This is the accepted manuscript (post-print version) of the article.
The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication
Please cite the final published version:


Publication metadata

Title: Global hemispheric temperatures and co-shifting: A vector shifting-mean autoregressive analysis.
Author(s): Holt, Matthew T. ; Terasvirta, Timo.
Journal: Journal of Econometrics
DOI/Link: 10.1016/j.jeconom.2019.05.011
Document version: Accepted manuscript (post-print)
Document license: CC-BY-NC-ND

General Rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

This cover sheet template is made available by AU Library
Version 2.1, September 2020
Global Hemispheric Temperatures and Co–Shifting: A Vector Shifting–Mean Autoregressive Analysis∗

Matthew T. Holt†
Department of Agricultural and Applied Economics
Virginia Tech

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

January 23, 2019

Abstract
This paper examines local changes in annual temperature data for the northern and southern hemispheres (1850–2017) 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.

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 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 and three anonymous referees for thorough reading of the paper and many suggestions. Errors and shortcomings in this work remain our own responsibility. Authors names appear in alphabetical order–senior authorship is not assigned.

†Professor and Head, Department of Agricultural and Applied Economics, Virginia Tech, 250 Drillfield Drive, 208 Hutcheson Hall, Blacksburg, VA 24061, USA. Telephone: 540-231-6301. Fax: 540-231-7417 . E-mail: mattholt@vt.edu.

‡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
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 permeant 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 relatively short 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 a longer perspective, however, nonstationarity may not hold. Davidson, Stephenson and Turasieb (2016), analysing a long paleoclimate temperature series, find it to be

---

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.
weakly stationary and thus bounded in probability.

In this paper we model these temperature series differently. Instead of global common features, we are interested in local similarities between them. The basic idea is to assume nonstationarity in 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. The purpose of the shifts is to describe and underline conspicuous movements of the series in-sample; extrapolating them into the future may not be very informative. This is of course true for deterministic trends, linear and nonlinear, as well.

We follow many of the aforementioned studies in that we only consider annual hemispheric temperature series and present a new way of looking at them. While a more detailed analysis might seek to include a suite of radiative forcing variables including those such as well-mixed greenhouse gases, total radiative forcing, and radiative forcing from stratospheric aerosols, as in, for example, Estrada, Martins and Perron (2017), here we simply seek to illustrate our methodology. As such, and not unlike a number of related studies in the literature, we make no attempt to assign causality, either anthropogenic or otherwise, to the hemispheric temperature shifts (co-shifts) that we identify and explore—doing so remains as an important topic for future research.

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. Our model implicitly assumes the series to be bounded but, as already suggested, the model is not designed for long-range out-of-sample forecasting.

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, while proofs and additional results appear in the Appendix.
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:

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

(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:

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

(2)

where

\[ g(t/T; \gamma_i, c_i) = \left(1 + \exp\{-\gamma_i (t/T - c_i)\}\right)^{-1}, \quad \gamma_i > 0, \]  

(3)

and \( T \) is the number of observations. It follows that the value of \( \delta(t/T) \) changes (possibly nonmonotonically) from \( \delta_0 \) 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 \( \gamma_j \) and \( c_j \). Rescaled time in the argument of \( g(t/T; \gamma_j, c_j) \) leaves the relative location of transitions intact as \( T \to \infty \). 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

\[ \mathbb{E}y_t = \phi^{-1}(1)\delta(t/T) + O(1/T), \]

where the asymptotically vanishing term is due to the fact that \( \delta(t/T) \) is a nonlinear function, see, for example, He, Kang, Teräsvirta and Zhang (2018) for discussion. 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 \( \gamma_i \) imply that the underlying shift, whose mid-point occurs at time \( c_i \), becomes rather abrupt. Alternatively, for small values of \( \gamma_i \) (and assuming for the moment that \( q = 1 \)), the shift from \( \delta_0 \) 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

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

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

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

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

\[ \delta(t/T) = \delta_0 + \sum_{i=1}^{q} \mathbf{G}_i(t/T) \delta_i, \]  

where \( \delta_i = (\delta_{i1}, ..., \delta_{ik})' \) is a \( k \times 1 \) parameter vector, \( i = 0, 1, ..., q \). Furthermore, \( \mathbf{G}_i(t/T) \) are diagonal matrices defined as:

\[ \mathbf{G}_i(t/T) = \text{diag}(g(t/T; \gamma_{i1}, c_{i1}), ..., g(t/T; \gamma_{ki}, c_{ki})), \]  

where, as before,

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

for \( i = 1, ..., q \) and \( j = 1, ..., k \). For some of the equations in (6) it is possible that \( q_j < q = \max_{j=1,...,k} q_j \), which is not explicit from the notation. Equations (4)–(8) define the Vector Shifting–Mean Autoregressive Model.

We make the following assumptions:

**Assumption A1.** The roots of \( |\mathbf{I} - \sum_{j=1}^{p} \Phi_j z^j| = 0 \) lie outside the unit circle.

**Assumption A2.** In the transition function \( g(t/T; \gamma_{ji}, c_{ji}) \), \( \gamma_{ji} > 0, \; i = 1, ..., q; \; c_{ji} < ... < c_{jq_j} \). This implies that in the \( j \)th equation \( g(t/T; \gamma_{ji}, c_{ji}) \neq g(t/T; \gamma_{ji+1}, c_{ji+1}) \). In addition, \( \delta_{ji} \neq 0 \) for \( i = 1, ..., q_j \).

If this assumption is relaxed such that \( c_{ji} = c_{j,i+1} \) for some \( i \), then \( g(t/T; \gamma_{ji}, c_{ji}) \neq g(t/T; \gamma_{ji+1}, c_{j,i+1}) \) requires \( \gamma_{ji} \neq \gamma_{j,i+1} \).
Assumption A3. Parameter space $\Theta^*$ is compact, the true parameter $\theta_0^*$ is an interior point of $\Theta^*$.

Assumption A4. The density $f(\varepsilon_t|\theta^*)$ is positive (bounded away from zero) for all $\theta^* \in \Theta^*$.

Assumption A5. The errors $\varepsilon_t \sim (iid)\mathcal{N}(0, \Sigma)$, where $\Sigma$ is positive definite.

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

$$Ey_t = \Psi \delta(t/T) + O(1/T)1_k,$$

where $1_k$ is a $k \times 1$ vector of unit values. According to A3, the elements of $\delta(t/T)$ are bounded. It follows that the shifting mean (9) is bounded as well, and $y_t$ 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). It should be noted, however, that all aforementioned models, including the VSM-AR model, are local approximations to the unknown data-generating process and, as such, are not necessarily suited for long-term forecasting.

