shift. This is called (strong) co-shifting.

The latter condition of linear dependence is indeed quite strong, in particular if the number of transitions is large. Therefore, defining partial or weak co-shifting is of interest. For example, assume

$$\mathbf{G}_1(t/T) = g(t/T; \gamma_1, c_1)\mathbf{I}_2.$$

Then there exists a nonzero vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\delta}_1 = 0$, which eliminates the first but not necessarily the second shift:

$$\boldsymbol{\alpha}'\mathbf{y}_t = \boldsymbol{\alpha}'\boldsymbol{\delta}_0 + \boldsymbol{\alpha}'\mathbf{G}_2(t/T)\boldsymbol{\delta}_2 + \sum_{j=1}^p \boldsymbol{\alpha}'\boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\alpha}'\boldsymbol{\varepsilon}_t.$$

This is an example of weak co-shifting. If also

$$\mathbf{G}_2(t/T) = g(t/T; \gamma_{21}, c_{21})\mathbf{I}_2$$

there exists another nonzero vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\boldsymbol{\delta}_2 = 0$ and

$$\boldsymbol{\beta}'\mathbf{y}_t = \boldsymbol{\beta}'\boldsymbol{\delta}_0 + \boldsymbol{\beta}'\mathbf{G}_1(t/T)\boldsymbol{\delta}_1 + \sum_{j=1}^p \boldsymbol{\beta}'\boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t.$$

As already discussed, a single vector $\boldsymbol{\alpha}$ can eliminate both shifts only if $\boldsymbol{\delta}_2 = \lambda\boldsymbol{\delta}_1$.

Note that co-shifting does not mean that the shifts contribute to holding the two series together. This property of linear cointegration is not present in co-shifting. This is because the elements of $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are not restricted to mimic this feature of cointegrated random variables.

The definition of weak co-shifting accords well with the definition of contemporaneous mean co-breaking, as stated in Hendry and Mizon (1998) and Hendry and Massmann (2007). These authors consider a location shift in the unconditional mean of a vector autoregressive process at time t : $\mathbf{E}(\mathbf{y}_t - \boldsymbol{\rho}_0) = \boldsymbol{\tau}_t$. If there is a time-point t such that $\boldsymbol{\tau}_t \neq \mathbf{0}$, this defines a structural break. If there is a vector $\boldsymbol{\alpha}$ that cancels the break, $\boldsymbol{\alpha}'\boldsymbol{\tau}_t = 0$, then $\boldsymbol{\alpha}$ is contemporaneous co-breaking for the sequence \mathbf{y}_t of order one. The definition in Hendry and Mizon (1998) is more general than what this simple example suggests. It stipulates that the break may be eliminated by reducing the dimension of the system from k to ℓ , $\ell < k$, and not necessarily to one as in our example. In our bivariate case, the mean $\boldsymbol{\rho}_t$ equals

$$\begin{aligned}\boldsymbol{\rho}_t &= \boldsymbol{\rho}^{(0)}I(t/T < c) + \boldsymbol{\rho}^{(1)}\{1 - I(t/T < c)\} \\ &= \boldsymbol{\rho}^{(1)} + (\boldsymbol{\rho}^{(0)} - \boldsymbol{\rho}^{(1)})I(t/T < c), \quad \boldsymbol{\rho}^{(0)} \neq \boldsymbol{\rho}^{(1)}\end{aligned}$$

where $I(A)$ is an indicator function: $I(A) = 1$ when A holds, zero otherwise. The break is eliminated when there exists $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'(\boldsymbol{\rho}^{(0)} - \boldsymbol{\rho}^{(1)}) = 0$. This can occur only if $\boldsymbol{\rho}^{(0)} = \lambda\boldsymbol{\rho}^{(1)}$, $\lambda \neq 0$.

4.4 Testing Co-shifting Restrictions

Co-shifting may be viewed as a special case of common nonlinearity as defined by Anderson and Vahid (1998). They derived a general test of common nonlinearity as a test of overidentifying restrictions in the generalized method of moments framework. Our test of co-shifting is simply a test of parameter restrictions in the VSM-AR model. If we test against strong co-shifting, the null hypothesis is

$$H_0 : (\gamma_{i1}, c_{i1}) = \dots = (\gamma_{ik}, c_{ik}), \quad i = 1, \dots, q, \quad (14)$$

and $\boldsymbol{\delta}_j = \lambda_j\boldsymbol{\delta}_1$, $\lambda_j \neq 0$, $j = 2, \dots, q$. This amounts to testing $k(3q - 1)$ restrictions in (4). If only a subset of shifts are under test, the number of restrictions decreases. As already mentioned, the number of transitions need not be the same in all equations of (4).

To implement tests of co-shifting in the VSM-AR model it is natural to consider a likelihood ratio test. The test statistic is defined as:

$$LR = T \left[\ln |\tilde{\Omega}| - \ln |\hat{\Omega}| \right], \quad (15)$$

where LR has an asymptotic $\chi_{k(3q-1)}^2$ distribution under the null hypothesis. This requires $\hat{\boldsymbol{\theta}}$ to be consistent and asymptotically normal; see Appendix A. In (15) $\hat{\Omega}$ denotes the maximum likelihood estimate of the residual covariance matrix for the general (unrestricted) model and $\tilde{\Omega}$ the corresponding estimates for the co-shifting (restricted) model that involves $k(3q-1)$ restrictions on the parameters of (4).

A well known problem with the test statistic in (15) is that its asymptotic $\chi_{k(3q-1)}^2$ null distribution is not a good approximation to its finite-sample null distribution when the dimension of the model and the null hypothesis are large compared to the length of the time series. In that case the test suffers from positive size distortion; see, for example, Candelon and Lütkepohl (2001) and Shukur and Edgerton (2002). A number of remedies have been proposed, but simulations almost invariably show that the best solution to the problem is to use Rao's F -test, see for example (Rao, 1973, p. 556). This does require, however, that the errors can be assumed independent and identically distributed. If the presence of conditional heteroskedasticity is suspected, one can generate a finite-sample null distribution of the test statistic by wild bootstrap; see, for example, Ahlgren and Catani (in press). This can be done even if there is no conditional or unconditional heteroskedasticity, but in that case computational ease speaks for Rao's F -test. Additional details on computing and using Rao's F in the context of an LM-based test for stationary vector autoregressive systems are provided by Teräsvirta, Tjøstheim and Granger (2010, pp. 100–102).

4.5 Evaluation

The estimated VSM-AR model has to be evaluated. One has to check whether or not the stability condition concerning the roots of $|\mathbf{I} - \sum_{j=1}^p \boldsymbol{\Phi}_j z^j| = 0$ holds. Misspecification tests need to be carried out. They include the multivariate test of normality, see Lütkepohl and Krätzig (2004, p. 128), the multivariate test of no error autocorrelation adopted from Yang (2012), and the test of constancy of the error covariance matrix by Eklund and Teräsvirta (2007). Their framework can be used for testing constancy against various alternatives. In this work the alternative is that the variances are changing smoothly over time whereas the correlations remain constant. It would also be possible to develop a test of linearity for the VAR component against vector smooth transition AR. In our application, however, this component is rather minor, so we do not require such a test here.

5 Data

The data used in the empirical analysis are annual average hemispheric temperature series from 1850–2014, and are described in detail in Brohan et al. (2006). As mentioned previously, versions of these time series have been used by various authors in recent years to empirically investigate features of global warming, including Gay-Garcia, Estrada and Sánchez (2009), Kaufmann, Kauppi and Stock (2010a), Estrada, Gay and Sánchez (2010), and Mills (2010). A time series plot of the basic data is reported in Figure 1. As illustrated there, temperatures in both hemispheres appear to have a slight downward trend between approximately 1880 and 1910, with temperatures in the southern hemisphere showing what appears to be a slightly steeper decline. Then, from about 1910 until approximately 1945 both series exhibit an upward trend, with temperatures in the northern hemisphere appearing to increase more rapidly than those in the southern. There is then a leveling off between the 1940s and approximately 1980, after which both series exhibit a rather steep upward trend and, moreover, appear to increase at approximately the same rate. These observations are based on a casual inspection of the data and trends in Figure 1; formal model specification, estimation, and testing will follow next.

6 VSM–AR Estimation and Results

We first specify and estimate our VSM-AR model by `QuickShift` and use the resulting estimates as starting values for the nonlinear estimation algorithm.³ The system lag length, p , is determined by using a sequence of likelihood ratio tests; coincident with results reported by Harvey and Mills (2001), we find that $p = 3$ is adequate. In specifying the VSM–AR model, we initially followed Kaufmann and Stern (1997) and Harvey and Mills (2001), who reported that lags of northern temperatures could be excluded from the southern equation. A Granger non-causality test performed with respect to lags of northern temperatures in the southern hemisphere equation strongly indicates, however, that these exclusion restrictions should be rejected. This result is a consequence of the fact that the deterministic component in our model is different from that of the previous models. To that end, the model was respecified and re-estimated to include lags of northern temperatures in the southern hemisphere equation.

³In obtaining ML estimates we follow established practice (see, e.g., van Dijk, Strikholm and Teräsvirta, 2003) and constrain values of γ_i to be bounded above, in this case at 50. Doing so helps avoid potential numerical problems. As with our implementation of `QuickShift`, we also constrain the values of c_i parameters to be bounded on the $[0.05, 0.95]$ interval, although this may not be necessary.

The model with additional mean shifts is only identified under the alternative hypothesis, and the ensuing identification problem is circumvented by approximating the additional transitions by a (third-order) Taylor expansion as in Luukkonen, Saikkonen and Teräsvirta (1988). The null hypothesis cannot be rejected at any plausible significance level. Based on these results we conclude that the VSM-AR model with four mean shifts in the northern equation; with three mean shifts in the southern equation; and with lagged temperatures from the north appearing in the equation for the south, is a reasonable representation of the hemispheric temperature data.

