Statistica Sinica **27** (2017), 23-42
doi:http://dx.doi.org/10.5705/ss.202014.0024

THE CEPSTRAL MODEL FOR MULTIVARIATE TIME SERIES: THE VECTOR EXPONENTIAL MODEL

Scott H. Holan, Tucker S. McElroy and Guohui Wu

University of Missouri, U.S. Census Bureau and SAS Institute Inc.

Abstract: Vector autoregressive models have become a staple in the analysis of multivariate time series and are formulated in the time domain as difference equations, with an implied covariance structure. In many contexts, it is desirable to work with a stable, or at least stationary, representation. To fit such models, one must impose restrictions on the coefficient matrices to ensure that certain determinants are nonzero which, except in special cases, can prove burdensome. To circumvent these difficulties, we propose a flexible frequency domain model expressed in terms of the spectral density matrix. Specifically, this paper treats the modeling of covariance stationary vector-valued time series via an extension of the exponential model for the spectrum of a scalar time series. We discuss the modeling advantages of the vector exponential model and its computational facets, such as how to obtain moving average coefficients from given cepstral coefficients. We demonstrate the utility of our approach through simulation as well as two illustrative data examples focusing on multi-step ahead forecasting and estimation of squared coherence.

Key words and phrases: Autocovariance matrix, Bayesian estimation, cepstral, coherence, moving average coefficients, spectral density matrix, stochastic search variable selection.

1. Introduction

This paper treats the modeling of covariance stationary vector-valued time series through an extension of the exponential model of Bloomfield (1973). Such a process will be called a Vector EXPonential (VEXP). In contrast to Vector AutoRegressive Moving Average (VARMA) models, the VEXP processes are always invertible, which means that the (causal) Moving Average (MA) form of the process can be inverted into a (stable) Vector AutoRegressive (VAR) form – or equivalently, that the spectral density matrix of the VEXP is non-singular at all frequencies. Necessarily, a VEXP process is also stable, or stationary, which here means that the spectral density matrix has finite determinant at all frequencies. We note that, when estimation proceeds in an unconstrained fashion (e.g., by ordinary least squares) a VAR or VARMA process need not be stable or invertible; see Lütkepohl (2007) for a basic treatment. Nevertheless, there are

practical scenarios where these stationarity restrictions on the vector process are actually necessary.

Although Gaussian maximum likelihood estimation can still proceed when a vector process is non-invertible, so long as the singularities occur at a set of frequencies that have Lebesgue measure zero (see McElroy and Trimbur (2015) for proof and discussion), Whittle estimation, the Gaussian quasi-maximum likelihood procedure described in Taniguchi and Kakizawa (2000), becomes intractable. In addition, the long-term forecasting filters are not well-defined (see the discussion in McElroy and McCracken (2014)), because such filters rely on the ability to recover the innovations from the MA form of the process. Another motivation for using invertible processes arises from a popular model for co-integration (Engle and Granger (1987)), called the common trends formulation (Stock and Watson (1988)). As shown in McElroy and Trimbur (2015), a co-integrated data process can naturally arise from a co-linear trend process so long as the noise process (after differencing, if appropriate) is invertible. If the noise process spectrum has singularities, then the resulting spectrum of the data process can have singularities as well, which may be an undesirable feature.

By reparametrization, it is possible to enforce that a matrix polynomial is stable (i.e., its determinant has roots outside the unit circle), which can be used to ensure that MA polynomials are invertible. For a VEXP process both stability and invertibility are automatic, while the parameters are completely unconstrained in $\mathbb{R}^Q$, where Q is the total number of parameters. Moreover, the VEXP class of processes is arbitrarily dense in the space of stable invertible vector processes, much in the same way that the EXP process can approximate a stationary univariate process arbitrarily well. This approximation can be made arbitrarily accurate, and the novel algorithms developed herein allow for efficient computation of the cepstral representation.

Many situations arise in which modeling the data by a stationary (stable) vector model is desirable. For example, this might occur if the data had already been made stationary by differencing, or perhaps by utilizing a common trend structure for an unobserved component (e.g., see Harvey (1990) or Nyblom and Harvey (2000)). For the VAR class, one would need to impose restrictions on the parameters to ensure a stable result, or have recourse to using the Yule-Walker estimates (see the discussion in Lütkepohl (2007)), which guarantee stable outcomes. However, if a Bayesian treatment is desired, prior elicitation becomes a quagmire, since the implicit restrictions imply that the parameters must be supported on a complicated manifold. The Bayesian treatment for the VARMA class of models is even more challenging. However, the cepstral approach of the VEXP allows for the entries of each parameter matrix to be any real number, so that taking independent vague Gaussian priors is a sensible and coherent choice that guarantees a stable outcome.