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 $\gamma_{ij} = 0$, $\delta_{ij}$ 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

\[
S_j = \left\{ (1 + \exp\left\{ -\gamma_{ji}(t/T - c_{ji}) \right\})^{-1}, \ i = 1, \ldots, M \right\},
\]

where \( M \) is large. The set is obtained by constructing a grid of points \((\gamma_{ji}, c_{ji})\) 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 \( \alpha_0 \), 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 \( y_t \), 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 \( \Phi_j = [\phi_{j,1}', \ldots, \phi_{j,k}]' \), where \( \phi_{jm} \) is the \( m \)th row of \( \Phi_j \). Setting \( y_{t-m} = (y_{1,t-m}', \ldots, y_{k,t-m}')' \), the \( i \)th equation
of (4) can be written as

\[ y_{jt} = \delta_j(t/T) + \sum_{m=1}^{p_j} \phi'_{jm} y_{t-m} + \xi_{jt}, \quad (11) \]

\( j = 1, \ldots, k. \) Let \( \theta_0 \) be the true vector of parameters in (11). Asymptotic properties of the maximum likelihood estimator \( \hat{\theta} \) of \( \theta_0 \) 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 \( \gamma_{ji} = \exp(\eta_{ji}) \) and estimate \( \eta_{ji}. \)

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 a common feature. To illustrate this in our framework, consider a \( k \)-dimensional VSM-AR model with two shifts in the intercept:

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

where

\[ \mathbf{G}_i(t/T) = \text{diag}(g(t/T; \gamma_{i1}, c_{i1}), \ldots, g(t/T; \gamma_{ik}, c_{ki})), \]

and \( \mathbf{\delta}_i = (\delta_{i1}, \ldots, \delta_{ik})', \ i = 1, 2. \) The shifting intercept is a common feature if there is a \( k \times m \) matrix \( \mathbf{A} \) such that \( \mathbf{A}' \mathbf{y}_t \) is linear, that is, \( \mathbf{A}'\mathbf{G}_1(t/T)\mathbf{\delta}_1 = \mathbf{A}'\mathbf{G}_2(t/T)\mathbf{\delta}_2 = \mathbf{0}. \)

A necessary condition for this to happen is that \( g(t/T; \gamma_{ji}, c_{ji}) = g(t/T; \gamma_i, c_i), \ j = 2 \)

\(^{2}\)This parametrization for \( \gamma_{ji} \) automatically ensures that \( \gamma_{ji} > 0 \) holds, which implies that the identification condition \( \gamma_{ji} > 0 \) holds. Moreover, it facilitates a grid search wherein equal spacings between (large) \( \gamma_{ji} \) values are not optimal.
1, ..., k, or, in matrix notation:

\[ G_i(t/T) = g(t/T; \gamma_i, c_i)I_k, \ i = 1, 2, \]  

(13)

where \( I_k \) is a \((k \times k)\) identity matrix. This implies that the slope and location parameters of the two transition functions across the \( k \) equations are the same. If the necessary condition (13) is completed by the condition \( \delta_2 = \pm \lambda \delta_1, \lambda \neq 0 \), there exists a \( k \times m, m < k \), matrix \( A \neq 0 \) with rank \( m \) that eliminates the shift. The resulting model for \( x_t = A'y_t \) is linear. This is called (strong) co-shifting.

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

\[ G_1(t/T) = g(t/T; \gamma_1, c_1)I_k. \]

and that there exists a \( k \times m, m < k \), matrix \( A \) with rank \( m \) such that \( A'\delta_1 = 0 \). Left-multiplication of (12) by \( A' \) eliminates the first but not necessarily the second shift:

\[ A'y_t = A'\delta_0 + A'G_2(t/T)\delta_2 + \sum_{j=1}^{p} A'\Phi_jy_{t-j} + A'\epsilon_t. \]

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

\[ G_2(t/T) = g(t/T; \gamma_{12}, c_{12})I_k \]

there may exist another \( k \times n, n < k \), matrix \( B \) of rank \( n \) such that \( B'\delta_2 = 0 \) and

\[ B'y_t = B'\delta_0 + B'G_1(t/T)\delta_1 + \sum_{j=1}^{p} B'\Phi_jy_{t-j} + B'\epsilon_t. \]

As already discussed, a single matrix \( A \) can eliminate both shifts only if \( \delta_2 = \pm \lambda \delta_1 \). 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 \( \delta_1 \) and \( \delta_2 \) are not restricted to mimic this feature of cointegrated random variables.
The definition of strong co-shifting resembles the definition of contemporaneous mean co-breaking, as stated in Hendry and Mizon (1998) and Hendry and Massmann (2007). These authors assumed that there is a (typically linear) ‘baseline’ vector autoregressive process for the $k \times 1$ vector $y_t$ such that $E y_t = \rho_0$. They considered a deterministic location shift in the unconditional mean of a vector autoregressive process at time $t$: $E (y_t - \rho_0) = \tau_t \neq 0$ with $\tau_t \neq \tau_{t-1}$. This defines a structural break at $t$. If there is a $k \times m$, $m < k$, matrix $A$ with rank $m$ that cancels the break, $A' \tau_t = 0$, then $A$ is contemporaneous co-breaking for the sequence $y_t$ of order $m$. There may be several breaks represented by $\tau_{t_1}, \ldots, \tau_{t_r}$ but these vectors have to be linearly dependent for $A$ to satisfy the definition of contemporaneous co-breaking. Even here, co-breaking does not mean that the breaks hold the series together.

Strong co-shifting gets closer to co-breaking when (13) holds and $\gamma_i \to \infty$. Then smooth shifts become contemporaneous breaks as the location parameter $c_i$ has the same value for all $k$ equations. Linear dependence of $\delta_1$ and $\delta_2$ in (12) is then required for contemporaneous co-breaking. However, co-breaking is more general than co-shifting in the sense that the location shift may not be a deterministic break. It can also for example be a trend break or a shift or shifts in a trend. Even a contemporaneous outlier in the sense that $\tau_t = 0$ except for some $t = t_0$ fits the definition if the aforementioned matrix $A$ cancels it.

Hendry and Massmann (2007) extended co-breaking into cointegration, whereas in our case the ‘baseline’ model (a model without shifts) is a stationary vector autoregression. However, an extension similar to the one in Hendry and Massmann (2007) might be possible. The model with smooth trends developed and analysed by Ripatti and Saikkonen (2001) would provide a proper starting-point for such an extension.