Next we turn to our main question: do the two temperature series exhibit any form of co-shifting? Visual inspection gives at least some reason to believe that the hemispheric temperature data have at least one and likely two intercept shifts in common. The possibilities include a long, relatively slow shift that started in the second half of the 20th century and another shift that occurred during the 1930s. The null hypothesis to be tested is that these two logistic transitions (but not their coefficients) are identical. The likelihood ratio test statistic defined in (15) thus has four degrees of freedom as each transition function contains two parameters.

Test results associated with this co-shifting hypothesis are reported in Table 1 where it is seen that the null hypothesis is not rejected. Since the model is only two-dimensional, the asymptotic χ^2 -test and Rao's F yield almost identical p -values. The model is re-estimated with these restrictions imposed. Table 2 contains misspecification tests of the estimated equations of this model and of the whole system. Normality of errors is tested both for each equation separately and for the whole system and is not rejected. It appears that the errors are not autocorrelated and that the error covariance matrix is stable over time.

Estimation results for the VSM-AR are reported in Table 2; information regarding the timing and nature of the estimated logistic function shifts in the model is summarized in Table 3. The estimation results have interesting interpretations. The coefficients of the common slow transition with $c = 0.95$ has different signs in the northern and southern hemisphere equations. The positive sign in the northern hemisphere equation clearly implies that the shift describes the long positive trend-like movement in that temperature series visible in Figure 2. Furthermore, this figure suggests that the same shift in the southern hemisphere equation contributes to explaining the long but slow downward movement from 1850 to 1920. To compensate for this, the southern hemisphere equation contains a positive shift with $\hat{c} = 0.908$. The slope coefficient of this transition is much larger than that of the slow transition, so the net effect remains positive. This is a case in which co-shifting does not imply that the two series are moving together in a similar fashion but rather that they have similar episodes *at different periods of time*. An extra shift is required for describing the long-run increase in the two series.

The second common transition has positive coefficients in both equations. They are not, however, equal as the coefficient estimate in the southern equation is greater than the corresponding one in the northern equation. The shift is therefore more pronounced in the latter, as Figure 2 shows.

Of the remaining transitions the first one in the northern hemisphere equation is there to counteract the slow shift with a positive coefficient, as the upward movement from 1850 to 1880 changes to a downward drift to match the downward movement in the southern hemisphere series. The sharp shift around 1945 is taking care of the rapid temperature change in the northern hemisphere at that time. Put together, the shifting mean indicates that while the two series undoubtedly move together, they do display individual local tendencies.

It is seen from (9) that all transitions, common or not, contribute to both equations through the lag structure. It is also obvious that the shifts belonging to a certain equation dominate the explanation in that equation, because the dynamic structure is not very strong. Indeed, the largest root of the VAR component is real but only equals 0.508, which means that the effects of shifts belonging to the other equation remain small. In fact, the infinite sum in (9) can be well approximated by truncating it already after the fourth lag.

7 Forecasting

We construct forecasts from our VSM-AR model through 2035 by simulation. The simulated random errors are drawn from the multivariate normal distribution with the estimated covariance matrix $\hat{\Omega}$ for the VSM-AR model. The forward simulations are repeated 25,000 times. Actual values over the sample period (1860–2014) and simulated values (2015–2035) are illustrated in Figure 3. Approximate 95% confidence intervals are represented by the shaded areas during the forecasting period. Uncertainty due to estimated parameters is not, however, accounted for.

According to the results shown in Figure 3, temperatures in both hemispheres will likely continue to increase for at least the next fifteen years. Temperatures in the south are predicted to increase at the rate that is slowing relative to that in the north. As noted in Table 3, this is a consequence of the fact that the latest shift in the southern hemisphere equation reaches its inflection point around 1999. Because this shift is not as slow as the positive shift in the northern hemisphere equation, the transition function and thus the predicted shifting mean begin to bend much more quickly than the corresponding transition in the northern equation. This causes the impression that the southern temperature will permanently increase more slowly than the northern one. That hardly seems plausible and demonstrates the fact that

these forecasts may only have relevance in the short but not in the long run. Their confidence intervals, however, do overlap. In the long run, new transitions cannot be excluded either. That possibility is not included in the forecasts or, rather, forward projections of the present tendencies.

8 Conclusions

In this paper we consider local similarities between the northern and southern hemisphere annual temperature series. In order to do that we generalise the switching-mean autoregressive model by González and Teräsvirta (2008) into a vector model. Within this model we define the concept of co-shifting with which considering local similarities becomes possible.

Under regularity conditions we prove consistency and asymptotic normality for maximum likelihood estimators of the parameters of the VSM-AR model. Since the model contains deterministic components, the asymptotic theory we apply is triangular array asymptotics.

In the empirical analysis the modelling sequence begins by using the QuickShift procedure developed by González and Teräsvirta (2008) to identify the number of relevant shifts in each series. When we apply this approach to the hemispheric temperature series we find that four logistic function components are adequate to characterize the shifts in the mean of the northern series whilst only three are required for the southern series. After estimating the parameters of the model by maximum likelihood, subsequent testing reveals that two logistic function components are common to both equations. Interpreting the model with (weak) co-shifting is not straightforward because one of the common shifts does not have the same sign in both equations.

Importantly, the estimated VSM-AR has been subjected to a battery of diagnostic and evaluative tests. The estimated model passes them. We are therefore confident that the estimated VSM-AR model of hemispheric temperatures provides a reasonable representation of the data and, moreover, an accurate representation of the mean shifts that have occurred in these series over time.

Appendix A

The Model

The i th equation of the VSM-AR model (4)-(??) has the following form:

$$y_{it} = \delta_i(t/T) + \sum_{j=1}^p \phi'_{ij} \mathbf{y}_{t-j} + \varepsilon_{it}, \quad (\text{A.1})$$

where ϕ'_{ij} is the i th row of Φ_j , and the i th element of the intercept vector equals

$$\delta_i(t/T) = \delta_{i0} + \sum_{j=1}^q \delta_{ij} g(t/T; \gamma_{ij}, c_{ij}).$$

In what follows we omit the subscript i .

The log-likelihood function (T observations) of (A.1) is defined as follows:

$$L_T(\boldsymbol{\theta}, \boldsymbol{\varepsilon}) = \sum_{t=1}^T \ell(\varepsilon_t; \boldsymbol{\theta}), \quad (\text{A.2})$$

where

$$\ell(\varepsilon_t; \boldsymbol{\theta}) = k - (1/2) \ln \sigma^2 - \frac{\varepsilon_t^2}{2\sigma^2}. \quad (\text{A.3})$$

In (A.3), σ^2 is the i th diagonal element of Σ . The parameter vector $\boldsymbol{\theta} = (\phi'_1, \dots, \phi'_p, \boldsymbol{\delta}', \boldsymbol{\gamma}', \mathbf{c}')$ $\in \Theta \subset \mathbb{R}^{3q+p+1}$, where $\phi_i = (\phi_{i1}, \dots, \phi_{ik})' = \text{vec}(\Phi'_i)$ is a $k \times 1$ vector, $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots, \delta_q)'$ is a $(q+1) \times 1$ vector, and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)'$ and $\mathbf{c} = (c_1, \dots, c_q)'$ are $q \times 1$ vectors. Let $\boldsymbol{\theta}_0 = (\phi'_{10}, \dots, \phi'_{p0}, \boldsymbol{\delta}'_0, \boldsymbol{\gamma}'_0, \mathbf{c}'_0)'$ be the corresponding true parameter vector.

In order to consider the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}_0$, we first define the score of (A.3). It appears (for observation t) in the following lemma:

Lemma A.1. *The $(kp+3q+1) \times 1$ score function $\partial \ell(\varepsilon_t; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ of (A.3) for observation t has the form*

$$\partial \ell(\varepsilon_t; \boldsymbol{\theta}) / \partial \boldsymbol{\theta} = \frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\varepsilon_t | \boldsymbol{\theta}) = - \frac{\varepsilon_t}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\theta}} = \frac{\varepsilon_t}{\sigma^2} \mathbf{g}(t/T; \boldsymbol{\theta}),$$

where $\mathbf{g}(t/T; \boldsymbol{\theta}) = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, \mathbf{g}'_{\delta}(t/T), \mathbf{g}'_{\gamma}(t/T), \mathbf{g}'_{\mathbf{c}}(t/T))'$, and T is the number of

observations. The blocks of $\mathbf{g}(t/T; \boldsymbol{\theta})$ are

$$\begin{aligned}\mathbf{y}_{t-j} &= \partial g(t/T)/\partial \boldsymbol{\phi} = (y_{1,t-j}, \dots, y_{k,t-j})', \quad j = 1, \dots, p, \\ \mathbf{g}_\delta(t/T) &= \partial g(t/T)/\partial \boldsymbol{\delta} = (1, g_1(t/T), \dots, g_q(t/T))', \\ \mathbf{g}_\gamma(t/T) &= \partial g(t/T)/\partial \boldsymbol{\gamma} = (g_{\gamma 1}(t/T), \dots, g_{\gamma q}(t/T))', \\ \mathbf{g}_c(t/T) &= \partial g(t/T)/\partial \mathbf{c} = (g_{c1}(t/T), \dots, g_{cq}(t/T))',\end{aligned}$$

where $g_{\gamma j}(t/T) = \delta_j g_j(t/T) \{1 - g_j(t/T)\} (t/T - c_j)$ and $g_{c j}(t/T) = -\gamma_j \delta_j g_j(t/T) \{1 - g_j(t/T)\}$ for $j = 1, \dots, q$.