Following the pioneering paper of Bloomfield (1973), the literature on cepstral models has experienced substantial growth. For example, Holan (2004), Holan, McElroy and Chakraborty (2009), McElroy and Holan (2012), and the references therein, discuss univariate short memory, long memory, and seasonal long memory cepstral models, respectively. In contrast, McElroy and Holan (2014) provides a comprehensive treatment of the asymptotic theory for cepstral random fields. Nevertheless, currently no multivariate cepstral model exists; while Theorem 3 of Taniguchi, Puri and Kondo (1996) defines the spectral density as a matrix exponential, this is not developed as a time series model. Instead, they examine the matrix exponential for the spectral density, whereas we examine the matrix exponential of the MA filter. Given this lacuna, we propose a novel class of multivariate cepstral models and illustrate their utility. Importantly, we provide precise mathematical development and computational algorithms for the proposed class of models, which differs significantly from both the univariate time series and random field cases.

There are several facets of this VEXP process that are fascinating and non-intuitive. In particular, because we are studying vector time series, the algebra that relates the cepstral coefficients to the MA coefficients is no longer Abelian, and great care is needed in working with the matrix exponential. Background material as well as our basic VEXP model is provided in Section 2 – moving from mathematical foundations to the explicit definition and on to algorithmic considerations. Section 3 discusses different aspects associated with modeling using the VEXP and provides details surrounding Bayesian estimation. Subsequently, two bivariate data illustrations involving multi-step ahead forecasting and squared coherence estimation are presented in Section 4. Section 5 provides concluding discussion. For convenience of exposition, all proofs and derivations, discussion of several VEXP modeling applications, and two simulated examples are given in a Supplemental Appendix.

2. The VEXP Model

2.1. The VEXP process

General discussion concerning vector time series is provided in Hannan (1970) and Brockwell and Davis (1991). Here, we use $'$ for transpose and * for conjugate transpose of a complex-valued matrix. For a m-variate time series, the spectral density matrix f is a $m \times m$ dimensional matrix function of frequency λ, $f(\lambda)$ is always Hermitian nonnegative definite, and is often positive definite (pd). Moreover, the autocovariance function (acvf) for a mean-zero process is defined via $\Gamma_h = \mathbb{E}[X_{t+h}X_t']$, and is related to the spectral density matrix (sdm) via the inverse Fourier transform (FT):

$$\Gamma_h = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda) z^{-h}\, d\lambda,$$

where $z = \exp(-i\lambda)$. This integration works component-wise on each entry of the sdm. This relationship can be re-expressed in terms of the FT as

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \Gamma_h z^h.$$

This relation is indicative of a more general Hilbert Space expansion of spectral matrix functions, where a generic function, g, of frequency can be expanded in terms of the orthonormal basis $\{z^h\}$, yielding coefficient matrices given by the inner product of g with z^h.

There is a MA decomposition for vector time series, which amounts to a particular form for the sdm; see Brockwell and Davis (1991) for a comprehensive discussion. Let $\Psi(z) = \sum_{j\geq 0} \Psi_j z^j$ be the causal representation of the time series (and for identifiability, we have $\Psi_0 = I$, the identity matrix), such that $X_t = \Psi(B)\epsilon_t$, for some vector white noise $\{\epsilon_t\}$ with (lag zero) covariance matrix $\underline{\Sigma}$. Then the sdm is $f(\lambda) = \Psi(z)\underline{\Sigma}\Psi'(z)$. Given the above definitions, the acvf is related to the MA filter $\Psi(B)$ by $\Gamma_h = \sum_{j\geq 0} \Psi_{j+h}\underline{\Sigma}\Psi_j'$.