4.4 Testing Co-shifting Restrictions

Co-shifting may also 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

$$H_0: (\gamma_{i_1}, c_{i_1}) = \ldots = (\gamma_{k_i}, c_{k_i}), \ i = 1, \ldots, q,$$

and $\delta_i = \lambda_i \delta_1$, $\lambda_i \neq 0$, $i = 2, \ldots, q$. 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.
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:

\[
LR = T \left[ \ln |\Omega| - \ln |\hat{\Omega}| \right],
\]

where \( LR \) has an asymptotic \( \chi^2_{k(3q-1)} \) distribution under the null hypothesis. This requires \( \hat{\theta} \) 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 \( \hat{\Omega} \) 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 \( \chi^2_{k(3q-1)} \) 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. Even so, this assumption seems to be satisfied in our application; see Section 6.

### 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 \( |I - \sum_{j=1}^{p} \Phi_j z^j| = 0 \) 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 in the sense that the roots of the aforementioned lag polynomial lie far from the unit circle, so we do not require such a test here.
5 Empirical Application: Discussion and Data

The example we focus on in the empirical analysis involves shifting (co–shifting) in hemispheric temperature data. Importantly, the present analysis is considered to be illustrative regarding a class of models frequently examined in the recent time–series–related literature on long–term hemispheric climate trends. As already mentioned in the Introduction, exogenous variables such as greenhouse gases, total radiative forcing, and radiative forcing from stratospheric aerosols, are not included in our analysis. The purpose here is simply to illustrate our methodology and thus concentrate on parametric descriptions of similarities between the two temperature series.

The data used in the empirical analysis are annual average hemispheric temperature series from 1850–2017, and are described in detail in Morice, Kennedy, Rayner and Jones (2012). As mentioned previously, versions of these time series have been used by various authors in to empirically investigate features of global warming, including Gay-García, 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)

\(^3\)In obtaining ML estimates we follow established practice (see, e.g., van Dijk, Strikholm and Teräsvirta, 2003) and constrain values of \(\gamma_i\) 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.
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 confirms that these restrictions should be maintained.

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 three mean shifts in the northern equation; with two mean shifts in the southern equation; and with lagged temperatures from the north excluded from the equation for the south, is a reasonable representation of the hemispheric temperature data. Estimation results for the unrestricted VSM–AR are reported in Appendix B.

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 intercept shift in common. The possibilities include a long, relatively slow shift that started in the second half of the 20th century. The null hypothesis to be tested is that the logistic transitions (but not their coefficients) are identical. The likelihood ratio test statistic defined in (15) thus has two degrees of freedom as each transition function contains two parameters that are, under the null, restricted to be identical.

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 \( \chi^2 \)-test and Rao’s \( F \) yield similar \( p \)-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. Finally, the implied shifting means for both the northern and southern temperature series are depicted by the dashed red lines in 2.\(^4\) The estimation results have interesting interpretations. The coefficients of the common slow transition with \( c = 0.889 \), starting in the 1940s and centered on 1998, have the same sign in the northern and southern hemisphere equations, but with the magnitude of the intercept shift being over 3.5 times larger in the northern equation. The positive

\(^{4}\)The values for the shifting means recorded in (2) are constructed by using (9) as applied to the estimated VSM–AR model, but where the asymptotically vanishing term is ignored.
sign associated with this intercept shift in both hemisphere equations clearly implies that the shift describes the long positive trend-like movement in the temperature series starting in the second half of the 20th century and visible in Figure 2. Even so, the timing around the beginning of the cumulative upward trend is not equal in both hemispheres. Specifically, and as also illustrated in Figure 2, the gradual upward trend in temperatures in the southern hemisphere begins in the 1960s. The corresponding shift in the north does not effectively start, however, until the 1970s. This delay in the northern hemisphere is accounted for by the third (independent) intercept shift in the northern equation, which not coincidentally is centered on 1968 and, in turn, allows the shifting mean in the northern equation to take a temporary downturn in the 1960s, which is also consistent with the data patterns observed in Figure 2.

Of the remaining transitions, the second one in the northern hemisphere equation is there to account for the observed upward shift in temperatures that occurred in the 1920s. As suggested in Figure 2, this shift in the north is reinforced by the second and somewhat more gradual shift in the southern equation, occurring approximately between 1916 and 1942. See Table 3. That is, as seen from (9) all transitions, common or not, will contribute in this case to the northern equation through the lag structure (e.g., lagged values of southern temperatures help predict northern temperatures). Moreover, the cross-effects of the second shift in the south on the shifting mean in the north is not large but neither is it trivial; the largest root of the VAR component is 0.603. As already noted, the third shift in the northern equation accounts for a temporary downturn in temperatures in this hemisphere in the 1960s. Taken together, the results for shifting means indicate that while the two series undoubtedly move together, they do display individual local tendencies.

How might we connect our results to factors affecting the climate? Based on an abundance of evidence, the steadily increasing deterministic component in Figure 2, common to both equations, most likely represents anthropogenic effects, mainly greenhouse gases, on global temperatures. See, for example, Tett, Stott, Allen, Ingram and Mitchell (1999). Meehl, Hu, Arblaster, Fasullo and Trenberth (2013) argued that these effects are accentuated by internally generated factors, and they emphasized the role of the interdecadal Pacific oscillation of sea surface temperatures. More generally, researchers have found, see, for example, Schlesinger and Ramankutty (1994), a 65–70 year cyclical component in global temperatures for the last 150 years or so. By using principal components and spectral techniques, Barcikowska, Knutson and Zhang (2017) find a strong 65–year cycle in these temperatures that apparently links, moreover, to the reconstructed Atlantic Multidecadal Oscillation (AMO) index. It is known that the AMO affects temperatures not only in the Northern but also in the Southern Hemisphere. The two positive shifts in Figure 2 and Table 3, centered around 1923 in the Northern Hemisphere and around 1929 in the Southern Hemisphere, may be associated with an upswing in this cycle and the AMO. The re-
remaining shift in the Northern Hemisphere equation, centered at 1968, occurs rapidly. This shift does not completely fit with conclusions reported in Barcikowska, Knutson and Zhang (2017), whose 65-year cycle suggests slow cooling from the 1940s through the 1970s. Even so, the negative sign associated with the shift indicates a downward temperature pattern, albeit at a faster rate than the one produced by the constructed cycle. As already mentioned, in the Southern Hemisphere series the corresponding cooling is much less dramatic than in the north, and no shift is needed to describe this phase. The shifting mean does slow down, but as Figure 2 shows, its first derivative remains positive during this period.

Determining causes for the fluctuations in the AMO index are beyond the scope of this study. However, Knudsen, Jacobsen, Seidenkrantz and Olsen (2014) pointed out two possible drivers evident from the 1770s: solar radiation and volcanic activity. Both factors seemingly have a strong, positive correlation with the AMO; in fact, they lead the AMO by about five years. Even so, because we have modelled the two hemispheric time series without including any forcing variables, we can say nothing concrete regarding these possible relationships.