The Hessian matrix of (A.3) for observation t is given in the following lemma:

Lemma A.2. *The Hessian $\partial^2 \ell(\varepsilon_t; \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ for observation t equals*

$$\partial^2 \ell(\varepsilon_t; \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' = -\frac{1}{\sigma^2} \left\{ \mathbf{g}(t/T; \boldsymbol{\theta}) \mathbf{g}'(t/T; \boldsymbol{\theta}) + \varepsilon_t(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{g}(t/T; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\},$$

where

$$\begin{aligned}\mathbf{g}(t/T; \boldsymbol{\theta}) \mathbf{g}'(t/T; \boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{M}_{\phi\phi}(t/T) & \mathbf{M}_{\phi\delta}(t/T) & \mathbf{M}_{\phi\gamma}(t/T) & \mathbf{M}_{\phi c}(t/T) \\ & \mathbf{M}_{\delta\delta}(t/T) & \mathbf{M}_{\delta\gamma}(t/T) & \mathbf{M}_{\delta c}(t/T) \\ & & \mathbf{M}_{\gamma\gamma}(t/T) & \mathbf{M}_{\gamma c}(t/T) \\ & & & \mathbf{M}_{cc}(t/T) \end{bmatrix}, \quad (\text{A.4})\end{aligned}$$

with

$$\mathbf{M}_{\phi\phi}(t/T) = \begin{bmatrix} \mathbf{M}_{\phi\phi 11}(t/T) & \dots & \mathbf{M}_{\phi\phi 1p}(t/T) \\ \mathbf{M}_{\phi\phi p1}(t/T) & \dots & \mathbf{M}_{\phi\phi pp}(t/T) \end{bmatrix}, \quad (\text{A.5})$$

and

$$\mathbf{M}_{\phi\alpha}(t/T) = [\mathbf{M}'_{\phi\alpha 1}(t/T) \quad \dots \quad \mathbf{M}'_{\phi\alpha p}(t/T)]'. \quad (\text{A.6})$$

In (A.5) $\mathbf{M}_{\phi\phi ij}(t/T) = \mathbf{y}_{t-i} \mathbf{y}'_{t-j}$, and in (A.6), $\mathbf{M}_{\phi\alpha j}(t/T) = \mathbf{y}_{t-j} \mathbf{g}'_{\phi\alpha}(t/T)$, $i, j = 1, \dots, p$ and $\alpha = \delta, \gamma, c$. Finally, in (A.4),

$$\begin{aligned}\mathbf{M}_{\delta\delta}(t/T) &= \mathbf{g}_\delta(t/T) \mathbf{g}'_\delta(t/T), \\ \mathbf{M}_{\delta\alpha}(t/T) &= \mathbf{g}_\delta(t/T) \mathbf{g}'_\alpha(t/T) + \varepsilon_t \mathbf{D}_{\delta\alpha}(t/T), \quad \alpha = \gamma, c, \\ \mathbf{M}_{\gamma\alpha}(t/T) &= \mathbf{g}_\gamma(t/T) \mathbf{g}'_\alpha(t/T) + \varepsilon_t \text{diag}(g_{\gamma\alpha 1}(t/T), \dots, g_{\gamma\alpha q}(t/T)), \quad \alpha = \gamma, c, \\ \mathbf{M}_{cc}(t/T) &= \mathbf{g}_c(t/T) \mathbf{g}'_c(t/T) + \varepsilon_t \text{diag}(g_{cc 1}(t/T), \dots, g_{cc q}(t/T)).\end{aligned}$$

where $\mathbf{D}_{\delta\alpha}(t/T) = [\mathbf{0}, \mathbf{g}_\alpha(t/T)]'$. The diagonal elements in the three diagonal matrices

are

$$\begin{aligned}
g_{\gamma\gamma_j}(t/T) &= \delta_j g_j(t/T) \{1 - g_j(t/T)\} \{1 - 2g_j(t/T)\} (t/T - c_j)^2, \\
g_{cc_j}(t/T) &= \delta_j \gamma_j^2 g_j(t/T) \{1 - g_j(t/T)\} \{1 - 2g_j(t/T)\}, \\
g_{\gamma c_j}(t/T) &= -\delta_j \gamma_j g_j(t/T) \{1 - g_j(t/T)\} \{1 - 2g_j(t/T)\} (t/T - c_j),
\end{aligned}$$

for $j = 1, \dots, q$.

Consistency

We begin by proving consistency of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ by using triangular array asymptotics; for a useful exposition, see Hillebrand, Medeiros and Xu (2013). We have the following result:

Theorem A1. *Consider equation (A.1) and suppose that Assumptions A1–A5 hold. Then the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$.*

Proof. We prove this result by verifying the conditions of Theorem 2.5 in Newey and McFadden (1994, p. 2131):

Theorem (Newey and McFadden). Suppose that ε_t ($i = 1, \dots, T$) are *iid* with probability distribution function $f(\varepsilon_t|\boldsymbol{\theta}_0)$. If

- (i) $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then $f(\varepsilon_t|\boldsymbol{\theta}_0) \neq f(\varepsilon_t|\boldsymbol{\theta})$,
- (ii) $\boldsymbol{\theta}_0 \in \Theta$ which is compact, and the density is positive (bounded away from zero) for all $\boldsymbol{\theta} \in \Theta$,
- (iii) $\ln f(\varepsilon_t|\boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta} \in \Theta$ with probability one,
- (iv) $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\ln f(\varepsilon_t|\boldsymbol{\theta})| < \infty$,

then the maximum likelihood estimator $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ as $T \rightarrow \infty$.

Assumption (i) is satisfied due to A2, (ii) is valid due to A3 and A4, and (iii) follows from Lemma A.1. To show (iv), apply the mean value theorem, the triangle inequality, and the Cauchy–Schwarz inequality to $|\ln f(\varepsilon_t|\boldsymbol{\theta})|$. This yields:

$$\begin{aligned}
|\ln f(\varepsilon_t|\boldsymbol{\theta})| &= |\ln f(\varepsilon_t|\boldsymbol{\theta}_0) + \frac{\partial}{\partial \boldsymbol{\theta}'} \ln f(\varepsilon_t|\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \\
&\leq |\ln f(\varepsilon_t|\boldsymbol{\theta}_0)| + \left\| \frac{\partial}{\partial \boldsymbol{\theta}'} \ln f(\varepsilon_t|\bar{\boldsymbol{\theta}}) \right\| \times \|(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \\
&\leq C_1 + C_2 C_3 < \infty,
\end{aligned} \tag{A.7}$$

where $\bar{\boldsymbol{\theta}}$ is an intermediate value between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, $|\ln f(\varepsilon_t|\boldsymbol{\theta}_0)| \leq C_1$ because of A4, and $\|\frac{\partial}{\partial \boldsymbol{\theta}'} \ln f(\varepsilon_t|\bar{\boldsymbol{\theta}})\| \leq C_2$ because of Lemma A.1. Furthermore, $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq C_3$ follows from the fact that Θ is compact (A3), so the elements of $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$ are bounded for $\boldsymbol{\theta} \in \Theta$. In (A.7), C_i , $i = 1, 2, 3$, are generic positive constants. As $|\ln f(\varepsilon_t|\boldsymbol{\theta})|$ is finite for all $\boldsymbol{\theta} \in \Theta$ (A4), the expectation of its supremum is finite as well. ■

Note that A5 is quite strong. It would be sufficient to assume $\varepsilon_t \sim \text{iid}(0, \sigma^2)$. Not assuming normality would mean that (A.7) would be a quasi log-likelihood for observation t , and the resulting estimator would be a quasi ML estimator.

Asymptotic Normality

In order to consider asymptotic normality of $\hat{\boldsymbol{\theta}}$, in addition to Lemma A.2 we need the five lemmas below.

Lemma A.3. *Let \mathbf{x}_t be a stationary and ergodic VAR(p) process with zero mean:*

$$\mathbf{x}_t = \sum_{j=1}^p \Phi_j \mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t,$$

where $\{\boldsymbol{\varepsilon}_t\} \sim \text{iid}\mathcal{N}(\mathbf{0}, \Sigma)$. It has the infinite-order moving average representation

$$\mathbf{x}_t = \sum_{j=0}^{\infty} \Psi_j \boldsymbol{\varepsilon}_{t-j},$$

where $\Psi_0 = \mathbf{I}$, $\sum_{j=0}^{\infty} |\Psi_j| < \infty$ (ergodicity), and

$$\mathbf{E} \mathbf{x}_t \mathbf{x}'_{t-m} = \sum_{j=0}^{\infty} \Psi_j \Sigma \Psi'_{j+m}.$$

Proof. Omitted. ■

Lemma A.4. *Consider the VSM-AR model (4)–(8):*

$$\mathbf{y}_t = \boldsymbol{\delta}(t/T) + \sum_{j=1}^p \Phi_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t,$$

where Assumptions A1–A5 hold. Let $\mathbf{M}_{\phi\phi mnT} = (1/T) \sum_{t=1}^T \mathbf{y}_{t-m} \mathbf{y}'_{t-n}$, where $n =$

$m + s, s \geq 0$. Then

$$\lim_{T \rightarrow \infty} \mathbf{E} \mathbf{M}_{\phi \phi m n T} = \mathbf{\Psi} \int_0^1 \boldsymbol{\delta}(r) \boldsymbol{\delta}'(r) dr \mathbf{\Psi}' + \sum_{i=0}^{\infty} \mathbf{\Psi}_i \boldsymbol{\Sigma} \mathbf{\Psi}'_{i+s},$$

where $\mathbf{\Psi} = \sum_{i=0}^{\infty} \mathbf{\Psi}_i$.