This shows that knowledge of the MA filter is sufficient to determine the acvf, but the individual coefficients of the MA filter must satisfy various implicit constraints to generate an identifiable and invertible process; it is then not convenient to utilize these coefficients Ψ_j as parameters (as happens in a VMA model). By rewriting the MA filter in terms of a matrix exponential, we can automatically enforce identifiability and invertibility through a completely unconstrained parametrization. This idea is the matrix analogue of the univariate correspondence between MA and cepstral coefficients elucidated in Pourahmadi (1984). The matrix exponential is defined in Artin (1991), and many of its properties are provided in Chiu, Leonard and Tsui (1996). For any complex-valued square matrix A, the matrix exponential $\exp(A)$ is defined via the Taylor series expansion of $\exp(x)$ evaluated at $x = A$, and Proposition 8.3 of Artin (1991, p. 139) guarantees the convergence.

Therefore, we consider MA filters of the form $\Psi(z) = \exp\{\Omega(z)\}$ for some causal power series $\Omega(z) = \sum_{k\geq 1} \Omega_k z^k$, called the cepstral filter $\Omega(B)$. Consider the condition

$$\|I - \Psi(z)\| < 1 \quad \forall z \in D \tag{2.1}$$

for some matrix norm $\|\cdot\|$ and $D = \{z \in \mathbb{C} : |z| \leq 1\}$. Under this condition $\Psi(B)$ has a cepstral representation and, conversely, any cepstral filter $\Omega(B)$ generates a MA filter $\exp\{\Omega(B)\}$ that is invertible (its determinant has no roots inside the unit circle $D = \{z \in \mathbb{C} : |z| \leq 1\}$) with inverse $\exp\{-\Omega(B)\}$. This is shown in Corollary 1 of the Supplemental Appendix. Note that (2.1) is sufficient to show that $\Psi(z)^{-1}$ exists for all $z \in D$ (which is equivalent to $\det \Psi(z) \neq 0$ for all $z \in D$), but it is not necessary.

We propose to restrict attention to such processes for modeling stationary time series. When $\Omega(B)$ is a matrix polynomial, the corresponding process is called a VEXP, in analogy with the univariate EXP of Bloomfield (1973). We introduce the notation $[\Omega]_1^q(z) = \sum_{k=1}^q \Omega_k z^k$, and the corresponding process is called a VEXP(q). A comprehensive discussion motivating this definition appears in an earlier version of this manuscript; see Holan, McElroy and Wu (2014). We emphasize that each entry of each Ω_k can be any real number, including zero, although Ω_q does not have every entry equal to zero (else decrement q). The matrix polynomial $[\Omega]_1^q(z)$ always converges, so the VEXP(q) is always invertible and identifiable. In summary, the VEXP(q) process has MA filter

$$\Psi(z) = \exp\{[\Omega]_1^q(z)\}. \tag{2.2}$$

Although the elements of Ω_k are not as readily interpretable in the time domain as are the elements of the transition matrices of a VAR(p) model, typically this is not of prime concern. In practice, usually the primary concern resides in estimation of such target quantities as forecasts, trend estimates, and coherence, among others. In this sense, the elements of Ω_k can often be viewed as nuisance parameters and there is no need for direct interpretation. Issues of interpretability within the VEXP class of models is similar to that of MA coefficients in a VMA or VARMA model, though the cepstral representation may be less problematic directly in the frequency domain. Nevertheless, in some instances interest resides in interpreting the parameters of a VAR model (e.g., when making inference concerning Granger causality). In such cases we advocate estimating the VAR representation of the associated VEXP model or directly estimating a VAR model.

The white noise process has covariance matrix $\underline{\Sigma}$, which we can represent as the matrix exponential of some real symmetric matrix, say Ω_0, by Lemma 1 of Chiu, Leonard and Tsui (1996). Hence, we write $\underline{\Sigma} = \exp(\Omega_0)$, which is constrained to be symmetric. The formula $\det \underline{\Sigma} = \exp\{\mathrm{tr}\, \Omega_0\}$ follows from Proposition 5.11 of Artin (1991, p.286), and relates the eigenvalues of $\underline{\Sigma}$ to those of Ω_0. The sdm of the VEXP(q) can be written

$$f(\lambda) = \exp\{[\Omega]_1^q(z)\} \exp\{\Omega_0\} \exp\{[\Omega']_1^q(\overline{z})\}. \tag{2.3}$$

The VEXP(q) process is also dense in the space of causal processes with MA filter satisfying (2.1), in the sense that by taking q sufficiently large we can approximate any such process. Because our concept of density involves the mean square norm on vector processes, the appropriate matrix norm to utilize with condition (2.1) is the Frobenius norm; this implicitly defines a subset of the ℓ_2 space of causal power series, over which we formulate our results.