7 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 example involving hemispheric temperature data, 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 three logistic function components are adequate to characterize the shifts in the mean of the northern series whilst only two are required for the southern series. After estimating the parameters of the model by maximum likelihood, subsequent testing reveal that one logistic function component is common to both equations. Our analysis is limited to studying similarities and the dynamic behaviour of the two hemispheric temperature series. Considering the impact of exogenous variables on the system is left for future work.
Appendix A

The Model

The \( j \)th equation of the VSM–AR model (4)–(8) has the following form:

\[
y_{jt} = \delta_j(t/T) + \sum_{i=1}^{p} \phi'_{ji} y_{t-i} + \varepsilon_{jt}, \tag{A.1}
\]

where \( \phi'_{ji} \) is the \( j \)th row of \( \Phi_i \), and the \( j \)th element of the intercept vector equals

\[
\delta_j(t/T) = \delta_{j0} + \sum_{i=1}^{q} \delta_{ji} g(t/T; \gamma_{ji}, c_{ji}).
\]

In what follows we omit the subscript \( j \).

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

\[
L_T(\theta, \varepsilon) = \sum_{t=1}^{T} \ell(\varepsilon_t; \theta), \tag{A.2}
\]

where

\[
\ell(\varepsilon_t; \theta) = k - (1/2) \ln \sigma^2 - \frac{\varepsilon_t^2}{2\sigma^2}. \tag{A.3}
\]

In (A.3), \( \sigma^2 \) is the \( i \)th diagonal element of \( \Sigma \). The parameter vector \( \theta = (\phi'_1, ..., \phi'_p, \delta', \gamma', c')' \in \Theta \subset \mathbb{R}^{3q+p+1} \), where \( \phi_i = (\phi_{1i}, ..., \phi_{ki})' = \text{vec}(\Phi'_i) \) is a \( k \times 1 \) vector, \( \delta = (\delta_0, \delta_1, ..., \delta_q)' \) is a \( (q+1) \times 1 \) vector, and \( \gamma = (\gamma_1, ..., \gamma_q)' \) and \( c = (c_1, ..., c_q)' \) are \( q \times 1 \) vectors. Let \( \theta_0 = (\phi_{10}', ..., \phi_{p0}', \delta_0', \gamma_0', c_0')' \) be the corresponding true parameter vector.

In order to consider the maximum likelihood estimator \( \hat{\theta} \) of the parameter vector \( \theta_0 \), 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 \( \partial \ell(\varepsilon_t; \theta)/\partial \theta \) of (A.3) for observation \( t \) has the form

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

where \( g(t/T; \theta) = (y'_{t-1}, ..., y'_{t-p}, g'_0(t/T), g'_1(t/T), g'_q(t/T))' \), and \( T \) is the number of
The Hessian of (A.3) for observation \( t \) is given in the following lemma:

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

\[
\frac{\partial^2 \ell(\varepsilon_t; \theta)}{\partial \theta \partial \theta'} = -\frac{1}{\sigma^2} \{ g(t/T; \theta)g'(t/T; \theta) + \varepsilon_t(\theta) \frac{\partial^2 g(t/T; \theta)}{\partial \theta \partial \theta'} \},
\]

where

\[
g(t/T; \theta)g'(t/T; \theta)
= \begin{bmatrix}
M_{\phi\phi}(t/T) & M_{\phi\delta}(t/T) & M_{\phi\gamma}(t/T) & M_{\phi\psi}(t/T) \\
M_{\delta\delta}(t/T) & M_{\delta\gamma}(t/T) & M_{\delta\psi}(t/T) \\
M_{\gamma\gamma}(t/T) & M_{\gamma\psi}(t/T) & M_{\psi\psi}(t/T)
\end{bmatrix}, \quad (A.4)
\]

with

\[
M_{\phi\phi}(t/T) = \begin{bmatrix}
M_{\phi\phi_{11}}(t/T) & \ldots & M_{\phi\phi_{1p}}(t/T) \\
& \ddots & \\
M_{\phi\phi_{p1}}(t/T) & \ldots & M_{\phi\phi_{pp}}(t/T)
\end{bmatrix}, \quad (A.5)
\]

and

\[
M_{\phi\alpha}(t/T) = \begin{bmatrix}
M'_{\phi\alpha_{1}}(t/T) & \ldots & M'_{\phi\alpha_{p}}(t/T)
\end{bmatrix}', \quad (A.6)
\]

In (A.5) \( M_{\phi\phi_{ij}}(t/T) = y_{t-i}y'_{t-j}, \) and in (A.6), \( M_{\phi\alpha_{ij}}(t/T) = y_{t-j}g'_{\phi\alpha}(t/T), \) \( i, j = 1, \ldots, p \) and \( \alpha = \delta, \gamma, \psi. \) Finally, in (A.4),

\[
M_{\delta\delta}(t/T) = g_{\delta}(t/T)g'_{\delta}(t/T),
M_{\delta\alpha}(t/T) = g_{\delta}(t/T)g'_{\alpha}(t/T) + \varepsilon_t D_{\delta\alpha}(t/T), \quad \alpha = \gamma, \psi,
M_{\gamma\alpha}(t/T) = g_{\gamma}(t/T)g'_{\alpha}(t/T) + \varepsilon_t \text{diag}(g_{\gamma\alpha_{1}}(t/T), \ldots, g_{\gamma\alpha_{q}}(t/T)), \quad \alpha = \gamma, \psi,
M_{cc}(t/T) = g_{cc}(t/T)g'_{cc}(t/T) + \varepsilon_t \text{diag}(g_{cc_{1}}(t/T), \ldots, g_{cc_{q}}(t/T)).
\]

where \( D_{\delta\alpha}(t/T) = [0, g_{\alpha}(t/T)]'. \) The diagonal elements in the three diagonal matrices
are
\[
g_{\gamma j}(t/T) = \delta_j g_j(t/T)\{1 - g_j(t/T)\}\{1 - 2g_j(t/T)\}(t/T - c_j)^2, \\
g_{\varepsilon j}(t/T) = \delta_j^2 g_j(t/T)\{1 - g_j(t/T)\}\{1 - 2g_j(t/T)\}, \\
g_{\tau j}(t/T) = -\delta_j \gamma_j g_j(t/T)\{1 - g_j(t/T)\}\{1 - 2g_j(t/T)\}(t/T - c_j),
\]
for \(j = 1, \ldots, q\).

**Consistency**

We begin by proving consistency of the maximum likelihood estimator \(\hat{\theta}\) 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 \(\hat{\theta}\) is consistent for \(\theta_0\).