Proof. Write (4) as follows:

$$(\mathbf{I}_k - \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{L}^j) \mathbf{y}_t = \boldsymbol{\delta}(t/T) + \boldsymbol{\varepsilon}_t.$$

Since A1 holds, there exists the infinite-order moving average representation

$$\mathbf{y}_t = (\mathbf{I}_k - \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{L}^j)^{-1} \boldsymbol{\delta}(t/T) + \boldsymbol{\varepsilon}_t = \sum_{i=0}^{\infty} \mathbf{\Psi}_i \boldsymbol{\delta}\left(\frac{t-i}{T}\right) + \mathbf{x}_t,$$

where $\mathbf{x}_t = \sum_{i=0}^{\infty} \mathbf{\Psi}_i \boldsymbol{\varepsilon}_{t-i}$. Then

$$\mathbf{y}_t - \mathbf{E} \mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Psi}_i \left\{ \boldsymbol{\delta}\left(\frac{t-i}{T}\right) - \boldsymbol{\delta}\left(\frac{t}{T}\right) \right\} + \mathbf{x}_t$$

and, at time t ,

$$\begin{aligned} & \mathbf{E}(\mathbf{y}_{t-m} - \mathbf{E} \mathbf{y}_{t-m})(\mathbf{y}_{t-n} - \mathbf{E} \mathbf{y}_{t-n})' \\ &= \left[\sum_{i=0}^{\infty} \mathbf{\Psi}_i \left[\boldsymbol{\delta}\left(\frac{t-m-i}{T}\right) - \boldsymbol{\delta}\left(\frac{t-m}{T}\right) \right] \right] \left[\sum_{j=0}^{\infty} \mathbf{\Psi}_j \left[\boldsymbol{\delta}\left(\frac{t-n-j}{T}\right) - \boldsymbol{\delta}\left(\frac{t-n}{T}\right) \right] \right]' \\ & \quad + \mathbf{E} \mathbf{x}_{t-m} \mathbf{x}_{t-n}' \end{aligned}$$

where, from Lemma A.3,

$$\mathbf{E} \mathbf{x}_t \mathbf{x}'_{t-s} = \sum_{i=0}^{\infty} \mathbf{\Psi}_i \boldsymbol{\Sigma} \mathbf{\Psi}'_{i+s}.$$

Set $[Tr] = t$ and consider

$$\begin{aligned}
& \frac{1}{T-n} \sum_{t=n+1}^T \delta\left(\frac{t-m-i}{T}\right) \delta'\left(\frac{t-n-j}{T}\right) \\
&= \frac{1}{T-n} \sum_{t=n+1}^T \int_{t/T}^{(t+1)/T} \delta\left(\frac{[Tr]-m-i}{T}\right) \delta'\left(\frac{[Tr]-n-j}{T}\right) dr \\
&= \int_{(n+1)/T}^{(T+1)/T} \delta\left(\frac{[Tr]-m-i}{T}\right) \delta'\left(\frac{[Tr]-n-j}{T}\right) dr \\
&\rightarrow \int_0^1 \delta(r) \delta'(r) dr, \tag{A.8}
\end{aligned}$$

for $m = 0, 1, \dots, p$, $n \geq m$, as $T \rightarrow \infty$. It follows that

$$\frac{1}{T-n} \sum_{t=n+1}^T \mathbf{E}(\mathbf{y}_{t-m} - \mathbf{E}\mathbf{y}_{t-m})(\mathbf{y}_{t-n} - \mathbf{E}\mathbf{y}_{t-n})' \rightarrow \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi_{i+s}',$$

as $T \rightarrow \infty$. Next,

$$\frac{1}{T-n} \sum_{t=n+1}^T \mathbf{E}\mathbf{y}_{t-m} \mathbf{E}\mathbf{y}'_{t-n} = \Psi \delta\left(\frac{t-m}{T}\right) \delta'\left(\frac{t-n}{T}\right) \Psi',$$

so, from (A.8),

$$\lim_{T \rightarrow \infty} \frac{1}{T-n} \sum_{t=n+1}^T \mathbf{E}\mathbf{y}_{t-m} \mathbf{E}\mathbf{y}'_{t-n} = \Psi \int_0^1 \delta(r) \delta'(r) dr \Psi'.$$

Since

$$\mathbf{E}\mathbf{y}_{t-m} \mathbf{y}'_{t-n} = \mathbf{E}\mathbf{y}_{t-m} \mathbf{E}\mathbf{y}'_{t-n} + \mathbf{E}(\mathbf{y}_{t-m} - \mathbf{E}\mathbf{y}_{t-m})(\mathbf{y}_{t-n} - \mathbf{E}\mathbf{y}_{t-n})',$$

one obtains

$$\begin{aligned}
\mathbf{E}\mathbf{M}_{\phi\phi mn} &= \lim_{T \rightarrow \infty} \frac{1}{T-n} \sum_{t=n+1}^T \mathbf{E}\mathbf{y}_{t-m} \mathbf{y}'_{t-n} \\
&= \Psi \int_0^1 \delta(r) \delta'(r) dr \Psi' + \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi_{i+s}',
\end{aligned}$$

for $m = 1, \dots, p$, $n \geq m$. ■

The next two lemmas concern the limits of $\mathbf{E}\mathbf{M}_{\phi\alpha}(t/T)$ and $\mathbf{E}\mathbf{M}_{\alpha\beta}(t/T)$, $\alpha, \beta = \delta, \gamma, c$, as $T \rightarrow \infty$.

Lemma A.5. Consider equation (A.1) and let $\mathbf{M}_{\phi m \alpha T} = (1/T) \sum_{t=1}^T \mathbf{y}_{t-m} \mathbf{g}'_{\phi \alpha}(t/T)$, $\alpha = \delta, \gamma, c$. Then

$$\lim_{T \rightarrow \infty} \mathbf{E} \mathbf{M}_{\phi m \alpha T} = \Psi \int_0^1 \delta(r) \mathbf{g}'_{\phi \alpha}(r) dr,$$

for $m = 1, \dots, p$.

Proof. Write

$$\mathbf{E} \mathbf{y}_{t-m} \mathbf{g}'_{\phi \alpha}(t/T) = \Psi \delta\left(\frac{t-m}{T}\right) \mathbf{g}'_{\phi \alpha}(t/T).$$

Setting $t = [Tr]$,

$$\begin{aligned} & \frac{1}{T-m} \sum_{t=m+1}^T \mathbf{E} \mathbf{y}_{t-m} \mathbf{g}'_{\phi \alpha}(t/T) \\ &= \Psi \left\{ \frac{1}{T-m} \sum_{t=m+1}^T \delta\left(\frac{t-m}{T}\right) \right\} \mathbf{g}'_{\phi \alpha}(t/T) \\ &= \Psi \sum_{t=m+1}^T \int_{t/T}^{(t+1)/T} \delta\left(\frac{[Tr]-m}{T}\right) \mathbf{g}'_{\phi \alpha}([Tr]/T) dr \\ &= \Psi \int_{(m+1)/T}^{(T+1)/T} \delta\left(\frac{[Tr]-m}{T}\right) \mathbf{g}'_{\phi \alpha}([Tr]/T) dr \\ &\rightarrow \Psi \int_0^1 \delta(r) \mathbf{g}'_{\phi \alpha}(r) dr, \end{aligned}$$

for $m = 1, \dots, p$, as $T \rightarrow \infty$. ■

Lemma A.6. Consider equation (A.1) and let $\mathbf{M}_{\alpha \beta T} = (1/T) \sum_{t=1}^T \mathbf{g}_{\alpha}(t/T) \mathbf{g}'_{\beta}(t/T) + \varepsilon_t \text{diag}(g_{\alpha \beta 1}(t/T), \dots, g_{\alpha \beta q}(t/T))$, $\alpha, \beta = \gamma, c$. Then

$$\lim_{T \rightarrow \infty} \mathbf{E} \mathbf{M}_{\alpha \beta T} = \int_0^1 \mathbf{g}_{\alpha}(r) \mathbf{g}'_{\beta}(r) dr, \quad (\text{A.9})$$

where the elements of (A.9) are bounded.

Proof. Since $\mathbf{E} \text{diag}(g_{\alpha \beta 1}(t/T), \dots, g_{\alpha \beta q}(t/T)) = \mathbf{0}$, we have,

$$\mathbf{E} \mathbf{M}_{\alpha \beta T} = (1/T) \sum_{t=1}^T \mathbf{g}_{\alpha}(t/T) \mathbf{g}'_{\beta}(t/T),$$

$\alpha, \beta = \gamma, c$. Set $t = [Tr]$ and write

$$\begin{aligned} \mathbf{EM}_{\alpha\beta T} &= \sum_{t=1}^T \int_{t/T}^{(t+1)/T} \mathbf{g}_{\alpha}\left(\frac{[Tr]}{T}\right) \mathbf{g}'_{\beta}\left(\frac{[Tr]}{T}\right) dr \\ &= \int_{1/T}^{(T+1)/T} \mathbf{g}_{\alpha}\left(\frac{[Tr]}{T}\right) \mathbf{g}'_{\beta}\left(\frac{[Tr]}{T}\right) dr \\ &\rightarrow \int_0^1 \mathbf{g}_{\alpha}(r) \mathbf{g}'_{\beta}(r) dr \end{aligned}$$

for $\alpha, \beta = \delta, \gamma, c$, as $T \rightarrow \infty$. Since $g_{\alpha i}(r)$, $\alpha = \gamma, c$, $i = 1, \dots, q$; is bounded for $r \in (0, 1]$, see Lemma A.1, $\int_0^1 g_{\alpha i}(r) g_{\beta j}(r) dr < \infty$. \blacksquare

Lemma A.7. Consider equation (A.1) and let $\mathbf{M}_{\delta\delta T} = (1/T) \sum_{t=1}^T \mathbf{g}_{\delta}(t/T) \mathbf{g}'_{\delta}(t/T)$ and $\mathbf{M}_{\delta\alpha T} = (1/T) \sum_{t=1}^T \mathbf{g}_{\delta}(t/T) \mathbf{g}'_{\alpha}(t/T) + \varepsilon_t \mathbf{D}_{\delta\alpha}(t/T)$, $\alpha = \gamma, c$. Then

$$\lim_{T \rightarrow \infty} \mathbf{EM}_{\delta\delta T} = \int_0^1 \mathbf{g}_{\delta}(r) \mathbf{g}'_{\delta}(r) dr, \quad (\text{A.10})$$

and

$$\lim_{T \rightarrow \infty} \mathbf{EM}_{\delta\alpha T} = \int_0^1 \mathbf{g}_{\delta}(r) \mathbf{g}'_{\alpha}(r) dr, \quad (\text{A.11})$$

for $\alpha = \gamma, c$, as $T \rightarrow \infty$. The elements of (A.10) and (A.11) are bounded.