**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 \(\varepsilon_t\ (i = 1, \ldots, T)\) are \(iid\) with probability distribution function \(f(\varepsilon_t|\theta_0)\). If

(i) \(\theta \neq \theta_0\), then \(f(\varepsilon_t|\theta_0) \neq f(\varepsilon_t|\theta)\),

(ii) \(\theta_0 \in \Theta\) which is compact, and the density is positive (bounded away from zero) for all \(\theta \in \Theta\),

(iii) \(\ln f(\varepsilon_t|\theta)\) is continuous at each \(\theta \in \Theta\) with probability one,

(iv) \(E\sup_{\theta \in \Theta} |\ln f(\varepsilon_t|\theta)| < \infty\),

then the maximum likelihood estimator \(\hat{\theta} \overset{p}{\to} \theta_0\) as \(T \to \infty\).

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 \(|\ln f(\varepsilon_t|\theta)|\). This yields:

\[
|\ln f(\varepsilon_t|\theta)| = |\ln f(\varepsilon_t|\theta_0) + \frac{\partial}{\partial \theta'} \ln f(\varepsilon_t|\theta')(\theta - \theta_0)| \\
\leq |\ln f(\varepsilon_t|\theta_0)| + \|\frac{\partial}{\partial \theta'} \ln f(\varepsilon_t|\theta')\| \times \|\theta - \theta_0\| \\
\leq C_1 + C_2 C_3 < \infty,
\]

(A.7)
where \( \overline{\theta} \) is an intermediate value between \( \theta \) and \( \theta_0 \), \( |\ln f(\varepsilon_t|\theta_0)| \leq C_1 \) because of A4, and \( \|\frac{\partial}{\partial \theta} \ln f(\varepsilon_t|\overline{\theta})\| \leq C_2 \) because of Lemma A.1. Furthermore, \( \|\theta - \theta_0\| \leq C_3 \) follows from the fact that \( \Theta \) is compact (A3), so the elements of \( \|\theta - \theta_0\| \) are bounded for \( \theta \in \Theta \). In (A.7), \( C_i, i = 1, 2, 3, \) are generic positive constants. As \( |\ln f(\varepsilon_t|\theta)| \) is finite for all \( \theta \in \Theta \) (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 \( \hat{\theta} \), in addition to Lemma A.2 we need the five lemmas below.

**Lemma A.3.** Let \( x_t \) be a stationary and ergodic VAR(\( p \)) process with zero mean:

\[
x_t = \sum_{j=1}^{p} \Phi_j x_{t-j} + \varepsilon_t,
\]

where \( \{\varepsilon_t\} \sim \text{iid} N(0, \Sigma) \). It has the infinite-order moving average representation

\[
x_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j},
\]

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

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

**Proof.** Omitted.

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

\[
y_t = \delta(t/T) + \sum_{j=1}^{p} \Phi_j y_{t-j} + \varepsilon_t,
\]

where Assumptions A1–A5 hold. Let \( M_{\phi\phi m n T} = (1/T) \sum_{t=1}^{T} y_{t-m} y'_{t-n}, \) where \( n = \)
\[ m + s, \ s \geq 0. \ \text{Then} \]

\[
\lim_{T \to \infty} EM_{\phi_{mn}T} = \Psi \int_0^1 \delta(r)\delta'(r)dr \Psi' + \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi'_{i+s},
\]

where \( \Psi = \sum_{i=0}^{\infty} \Psi_i \).

**Proof.** Write (4) as follows:

\[
(I_k - \sum_{j=1}^{p} \Phi_j L^j) y_t = \delta(t/T) + \varepsilon_t.
\]

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

\[
y_t = (I_k - \sum_{j=1}^{p} \Phi_j L^j)^{-1} \delta(t/T) + \varepsilon_t = \sum_{i=0}^{\infty} \Psi_i \delta\left(\frac{t-i}{T}\right) + x_t,
\]

where \( x_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} \). Then,

\[
y_t - Ey_t = \sum_{i=0}^{\infty} \Psi_i \{\delta\left(\frac{t-i}{T}\right) - \delta\left(\frac{t}{T}\right)\} + O\left(\frac{1}{T}\right) + x_t.
\]

To simplify notation, in what follows we omit the asymptotically vanishing term \( O(1/T) \). We have

\[
E(y_{t-m} - Ey_{t-m})(y_{t-n} - Ey_{t-n})' = \left[ \sum_{i=0}^{\infty} \Psi_i \delta\left(\frac{t-m-i}{T}\right) - \delta\left(\frac{t-m}{T}\right) \right]\left[ \sum_{j=0}^{\infty} \Psi_j \delta\left(\frac{t-n-j}{T}\right) - \delta\left(\frac{t-n}{T}\right) \right]'
\]

\[ + Ex_{t-m}x_{t-n}, \]

where, from Lemma A.3,

\[
Ex_{t-s} = \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi'_{i+s}.
\]
Set $[Tr] = t$ and consider

$$
\frac{1}{T-n} \sum_{t=n+1}^{T} \delta\left(\frac{t-m-i}{T}\right)\delta'\left(\frac{t-n-j}{T}\right)
$$

$$
= \frac{1}{T-n} \sum_{t=n+1}^{T} \int_{t/T}^{(t+1)/T} \delta\left(\frac{[Tr] - m - i}{T}\right)\delta'\left(\frac{[Tr] - n - j}{T}\right) dr
$$

$$
= \int_{(n+1)/T}^{(T+1)/T} \delta\left(\frac{[Tr] - m - i}{T}\right)\delta'\left(\frac{[Tr] - n - j}{T}\right) dr
$$

$$
\rightarrow \int_{0}^{1} \delta(r)\delta'(r) dr,
$$

(A.8)

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

$$
\frac{1}{T-n} \sum_{t=n+1}^{T} E(y_{t-m} - E_{t-m})(y_{t-n} - E_{t-n})' \to \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi'_{i+s},
$$

as $T \to \infty$. Next,

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

so, from (A.8),

$$
\lim_{T \to \infty} \frac{1}{T-n} \sum_{t=n+1}^{T} E_{t-m}E_{t-n} = \Psi \int_{0}^{1} \delta(r)\delta'(r) dr \Psi'.
$$

Since

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

one obtains

$$
EM_{\phi mn} = \lim_{T \to \infty} \frac{1}{T-n} \sum_{t=n+1}^{T} E_{t-m}y_{t-n}
$$

$$
= \Psi \int_{0}^{1} \delta(r)\delta'(r) dr \Psi' + \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi'_{i+s},
$$

for $m = 1, ..., p$, $n \geq m$. 

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

20
Lemma A.5. Consider equation (A.1) and let $M_{\phi m T} = (1/T) \sum_{t=1}^{T} y_{t-m g'_{\phi}(t/T)}$, $\alpha = \delta, \gamma, c$. Then

$$\lim_{T \to \infty} EM_{\phi m T} = \Psi \int_{0}^{1} \delta(r) g'_{\phi}(r) dr,$$

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

Proof. Write

$$Ey_{t-m g'_{\phi}(t/T)} = \Psi \delta(\frac{t-m}{T}) g'_{\phi}(t/T).$$

Setting $t = [Tr]$,

$$\frac{1}{T-m} \sum_{t=m+1}^{T} Ey_{t-m g'_{\phi}(t/T)}$$

$$= \Psi \{ \frac{1}{T-m} \sum_{t=m+1}^{T} \delta(\frac{t-m}{T}) \} g'_{\phi}(t/T)$$

$$= \Psi \sum_{t=m+1}^{T} \int_{(m+1)/T}^{(t+1)/T} \delta(\frac{[Tr]-m}{T}) g'_{\phi}([Tr]/T) dr$$

$$= \Psi \int_{(m+1)/T}^{(T+1)/T} \delta(\frac{[Tr]-m}{T}) g'_{\phi}([Tr]/T) dr$$

$$\to \Psi \int_{0}^{1} \delta(r) g'_{\phi}(r) dr,$$

for $m = 1, \ldots, p$, as $T \to \infty$.

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

$$\lim_{T \to \infty} EM_{\alpha \beta T} = \int_{0}^{1} g_{\alpha}(r) g'_{\beta}(r) dr, \quad (A.9)$$

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

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

$$EM_{\alpha \beta T} = (1/T) \sum_{t=1}^{T} g_{\alpha}(t/T) g'_{\beta}(t/T),$$
\(\alpha, \beta = \gamma, c\). Set \(t = [Tr]\) and write

\[
EM_{\alpha\beta T} = \sum_{t=1}^{T} \int_{t/T}^{(t+1)/T} g_\alpha \left( \frac{[Tr]}{T} \right) g'_{\beta} \left( \frac{[Tr]}{T} \right) dr
\]

\[
= \int_{1/T}^{(T+1)/T} g_\alpha \left( \frac{[Tr]}{T} \right) g'_{\beta} \left( \frac{[Tr]}{T} \right) dr
\]

\[
\rightarrow \int_{0}^{1} g_\alpha (r) g'_{\beta} (r) dr
\]

for \(\alpha, \beta = \delta, \gamma, c\), as \(T \to \infty\). Since \(g_{\alpha i}(r), \alpha = \gamma, c, i = 1, \ldots, q\) is bounded for \(r \in (0, 1]\), see Lemma A.1, \(\int_{0}^{1} g_{\alpha i}(r) g_{\beta j}(r) dr < \infty\).

**Lemma A.7.** Consider equation (A.1) and let \(M_{\delta \alpha T} = (1/T) \sum_{t=1}^{T} g_{\delta}(t/T) g'_{\delta}(t/T) + \epsilon_t D_{\delta \alpha}(t/T), \alpha = \gamma, c\). Then

\[
\lim_{T \to \infty} EM_{\delta \delta T} = \int_{0}^{1} g_{\delta}(r) g'_{\delta}(r) dr,
\]

(A.10)

and

\[
\lim_{T \to \infty} EM_{\delta \alpha T} = \int_{0}^{1} g_{\delta}(r) g'_{\alpha}(r) dr,
\]

(A.11)

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

**Proof.** Since \(E \epsilon_t D_{\delta \alpha}(t/T) = 0\), \(\alpha = \gamma, c\), \(EM_{\delta \alpha T} = (1/T) \sum_{t=1}^{T} g_{\delta}(t/T) g'_{\alpha}(t/T)\), \(\alpha = \gamma, c\). Set \(t = [Tr]\) and write

\[
EM_{\delta \delta T} = \sum_{t=1}^{T} \int_{t/T}^{(t+1)/T} g_{\delta} \left( \frac{[Tr]}{T} \right) g'_{\delta} \left( \frac{[Tr]}{T} \right) dr
\]

\[
= \int_{1/T}^{(T+1)/T} g_{\delta} \left( \frac{[Tr]}{T} \right) g'_{\delta} \left( \frac{[Tr]}{T} \right) dr
\]

\[
\rightarrow \int_{0}^{1} g_{\delta} (r) g'_{\delta} (r) dr,
\]

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

\[
\lim_{T \to \infty} EM_{\delta \alpha T} = \int_{0}^{1} g_{\delta} (r) g'_{\alpha} (r) dr,
\]

(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 $\hat{\theta} \to \theta_0$ as $T \to \infty$. Assume further that Assumptions A1-A5 are valid. Then

$$\sqrt{T}(\hat{\theta} \to \theta_0) \overset{d}{\to} N(0, J^{-1})$$

where

$$J = \lim_{T \to \infty} \mathbb{E}J_T = (1/\sigma^2) \lim_{T \to \infty} \left(1/T \sum_{t=1}^{T} \mathbb{E}\frac{\partial}{\partial \theta} \ln f(\varepsilon_t | \theta_0) \frac{\partial}{\partial \theta} \{\ln f(\varepsilon_t | \theta_0)\}\right)$$

$$= (1/\sigma^2) \lim_{T \to \infty} \left[\begin{array}{cccc}
\mathbb{E}M_{\phi_1 T} & \mathbb{E}M_{\phi_2 T} & \mathbb{E}M_{\phi_3 T} & \mathbb{E}M_{\phi_4 T} \\
\mathbb{E}M_{\phi_5 T} & \mathbb{E}M_{\phi_6 T} & \mathbb{E}M_{\phi_7 T} & \mathbb{E}M_{\phi_8 T} \\
\mathbb{E}M_{\phi_9 T} & \mathbb{E}M_{\phi_{10} T} & \mathbb{E}M_{\phi_{11} T} & \mathbb{E}M_{\phi_{12} T}
\end{array}\right]. \quad \text{(A.13)}$$

In (A.13),

$$\mathbb{E}M_{\phi T} = \begin{bmatrix} \mathbb{E}M_{\phi_{11} T} & \cdots & \mathbb{E}M_{\phi_{13} T} \\ \mathbb{E}M_{\phi_{14} T} & \cdots & \mathbb{E}M_{\phi_{15} T} \end{bmatrix},$$

and

$$\mathbb{E}M_{\phi T}(t/T) = \begin{bmatrix} \mathbb{E}M_{\phi_{11} T} & \cdots & \mathbb{E}M_{\phi_{13} T} \\ \mathbb{E}M_{\phi_{14} T} & \cdots & \mathbb{E}M_{\phi_{15} T} \end{bmatrix}',$n