Proof. Since $\mathbf{E}\varepsilon_t \mathbf{D}_{\delta\alpha}(t/T) = \mathbf{0}$, $\alpha = \gamma, c$, $\mathbf{EM}_{\delta\alpha T} = (1/T) \sum_{t=1}^T \mathbf{g}_{\delta}(t/T) \mathbf{g}'_{\alpha}(t/T)$, $\alpha = \gamma, c$. Set $t = [Tr]$ and write

$$\begin{aligned} \mathbf{EM}_{\delta\delta T} &= \sum_{t=1}^T \int_{t/T}^{(t+1)/T} \mathbf{g}_{\delta}\left(\frac{[Tr]}{T}\right) \mathbf{g}'_{\delta}\left(\frac{[Tr]}{T}\right) dr = \int_{1/T}^{(T+1)/T} \mathbf{g}_{\delta}\left(\frac{[Tr]}{T}\right) \mathbf{g}'_{\delta}\left(\frac{[Tr]}{T}\right) dr \\ &\rightarrow \int_0^1 \mathbf{g}_{\delta}(r) \mathbf{g}'_{\delta}(r) dr, \end{aligned}$$

as $T \rightarrow \infty$. Since $g_{\delta 0}(r) = 1$, and $g_{\delta i}(r) = g_i(r)$, $i = 1, \dots, q$, are bounded between zero and one, $\int_0^1 g_{\delta i}(r) g_{\delta j}(r) dr < \infty$ for all i, j . Likewise,

$$\lim_{T \rightarrow \infty} \mathbf{EM}_{\delta\alpha T} = \int_0^1 \mathbf{g}_{\delta}(r) \mathbf{g}'_{\alpha}(r) dr, \quad (\text{A.12})$$

for $\alpha = \gamma, c$. The general element $\int_0^1 g_{\delta i}(r) g_{\alpha j}(r) dr$ of (A.12) is bounded because $g_{\alpha j}(r)$ is bounded by Lemma A.1 and the transition function $g_i(r)$ is bounded between zero and one. \blacksquare

We can now state and prove the following result:

Theorem A2. *Assume that the result of Theorem A1 holds, that is, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ as $T \rightarrow \infty$. Assume further that Assumptions A1-A5 are valid. Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}^{-1})$$

where

$$\begin{aligned} \mathbf{J} &= \lim_{T \rightarrow \infty} \mathbf{E} \mathbf{J}_T = (1/\sigma^2) \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbf{E} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\varepsilon_t | \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}'} \{ \ln f(\varepsilon_t | \boldsymbol{\theta}_0) \} \\ &= (1/\sigma^2) \lim_{T \rightarrow \infty} \begin{bmatrix} \mathbf{EM}_{\phi\phi T} & \mathbf{EM}_{\phi\delta T} & \mathbf{EM}_{\phi\gamma T} & \mathbf{EM}_{\phi c T} \\ & \mathbf{M}_{\delta\delta T} & \mathbf{EM}_{\delta\gamma T} & \mathbf{EM}_{\delta c T} \\ & & \mathbf{EM}_{\gamma\gamma T} & \mathbf{EM}_{\gamma c T} \\ & & & \mathbf{EM}_{cc T} \end{bmatrix}. \end{aligned} \quad (\text{A.13})$$

In (A.13),

$$\mathbf{EM}_{\phi\phi T} = \begin{bmatrix} \mathbf{EM}_{\phi\phi 11 T} & \dots & \mathbf{EM}_{\phi\phi 1p T} \\ & \dots & \\ \mathbf{EM}_{\phi\phi p1 T} & \dots & \mathbf{EM}_{\phi\phi pp T} \end{bmatrix},$$

and

$$\mathbf{EM}_{\phi\alpha}(t/T) = [\mathbf{EM}'_{\phi\alpha 1 T} \quad \dots \quad \mathbf{EM}'_{\phi\alpha p T}]',$$

with the same block division as in (A.5) and (A.6), respectively. Then

$$\lim_{T \rightarrow \infty} \mathbf{EM}_{\phi\phi mn T} = \boldsymbol{\Psi} \int_0^1 \delta(r) \delta'(r) dr \boldsymbol{\Psi}' + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{\Sigma} \boldsymbol{\Psi}'_{i+s}$$

where $s = n - m > 0$ and $\boldsymbol{\Psi} = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i$. Furthermore,

$$\lim_{T \rightarrow \infty} \mathbf{EM}_{\phi m \alpha T} = \boldsymbol{\Psi} \int_0^1 \delta(r) \mathbf{g}'_{\phi\alpha}(r) dr$$

for $\alpha = \gamma, c$ and $i = 1, \dots, q$; $\alpha = \delta$ and $i = 0, 1, \dots, q$. Finally,

$$\lim_{T \rightarrow \infty} \mathbf{M}_{\alpha\beta T} = [m_{\alpha\beta, ij}] = \int_0^1 \mathbf{g}_{\alpha}(r) \mathbf{g}'_{\beta}(r) dr$$

for $\alpha, \beta = \gamma, c$ and $i = 1, \dots, q$; $\alpha = \delta$ and $i = 0, 1, \dots, q$.

Proof. We prove Theorem A2 by verifying the conditions of Theorem 3.3 in Newey and McFadden (1994, p. 2146):

Theorem (Newey and McFadden). Suppose that hypotheses of Theorem 2.5 are sat-

isfied and

- (i) $\boldsymbol{\theta}_0$ is an interior point of Θ ,
- (ii) $f(\varepsilon_t|\boldsymbol{\theta})$ is twice continuously differentiable and $f(\varepsilon|\boldsymbol{\theta}) > 0$ in a neighbourhood \mathcal{N} of $\boldsymbol{\theta}_0$,
- (iii) $\int \sup_{\boldsymbol{\theta} \in \mathcal{N}} \|\frac{\partial}{\partial \boldsymbol{\theta}} f(\varepsilon|\boldsymbol{\theta})\| d\varepsilon < \infty$ and $\int \sup_{\boldsymbol{\theta} \in \mathcal{N}} \|\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} f(\varepsilon|\boldsymbol{\theta})\| d\varepsilon < \infty$,
- (iv) $\mathbf{J} = \mathbf{E}\{\ln \frac{\partial}{\partial \boldsymbol{\theta}} f(\varepsilon_t|\boldsymbol{\theta}_0)\}\{\frac{\partial}{\partial \boldsymbol{\theta}'} \ln f(\varepsilon_t|\boldsymbol{\theta}_0)\}$ exists and is nonsingular,
- (v) $\mathbf{E} \sup_{\boldsymbol{\theta} \in \mathcal{N}} \|\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f(\varepsilon|\boldsymbol{\theta})\| < \infty$, where \mathcal{N} is a neighbourhood of $\boldsymbol{\theta}_0$.

Then $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{J}^{-1})$.

Condition (i) follows from A3 and (ii) from A4 and Lemma A.2. Condition (iii) is satisfied. In order to verify (iv), we consider blocks of (A.13). First note that \mathbf{J}_T is nonsingular for $T > 3q + kp + 1$, so \mathbf{J}_T^{-1} exists. Lemma A.4 yields $\lim_{T \rightarrow \infty} \mathbf{E} \mathbf{M}_{\phi\phi}(t/T)$, Lemma A.5 provides the elements of $\lim_{T \rightarrow \infty} \mathbf{E} \mathbf{M}_{\phi\alpha}(t/T)$, $\alpha = \delta, \gamma, c$, Lemma A.6 does the same for $\lim_{T \rightarrow \infty} \mathbf{M}_{\alpha\beta}(t/T)$, $\alpha, \beta = \gamma, c$ and Lemma A.7 for $\lim_{T \rightarrow \infty} \mathbf{M}_{\delta\delta}(t/T)$ and $\lim_{T \rightarrow \infty} \mathbf{M}_{\delta\alpha}(t/T)$, $\alpha = \gamma, c$. Putting these together defines $\mathbf{J} = \lim_{T \rightarrow \infty} \mathbf{J}_T$. Condition (v) is satisfied because the deterministic components in the matrix $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln f(\varepsilon|\boldsymbol{\theta})$ are bounded. Furthermore, the elements of this matrix have finite expectations for all $\boldsymbol{\theta} \in \Theta$. Finally, since matrix inversion is a continuous transformation, it follows from the continuous mapping theorem that when $\mathbf{J}_T \rightarrow \mathbf{J}$, then $\mathbf{J}_T^{-1} \rightarrow \mathbf{J}^{-1}$. This concludes the proof. \blacksquare

References

- Ahlgren, Niklas and Paul Catani**, “Wild Bootstrap Tests for Autocorrelation in Vector Autoregressive Models,” *Statistical Papers*, in press.
- Anderson, Heather M. and Farshid Vahid**, “Testing Multiple Equation Systems for Common Nonlinear Components,” *Journal of Econometrics*, 1998, *84*, 1–36.
- Breusch, Trevor and Farshid Vahid**, “Global Temperature Trends,” Monash Econometrics and Business Statistics, Working Paper 04–11, Monash University, Department of Econometrics and Business Statistics, 2011.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones**, “Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Data Set from 1850,” *Journal of Geophysical Research*, 2006, *111*, D12106.
- Candelon, Bertrand and Helmut Lütkepohl**, “On the Reliability of Chow-type Tests for Parameter Constancy in Multivariate Dynamic Models,” *Economics Letters*, 2001, *73*, 155–160.
- Eklund, Bruno and Timo Teräsvirta**, “Testing Constancy of the Error Covariance Matrix in Vector Models,” *Journal of Econometrics*, 2007, *140*, 753–780.
- Estrada, Francisco, Carlos Gay, and Armando Sánchez**, “A Reply to “Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results” by R. K. Kaufmann et al,” *Climatic Change*, 2010, *101*, 407–414.
- Gay-Garcia, Carlos, Francisco Estrada, and Armando Sánchez**, “Global and Hemispheric Temperatures Revisited,” *Climatic Change*, 2009, *94*, 333–349.
- Gil-Alana, L.**, “Time Trend Estimation with Breaks in Temperature Time Series,” *Climatic Change*, 2008, *89*, 325–337. 10.1007/s10584-008-9407-z.
- Gil-Alana, Luis A.**, “Warming Break Trends and Fractional Integration in the Northern, Southern, and Global Temperature Anomaly Series,” *Journal of Atmospheric and Oceanic Technology*, 2008, *25*, 570–578.
- González, Andrés and Timo Teräsvirta**, “Modelling Autoregressive Processes with a Shifting Mean,” *Studies in Nonlinear Dynamics & Econometrics*, 2008, *12*, No. 1, Article 1. Retrieved from: <http://www.bepress.com/snede/vol12/iss1/art1>.
- González, Andrés, Kirstin Hubrich, and Timo Teräsvirta**, “Forecasting Inflation with Gradual Regime Shifts and Exogenous Information,” CREATES Research Papers 2009–03, School of Economics and Management, University of Aarhus, 2009.