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

$$\lim_{T \to \infty} \mathbb{E}M_{\phi mn T} = \Psi \int_0^1 \delta(r)\delta'(r) dr \Psi' + \sum_{i=0}^{\infty} \Psi_i \Sigma \Psi_{i+s},$$

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

$$\lim_{T \to \infty} \mathbb{E}M_{\phi m T} = \Psi \int_0^1 \delta(r)g_{\phi T}(r)dr$$

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

$$\lim_{T \to \infty} M_{\alpha T} = [m_{\alpha_i,ij}] = \int_0^1 g_{\alpha T} g_{\beta T}(r)dr$$

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

**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-
is satisfied and

(i) \( \theta_0 \) is an interior point of \( \Theta \),

(ii) \( f(\varepsilon|\theta) \) is twice continuously differentiable and \( f(\varepsilon|\theta) > 0 \) in a neighbourhood \( \mathcal{N} \) of \( \theta_0 \),

(iii) \( \int \sup_{\theta \in \mathcal{N}} \| \frac{\partial}{\partial \theta} f(\varepsilon|\theta) \| d\varepsilon < \infty \) and \( \int \sup_{\theta \in \mathcal{N}} \| \frac{\partial^2}{\partial \theta \partial \theta'} f(\varepsilon|\theta) \| d\varepsilon < \infty \),

(iv) \( J = \mathbb{E} \left\{ \ln \frac{\partial}{\partial \theta} f(\varepsilon|\theta_0) \right\} \{ \frac{\partial}{\partial \theta} \ln f(\varepsilon|\theta_0) \} \) exists and is nonsingular,

(v) \( \mathbb{E} \sup_{\theta \in \mathcal{N}} \| \frac{\partial^2}{\partial \theta \partial \theta'} \ln f(\varepsilon|\theta) \| < \infty \), where \( \mathcal{N} \) is a neighbourhood of \( \theta_0 \).

Then \( \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, J^{-1}) \).

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 \( J_T \) is nonsingular for \( T > 3q + kp + 1 \), so \( J_T^{-1} \) exists. Lemma A.4 yields \( \lim_{T \to \infty} \mathbb{E} M_{\phi}(t/T) \), Lemma A.5 provides the elements of \( \lim_{T \to \infty} \mathbb{E} M_{\phi}(t/T) \), \( \alpha = \delta, \gamma, c \), Lemma A.6 does the same for \( \lim_{T \to \infty} M_{\phi}(t/T) \), \( \alpha, \beta = \gamma, c \) and Lemma A.7 for \( \lim_{T \to \infty} M_{\phi}(t/T) \) and \( \lim_{T \to \infty} M_{\phi}(t/T) \), \( \alpha = \gamma, c \). Putting these together defines \( J = \lim_{T \to \infty} J_T \). Condition (v) is satisfied because the deterministic components in the matrix \( \frac{\partial^2}{\partial \theta \partial \theta'} \ln f(\varepsilon|\theta) \) are bounded. Furthermore, the elements of this matrix have finite expectations for all \( \theta \in \Theta \). Finally, since matrix inversion is a continuous transformation, it follows from the continuous mapping theorem that when \( J_T \to J \), then \( J_T^{-1} \to J^{-1} \). This concludes the proof.
Table B.1: Unrestricted System Estimates for the Hemispheric Temperature VSM–AR.

<table>
<thead>
<tr>
<th>Northern Hemisphere:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t^n = -0.148 + 1.194 g_1 (t/T; \hat{\alpha}<em>{11}, \hat{\gamma}</em>{11}) + 0.176 g_2 (t/T; \hat{\alpha}<em>{12}, \hat{\gamma}</em>{12}) - 0.242 g_3 (t/T; \hat{\alpha}<em>{13}, \hat{\gamma}</em>{13})$</td>
</tr>
<tr>
<td>$+ 0.117 y_{t-1}^n + 0.021 y_{t-2}^n - 0.174 y_{t-3}^n + 0.463 y_{t-1}^s - 0.237 y_{t-2}^s + 0.230 y_{t-3}^s + \tilde{\varepsilon}_t^n$</td>
</tr>
<tr>
<td>$R^2 = 0.885, \quad \hat{\sigma}^n = 0.116, \quad Sk = 0.073, \quad Ek = 0.071, \quad LJB = 0.182(0.913)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southern Hemisphere:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t^s = -1.254 + 3.257 g_4 (t/T; \hat{\alpha}<em>{21}, \hat{\gamma}</em>{21}) - 0.192 g_5 (t/T; \hat{\alpha}<em>{22}, \hat{\gamma}</em>{22})$</td>
</tr>
<tr>
<td>$+ 0.456 y_{t-1}^s - 0.167 y_{t-2}^s + 0.141 y_{t-3}^s + \tilde{\varepsilon}_t^s$</td>
</tr>
<tr>
<td>$R^2 = 0.881, \quad \hat{\sigma}^s = 0.093 \quad Sk = 0.008, \quad Ek = -0.051, \quad LJB = 0.020(0.990)$</td>
</tr>
<tr>
<td>$\hat{\rho}_{ns} = 0.568$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition Functions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1 (t/T; \hat{\alpha}<em>{11}, \hat{\gamma}</em>{11}) = (1 + \exp{-3.587[t/T - 0.907]/0.289})^{-1}$</td>
</tr>
<tr>
<td>$g_2 (t/T; \hat{\alpha}<em>{12}, \hat{\gamma}</em>{12}) = (1 + \exp{-5.0[t/T - 0.443]/0.289})^{-1}$</td>
</tr>
<tr>
<td>$g_3 (t/T; \hat{\alpha}<em>{13}, \hat{\gamma}</em>{13}) = (1 + \exp{-5.0[t/T - 0.712]/0.289})^{-1}$</td>
</tr>
<tr>
<td>$g_4 (t/T; \hat{\alpha}<em>{21}, \hat{\gamma}</em>{21}) = (1 + \exp{-0.254[t/T - 0.95]/0.289})^{-1}$</td>
</tr>
<tr>
<td>$g_5 (t/T; \hat{\alpha}<em>{22}, \hat{\gamma}</em>{22}) = (1 + \exp{-12.836[t/T - 0.249]/0.289})^{-1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System Statistics:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln L = 311.136, \quad AIC = -9.435, \quad BIC = -9.416$</td>
</tr>
</tbody>
</table>

**Note:** $\hat{\rho}_{ns}$ 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. AIC and BIC denote, respectively, the system Akaike information criterion and the Rissanen–Schwarz Bayesian information criterion. Values in parentheses beside test statistics are $p$-values. $LM_{13}$ denotes the system test, also based on Rao’s $F$, for no remaining mean shifts obtained by using a third–order Taylor approximation.
References


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

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>ln $L$</th>
<th>AIC</th>
<th>BIC</th>
<th>$K$</th>
<th>LR Test</th>
<th>Rao’s $F$ Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted(^\d)</td>
<td>311.136</td>
<td>-9.096</td>
<td>-8.550</td>
<td>29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$H_0 : \gamma_{11} = \gamma_{21}, c_{11} = c_{21}$</td>
<td>308.665</td>
<td>-8.843</td>
<td>-8.824</td>
<td>27</td>
<td>$\chi^2_{(2)} = 4.942(0.084)$</td>
<td>$F_{(2,151)} = 2.296(0.104)$</td>
</tr>
</tbody>
</table>

Weak Co–shifting

\(^\d\) Results for the model with no co–shifting restrictions imposed.

Note: There are 165 usable observations. The unrestricted model contains 29 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 $\chi^2_j$ is for a likelihood ratio test involving $j$ parameter restrictions. Rao’s $F$–test statistics are approximate system LM tests as described by Teräsvirta, Tjostheim and Granger (2010). Values in parentheses are asymptotic $p$–values.
Table 2: System Estimates for the Hemispheric Temperature VSM–AR with Co–Shifting Restrictions Imposed.

<table>
<thead>
<tr>
<th>Northern Hemisphere:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t^N = -0.173 + 1.390 g_1 (t/T; \hat{\alpha}<em>{11}, \hat{\gamma}</em>{11}) + 0.195 g_2 (t/T; \hat{\alpha}<em>{12}, \hat{\gamma}</em>{12}) - 0.269 g_3 (t/T; \hat{\alpha}<em>{13}, \hat{\gamma}</em>{13})$</td>
<td>( \begin{array}{l} (0.042) \ (0.232) \end{array} )</td>
</tr>
<tr>
<td>[ + 0.141 y_{t-1}^N + 0.047 y_{t-2}^N - 0.156 y_{t-3}^N + 0.419 y_{t-1}^S - 0.273 y_{t-2}^S + 0.185 y_{t-3}^S + \varepsilon_t^N ]</td>
<td>( \begin{array}{l} (0.071) \ (0.072) \ (0.068) \ (0.101) \ (0.105) \end{array} )</td>
</tr>
<tr>
<td>$R^2 = 0.886$, $\hat{\sigma}_n = 0.115$, $Sk = 0.094$, $Ek = 0.165$, $LJB = 0.431(0.806)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Southern Hemisphere:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t^S = -0.208 + 0.383 g_1 (t/T; \hat{\alpha}<em>{11}, \hat{\gamma}</em>{11}) + 0.153 g_5 (t/T; \hat{\alpha}<em>{22}, \hat{\gamma}</em>{22})$</td>
<td>( \begin{array}{l} (0.035) \ (0.101) \end{array} )</td>
</tr>
<tr>
<td>[ + 0.462 y_{t-1}^S - 0.166 y_{t-2}^S + 0.151 y_{t-3}^S + \varepsilon_t^S ]</td>
<td>( \begin{array}{l} (0.073) \ (0.076) \ (0.069) \end{array} )</td>
</tr>
<tr>
<td>$R^2 = 0.881$, $\hat{\sigma}_s = 0.093$, $Sk = 0.161$, $Ek = 0.144$, $LJB = 0.852(0.653)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition Functions:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1 (t/T; \hat{\alpha}<em>{11}, \hat{\gamma}</em>{11}) = (1 + \exp{ - 4.041 [t/T - 0.889] / 0.289 })^{-1}$</td>
<td>( \begin{array}{l} (0.669) \ (0.028) \end{array} )</td>
</tr>
<tr>
<td>$g_2 (t/T; \hat{\alpha}<em>{12}, \hat{\gamma}</em>{12}) = (1 + \exp{ - 45.009 [t/T - 0.443] / 0.289 })^{-1}$</td>
<td>( \begin{array}{l} (10.400) \ (0.009) \end{array} )</td>
</tr>
<tr>
<td>$g_3 (t/T; \hat{\alpha}<em>{13}, \hat{\gamma}</em>{13}) = (1 + \exp{ - 50[t/T - 0.710] / 0.289 })^{-1}$</td>
<td>( \begin{array}{l} (100.005) \end{array} )</td>
</tr>
<tr>
<td>$g_5 (t/T; \hat{\alpha}<em>{22}, \hat{\gamma}</em>{22}) = (1 + \exp{ - 7.982 [t/T - 0.474] / 0.289 })^{-1}$</td>
<td>( \begin{array}{l} (1.798) \ (0.027) \end{array} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System Statistics:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln L = 308.665$, $AIC = -8.843$, $BIC = -8.824$, $LM_{\Omega} = 4.624 \times 10^{-4}(0.999)$, $Sk_s = 1.439(0.487)$</td>
<td></td>
</tr>
<tr>
<td>$Ek_s = 0.792(0.673)$, $LJB_s = 2.231(0.693)$, $LM_{VAR(4)} = 4.56 \times 10^{-3}(0.999)$, $LM_{VAR(6)} = 4.48 \times 10^{-3}(0.999)$</td>
<td></td>
</tr>
<tr>
<td>$LM_{VAR(8)} = 5.01 \times 10^{-3}(0.999)$, $LM_{VAR(10)} = 6.02 \times 10^{-3}(0.999)$, $LM_{VAR(12)} = 5.72 \times 10^{-3}(0.999)$</td>
<td></td>
</tr>
<tr>
<td>$LM_{\delta} = 0.020(0.999)$</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* $\hat{\rho}_{ns}$ 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. $LM_{\Omega}$ 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. $LM_{\delta}$ denotes the system test, also based on Rao’s $F$, for no remaining mean shifts obtained by using a third–order Taylor approximation.
Table 3: Transitions and Shift Dates for the VSM–AR Model with Co–Shifting.

<table>
<thead>
<tr>
<th></th>
<th>Northern Hemisphere</th>
<th></th>
<th>Southern Hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i</td>
<td>$\gamma$</td>
<td>$\hat{c}$</td>
</tr>
<tr>
<td>1</td>
<td>4.014</td>
<td>0.889</td>
<td>1972</td>
</tr>
<tr>
<td>2</td>
<td>45.009</td>
<td>0.443</td>
<td>1921</td>
</tr>
<tr>
<td>3</td>
<td>50.000</td>
<td>0.710</td>
<td>1966</td>
</tr>
</tbody>
</table>

*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, 1850–2017.
Figure 2: Bivariate VSM-AR Results with Co-Trending Restrictions for Temperature Anomalies for the Northern (Panel A) and Southern (Panel B) Hemispheres, 1850–2017. The dashed red line indicates the shifting mean and the dash–dot lines indicate the estimated transition functions.