- Goodwin, Barry K., Matthew T. Holt, and Jeffrey P. Prestemon**, “North American Oriented Strand Board Markets, Arbitrage Activity, and Market Price Dynamics: A Smooth Transition Approach,” *American Journal of Agricultural Economics*, 2011, *93*, 993–1014.
- Harvey, David I. and Terence C. Mills**, “Modelling Global Temperature Trends Using Cointegration and Smooth Transitions,” *Statistical Modelling*, 2001, *1*, 143–159.
- and –, “Unit Roots and Double Smooth Transitions,” *Journal of Applied Statistics*, 2002, *29*, 675–683.
- Hendry, David F. and Grayham E. Mizon**, “Exogeneity, Causality, and Co-breaking in Economic Policy Analysis of a Small Econometric Model of Money in the UK,” *Empirical Economics*, 1998, *23*, 267–294.
- and **Michael Massmann**, “Co-Breaking: Recent Advances and a Synopsis of the Literature,” *Journal of Business & Economic Statistics*, 2007, *25*, 33–51.
- Hillebrand, Eric, Marcelo C. Medeiros, and Junuye Xu**, “Asymptotic Theory for Regressions with Smoothly Changing Parameters,” *Journal of Time Series Econometrics*, 2013, *5*, 133–162.
- Ivanov, Martin A. and Stilian N. Evtimov**, “1963: The Break Point of the Northern Hemisphere Temperature Trend During the Twentieth Century,” *International Journal of Climatology*, 2010, *30*, 1738–1746.
- Johansen, Søren**, “The Analysis of Nonstationary Time Series Using Regression, Correlation and Cointegration with an Application to Annual Mean Temperature and Sea Level,” CREATES Research Papers, School of Economics and Management, University of Aarhus, 2010.
- Kaufmann, Robert, Heikki Kauppi, and James Stock**, “Emissions, Concentrations, & Temperature: A Time Series Analysis,” *Climatic Change*, 2006, *77*, 249–278. 10.1007/s10584-006-9062-1.
- , – , and –, “The Relationship Between Radiative Forcing and Temperature: What Do Statistical Analyses of the Instrumental Temperature Record Measure?,” *Climatic Change*, 2006, *77*, 279–289. 10.1007/s10584-006-9063-0.
- , – , and –, “Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results,” *Climatic Change*, 2010, *101*, 395–405. 10.1007/s10584-009-9711-2.
- Kaufmann, Robert K. and David I. Stern**, “Evidence for Human Influence on Climate from Hemispheric Temperature Relations,” *Nature*, 1997, *388*, 39–44.

- **and** –, “Cointegration Analysis of Hemispheric Temperature Relations,” *Journal of Geophysical Research*, 2002, *107*(D2), 4012.
- , **Heikki Kauppi**, and **James H. Stock**, “Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results,” *Climatic Change*, 2010, *101* (3), 395–405.
- , **Kauppi**, **Michael L. Mann**, and **James H. Stock**, “Does Temperature Contain a Stochastic Trend: Linking Statistical Results to Physical Mechanisms,” *Climatic Change*, 2013, *118* (3), 729–743.
- Leybourne, Stephen, Paul Newbold, and Dimitrios Vougas**, “Unit Roots and Smooth Transitions,” *Journal of Time Series Analysis*, 1998, *19*, 83–97.
- Lin, Chien-Fu Jeff and Timo Teräsvirta**, “Testing the Constancy of Regression Parameters Against Continuous Structural Change,” *Journal of Econometrics*, 1994, *62*, 211–228.
- Liu, Hui and Gabriel Rodríguez**, “Human Activities and Global Warming: a Cointegration Analysis,” *Environmental Modelling & Software*, 2005, *20*, 761–773.
- Lütkepohl, Helmut**, *New Introduction to Multiple Time Series Analysis*, Springer-Verlag Berlin Heidelberg, 2005.
- Lütkepohl, Helmut and Markus Krätzig**, *Applied Time Series Econometrics* Cambridge Books, Cambridge: Cambridge University Press, 2004.
- Luukkonen, Ritva, Pentti Saikkonen, and Timo Teräsvirta**, “Testing Linearity Against Smooth Transition Autoregressive Models,” *Biometrika*, 1988, *75* (3), 491–499.
- Mills, Terence C.**, “‘Skinning a Cat’: Alternative Models of Representing Temperature Trends,” *Climatic Change*, 2010, *101*, 415–426.
- Newey, Whitney and Daniel McFadden**, “Large Sample Estimation and Hypothesis Testing,” in R. F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, North Holland, 1994, pp. 2113–2245.
- Perron, Pierre**, “The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis,” *Econometrica*, 1989, *57*, 1361–1401.
- , “Testing for a Unit Root in a Time Series with a Changing Mean,” *Journal of Business & Economic Statistics*, 1990, *8*, 153–162.
- Rao, C. Radhakrishna**, *Linear Statistical Inference and its Applications*, New York: Wiley, 1973.

- Rissanen, Jorma**, “Modeling by Shortest Data Description,” *Automatica*, 1978, 14 (5), 465–471.
- Schwarz, Gideon**, “Estimating the Dimension of a Model,” *The Annals of Statistics*, 1978, 6 (2), 461–464.
- Seidel, Dian J. and John R. Lanzante**, “An Assessment of Three Alternatives to Linear Trends for Characterizing Global Atmospheric Temperature Changes,” *Journal of Geophysical Research*, 2004, 109, D14108.
- Shukur, Ghazi and David Edgerton**, “The Small Sample Properties of the RESET Test as Applied to Systems of Equations,” *Journal of Statistical Computation and Simulation*, 2002, 72, 909–924.
- Teräsvirta, Timo, Dag Tjøstheim, and Clive W. J. Granger**, *Modelling Non-linear Economic Time Series*, Oxford: Oxford University Press, 2010.
- van Dijk, Dick, Birgit Strikholm, and Timo Teräsvirta**, “The Effects of Institutional and Technological Change and Business Cycle Fluctuations on Seasonal Patterns in Quarterly Industrial Production series,” *Econometrics Journal*, 2003, 6, 79–98.
- White, Halbert**, “Approximate Nonlinear Forecasting Methods,” in G. Elliott, C. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 1, Amsterdam: Elsevier, 2006, chapter 9, pp. 459–512.
- Yang, Yukai**, “Modelling Nonlinear Economic Time Series.” PhD dissertation, Aarhus University 2012.

Table 1: Results of Various System Tests for the Bivariate VSM-AR Model of Temperature Data in the Northern and Southern Hemispheres.

Hypothesis	$\ln L$	AIC	BIC	K	LR Test	Rao's F Test
Unrestricted [‡]	425.686	-10.461	-9.737	38	—	—
$H_0 : \gamma_1 = \gamma_5, c_1 = c_5$ Weak Co-shifting	425.285	-10.482	-9.796	36	$\chi^2_{(2)} = 0.802(0.670)$	$F_{(2,143)} = 0.356(0.701)$
$H_0 : \gamma_2 = \gamma_6, c_2 = c_6$ Weak Co-shifting	425.515	-10.485	-9.799	36	$\chi^2_{(2)} = 0.343(0.842)$	$F_{(2,143)} = 0.152(0.859)$
$H_0 : \gamma_1 = \gamma_5, c_1 = c_5,$ $\gamma_2 = \gamma_6, c_2 = c_6$ Weak Co-shifting	422.972	-10.478	-9.830	34	$\chi^2_{(4)} = 4.707(0.319)$ $\chi^2_{(2)} = 4.628(0.099)^*$ $\chi^2_{(2)} = 5.086(0.079)^\diamond$	$F_{(4,287)} = 1.050(0.382)$ $F_{(2,144)} = 2.094(0.127)^*$ $F_{(2,144)} = 2.305(0.103)^\diamond$
$H_0 : \text{No Remaining}$ Mean Shifts $ \gamma_1 = \gamma_5, c_1 = c_5,$ $\gamma_4 = \gamma_7, c_4 = c_7$	424.063 [†]	-10.417 [†]	-9.655 [†]	40 [†]	$\chi^2_{(6)} = 2.183(0.902)$	$F_{(6,285)} = 0.321(0.926)$

[‡] Results for the model with no co-shifting restrictions imposed.

* Test results relative to the unrestricted model where $\gamma_1 = \gamma_5, c_1 = c_5$ is maintained.

◊ Test results relative to the unrestricted model where $\gamma_2 = \gamma_6, c_2 = c_6$ is maintained.

† Results are for the model estimated under the alternative.

Note: There are 162 usable observations. The unrestricted model contains 38 free parameters. $\ln L$ denotes the maximized likelihood function value. AIC and BIC denote, respectively, the system Akaike information criterion and the Rissanen-Schwarz Bayesian information criterion. The test statistic associated with χ^2_j is for a likelihood ratio test involving j parameter restrictions. Rao's F -test statistics are approximate system LM tests as described by Teräsvirta, Tjøstheim and Granger (2010). Values in parentheses are asymptotic p -values. The test for no remaining mean shifts is conducted by using a third-order Taylor approximation.

Table 2: System Estimates for the Hemispheric Temperature VSM–AR with Co–Shifting Restrictions Imposed.

Northern Hemisphere:[†]

$$\begin{aligned}
 y_t^n &= -\frac{0.448}{(0.016)} + \frac{2.030}{(0.023)}(1 + \exp\{-\frac{0.682[t/T - 0.950]}{(-)} / 0.290\})^{-1} \\
 &\quad + \frac{0.069}{(0.021)}(1 + \exp\{-\frac{25.423[t/T - 0.452]}{(3.057)} / 0.290\})^{-1} \\
 &\quad - \frac{0.237}{(0.001)}(1 + \exp\{-\frac{9.528[t/T - 0.267]}{(0.760)} / 0.290\})^{-1} \\
 &\quad - \frac{0.174}{(0.015)}(1 + \exp\{-\frac{50[t/T - 0.589]}{(-)} / 0.290\})^{-1} \\
 &\quad + \frac{0.646}{(0.069)} y_{t-1}^n - \frac{0.580}{(0.053)} y_{t-2}^n + \frac{0.264}{(0.104)} y_{t-3}^n - \frac{0.136}{(0.088)} y_{t-1}^s + \frac{0.372}{(0.079)} y_{t-2}^s - \frac{0.173}{(0.079)} y_{t-3}^s + \widehat{\varepsilon}_t^n \\
 R^2 &= 0.890, \quad \widehat{\sigma}^n = 0.091, \quad Sk = 0.100, \quad Ek = 0.191, \quad LJB = 0.518(0.771)
 \end{aligned}$$

Southern Hemisphere:[†]

$$\begin{aligned}
 y_t^s &= -\frac{0.029}{(0.030)} - \frac{0.866}{(0.062)}(1 + \exp\{-\frac{0.682[t/T - 0.950]}{(0.078)} / 0.290\})^{-1} \\
 &\quad + \frac{0.273}{(0.031)}(1 + \exp\{-\frac{25.423[t/T - 0.452]}{(3.057)} / 0.290\})^{-1} \\
 &\quad + \frac{0.775}{(0.200)}(1 + \exp\{-\frac{3.769[t/T - 0.908]}{(0.662)} / 0.290\})^{-1} \\
 &\quad + \frac{0.652}{(0.090)} y_{t-1}^n - \frac{0.630}{(0.074)} y_{t-2}^n + \frac{0.356}{(0.144)} y_{t-3}^n - \frac{0.052}{(0.114)} y_{t-1}^s + \frac{0.433}{(0.103)} y_{t-2}^s - \frac{0.288}{(0.109)} y_{t-3}^s + \widehat{\varepsilon}_t^s \\
 R^2 &= 0.853, \quad \widehat{\sigma}^s = 0.115, \quad Sk = -0.243, \quad Ek = 0.508, \quad LJB = 3.346(0.188) \\
 \widehat{\rho}_{ns} &= 0.912
 \end{aligned}$$

System Statistics:

$$\begin{aligned}
 \ln L &= 422.972, \quad AIC = -10.478, \quad BIC = -9.830, \quad LM_{\Omega_t} = 2.096 \times 10^{-4}(0.999), \quad Sk_s = 1.840(0.398), \\
 Ek_s &= 1.553(0.460), \quad LJB_s = 3.393(0.494), \quad LM_{VAR(4)} = 7.39 \times 10^{-3}(0.999), \quad LM_{VAR(6)} = 8.06 \times 10^{-3}(0.999), \\
 LM_{VAR(8)} &= 6.33 \times 10^{-3}(0.999), \quad LM_{VAR(10)} = 7.53 \times 10^{-3}(0.999), \quad LM_{VAR(12)} = 6.80 \times 10^{-3}(0.999)
 \end{aligned}$$

[†] See footnote 6.

Note: $\widehat{\rho}_{ns}$ is the estimated correlation between the residuals. Sk denotes skewness and Ek excess kurtosis. LJB is the Lomnicki–Jarque–Bera test of normality of the residuals. These same statistics with a subscripted s are for the system, as described by (Lütkepohl and Krätzig, 2004, pp. 129–130). AIC and BIC denote, respectively, the system Akaike information criterion and the Rissanen–Schwarz Bayesian information criterion. LM_{Ω_t} denotes the system LM test of Eklund and Teräsvirta (2007) for a time-varying covariance matrix. $LM_{VAR(j)}$ denote system LM F -tests, based on Rao’s F , for remaining vector autocorrelation at lags $j = 4, 6, 8, 10, 12$. Values in parentheses beside test statistics are p -values.

Table 3: Transitions and Shift Dates for the VSM-AR Model with Co-Shifting.

i	Northern Hemisphere				Southern Hemisphere					
	$\hat{\gamma}$	\hat{c}	10%	Centre	90%	$\hat{\gamma}$	\hat{c}	10%	Centre	90%
1	0.682	0.950	1938	2006	2073	0.682	0.950	1938	2006	2073
2	25.423	0.452	1922	1924	1925	22.423	0.452	1922	1924	1925
3	9.528	0.267	1888	1894	1898	3.769	0.909	1987	1999	2011
4	50.000	0.589	1945	1946	1947					

Note: Columns titled 10% (90%) denote the dates for which the relevant logistic function is associated with a value of 0.1 (0.9). Likewise, columns headed Centre denote dates for which $t^* = \hat{c}$ for the respective shift function.

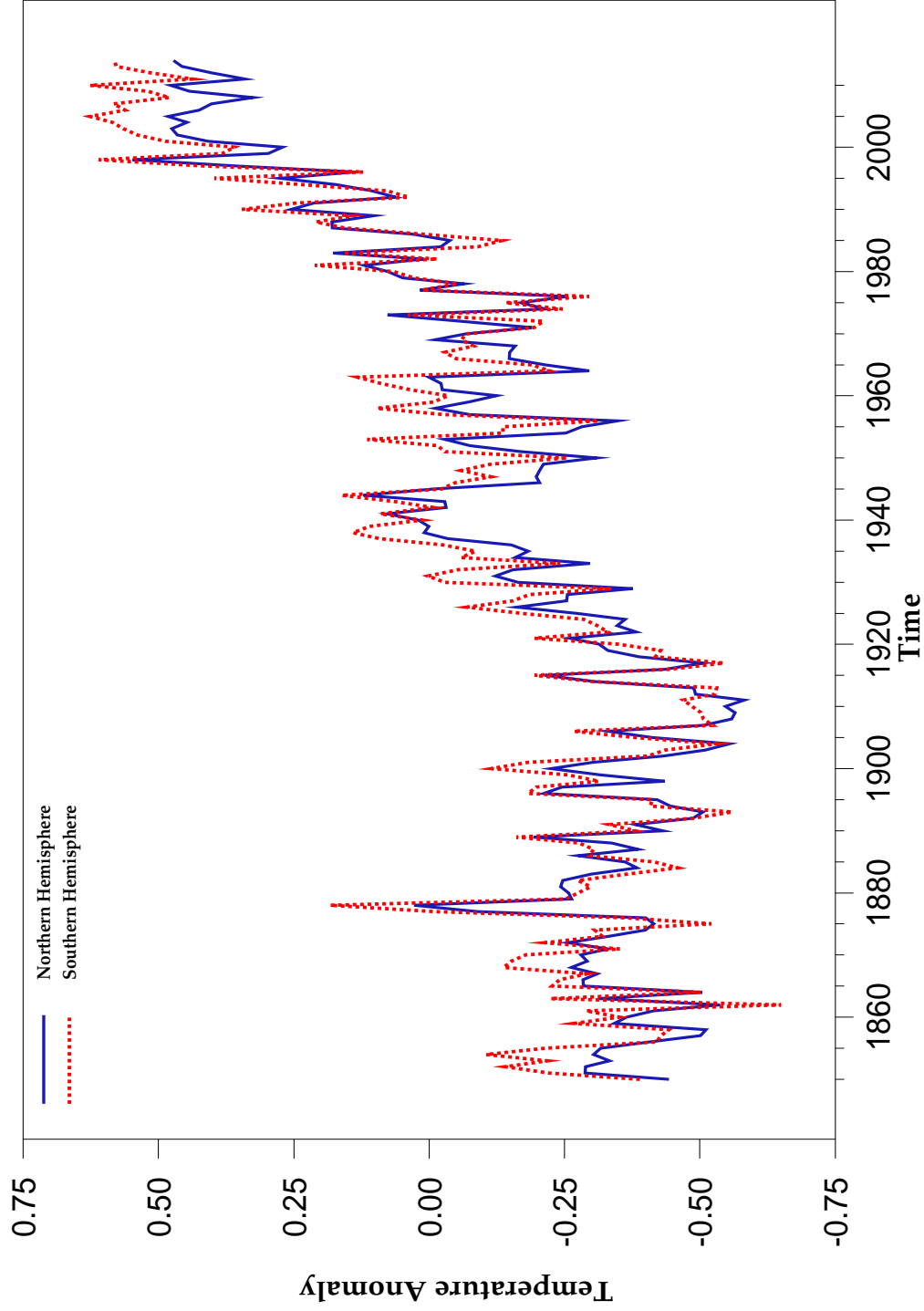


Figure 1: Temperature Anomalies in the Northern and Southern Hemisphere, 1850–2014.

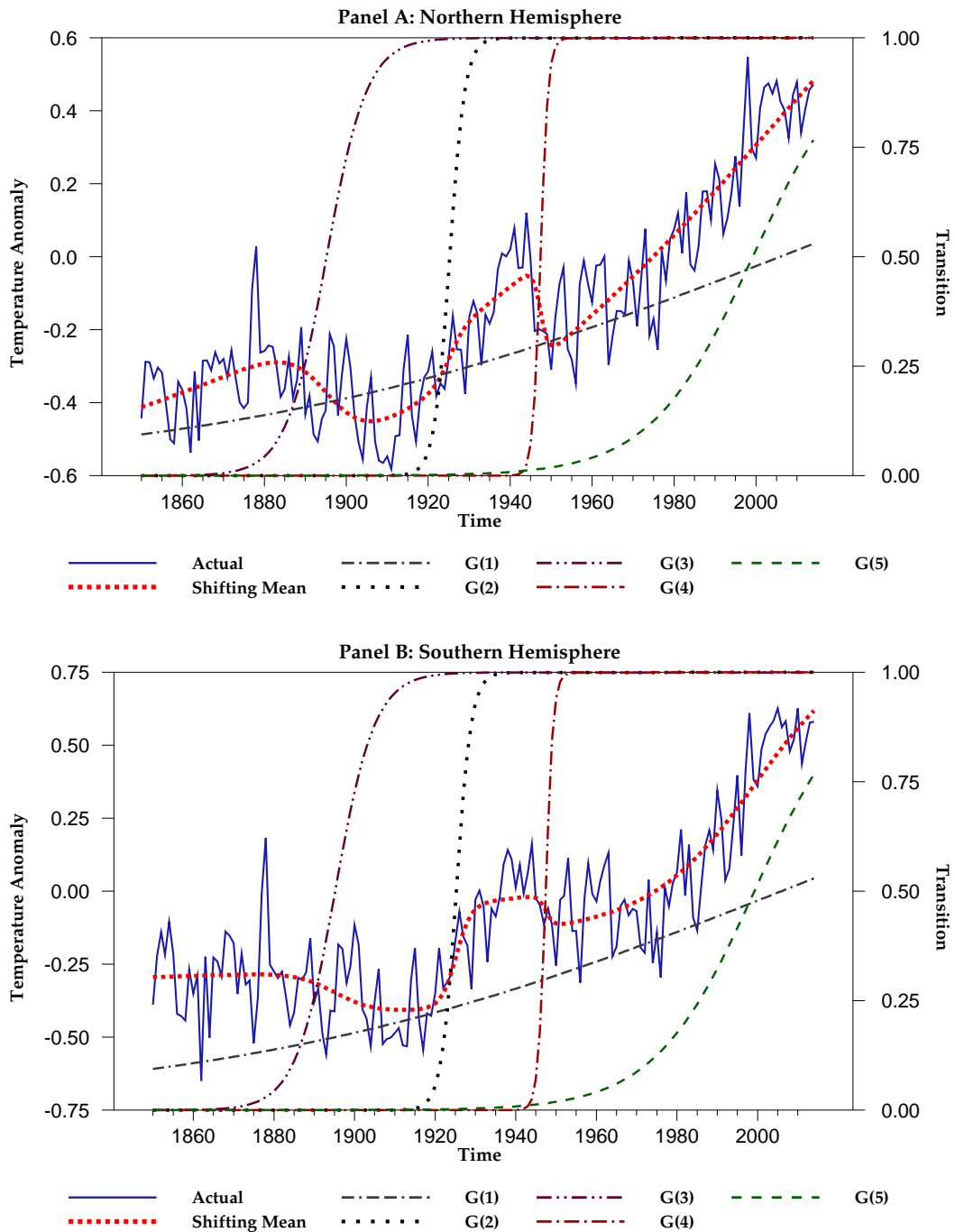


Figure 2: Bivariate VAR Results with Co-Trending Restrictions for Temperature Anomalies for the Northern (Panel A) and Southern (Panel B) Hemispheres, 1850–2014. The dashed line indicates the shifting mean and the dash-dot lines indicate the estimated transition functions.

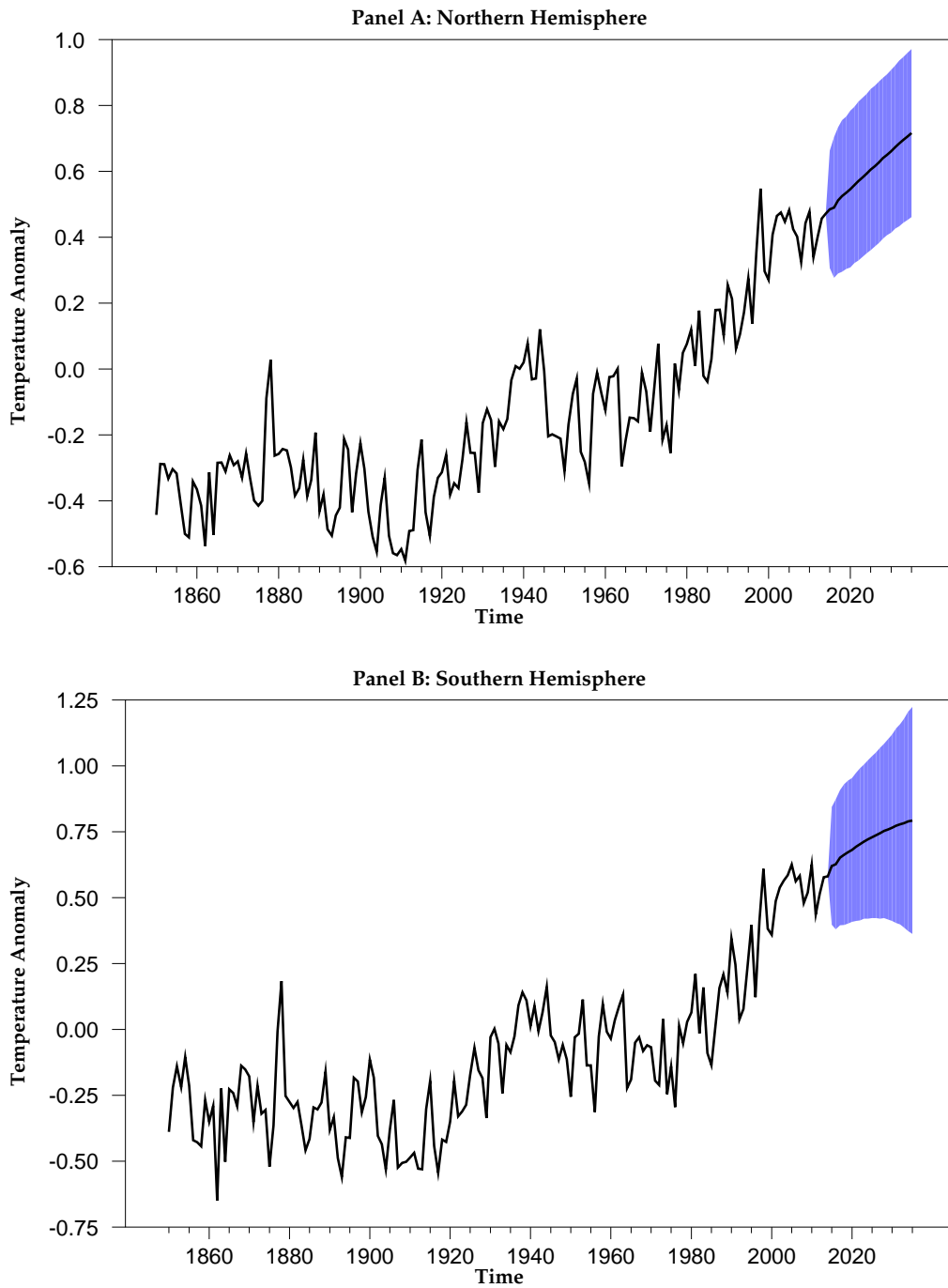


Figure 3: Temperature Anomalies in the Northern (Panel A) and Southern (Panel B) Hemispheres, Actual, 1850–2014, and Simulated, 2015–2035. Shaded areas denote 95–percent confidence intervals.

Research Papers 2016



- 2016-21: Mikkel Bennedsen: Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data
- 2016-22: Søren Johansen and Morten Ørregaard Nielsen: The cointegrated vector autoregressive model with general deterministic terms
- 2016-23: Yunus Emre Ergemen and Carlos Vladimir Rodríguez-Caballero: A Dynamic Multi-Level Factor Model with Long-Range Dependence
- 2016-24: Shin Kanaya: Convergence rates of sums of α -mixing triangular arrays: with an application to non-parametric drift function
- 2016-25: Gustavo Fruet Dias, Marcelo Fernandes and Cristina M. Scherrer: Improving on daily measures of price discovery
- 2016-26: Martin M. Andreasen, Tom Engsted, Stig V. Møller and Magnus Sander: Bond Market Asymmetries across Recessions and Expansions: New Evidence on Risk Premia
- 2016-27: Kim Christensen, Ulrich Hounyo and Mark Podolskij: Testing for heteroscedasticity in jumpy and noisy high-frequency data: A resampling approach
- 2016-28: Kim Christensen, Roel Oomen and Roberto Renò: The Drift Burst Hypothesis
- 2016-29: Hossein Asgharian, Charlotte Christiansen, Rangan Gupta and Ai Jun Hou: Effects of Economic Policy Uncertainty Shocks on the Long-Run US-UK Stock Market Correlation
- 2016-30: Morten Ørregaard Nielsen and Sergei S. Shibaev: Forecasting daily political opinion polls using the fractionally cointegrated VAR model
- 2016-31: Carlos Vladimir Rodríguez-Caballero: Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure
- 2016-32: Lasse Bork, Stig V. Møller and Thomas Q. Pedersen: A New Index of Housing Sentiment
- 2016-33: Joachim Lebovits and Mark Podolskij: Estimation of the global regularity of a multifractional Brownian motion
- 2017-01: Nektarios Aslanidis, Charlotte Christiansen and Andrea Cipollini: Predicting Bond Betas using Macro-Finance Variables
- 2017-02: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form
- 2017-03: Peter Exterkate and Oskar Knapik: A regime-switching stochastic volatility model for forecasting electricity prices
- 2017-04: Timo Teräsvirta: Sir Clive Granger's contributions to nonlinear time series and econometrics
- 2017-05: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperatures and Co-Shifting: A Vector Shifting-Mean Autoregressive Analysis