Proposition 1. *Let $\{X_t\}$ be given by $X_t = \Psi(B)\epsilon_t$ such that, with respect to the Frobenius norm, (2.1) holds, and also that $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$, so that the cepstral power series $\Omega(B)$ is well-defined. Let $\Psi^{(q)}(z)$ be the MA power series corresponding to the truncated cepstral polynomial $[\Omega]_1^q(B)$, and write $X_t^{(q)} = \Psi^{(q)}(B)\epsilon_t$ for each integer q. Then the time series $\{X_t^{(q)}\}$ forms a Cauchy sequence, and converges in mean square to $\{X_t\}$.*

This result gives us confidence that any time series having a causal MA representation satisfying (2.1) can be approximated arbitrarily well by a VEXP(q) by taking q suitably large. Of course, the same can be said of finite order VAR, VMA, or VARMA models, but such models require nuanced parameter restrictions to achieve stability, invertibility, and/or identifiability (see the discussion in Brockwell and Davis (1991), given at more length in Hannan and Deistler (1988)). Consider the case of a VMA (the discussion can be extended to VARMA, but is more complicated due to the possibility of cancelation of common factors) as described in Lütkepohl (2007), with polynomial $\Psi(z)$. Imposing $\det\Psi(z) \neq 0$ for all $z \in D$ ensures invertibility and identifiability as well; only imposing $\det \Psi(z) \neq 0$ for all $z \in \partial D$ still provides invertibility, but the model will not be identified. In contrast, the VEXP(q) corresponds to an infinite order VMA with $\det\Psi(z) \neq 0$ for all $z \in D$, so that invertibility is automatic; it is also identified.

Proposition 2. *A VEXP(q) process with $q < \infty$ is stable, invertible, and identifiable.*

In practice, the VEXP(q) model provides a particularly appealing candidate model in cases where no clear low-order VAR or VMA can be definitively selected. This occurs when neither the partial autocorrelation function (pacf) or autocorrelation function (acf) truncates after a low order lag. Instead, the pacf and acf are indicative that a VARMA or VEXP model would be more appropriate; the former is plagued with estimation challenges, whereas the latter has a Euclidean stationary parameter space.

2.2. The moving average-cepstral bijection

In the univariate case, one may differentiate (2.2) with respect to z, match coefficients, and arrive at the recurrence relations given in Pourahmadi (1984) and Hurvich (2002). This produces a recursive relation involving previously computed MA coefficients and a finite number of cepstral coefficients. Such an approach is demonstrably false in the multivariate case, because differentiation of the matrix exponential must allow for the non-Abelian algebra. In particular, the derivative of $\exp\{\Omega(z)\}$ is not equal to $\dot{\Omega}(z)\exp\{\Omega(z)\}$, except in the case that the terms in $\Omega(z)$ commute with each other. Instead, we can relate the MA coefficients

to cepstral matrices by expanding the matrix exponential using a Taylor series and matching corresponding powers of z. Then straightforward combinatorics provides the following relationship:

$$\Psi_k = \sum_{\ell \geq 1} \frac{1}{\ell!} \left(\sum_{\lambda \vDash k : |\lambda| = \ell} \Pi_{j=1}^{\ell} \Omega_{i_j} \right)$$

for $k \geq 1$. The symbols in the summation are defined as follows: $\lambda \vDash k$ denotes a partition of the integer k – actually $\vdash$ is typically used (Stanley (1997, p. 28)), but because we care about the order of the numbers occurring in the partition, we use the notation $\vDash$ instead. Also, $|\lambda| = \ell$ says that the number of elements in the partition is ℓ. So we sum over all partitions of the integer k into ℓ pieces, say $i_1, i_2, \ldots, i_\ell$ with $\sum_{j=1}^{\ell} i_j = k$. For example, the size two partitions of the integer 3 are given by $(1, 2)$ and $(2, 1)$, and these must be accounted as distinct terms in the summation, since $\Omega_1 \Omega_2$ is not equal to $\Omega_2 \Omega_1$. Actually, when all the Ω_k matrices commute with each other, all partitions of a given size and configuration produce the same result, and the above formula simplifies. However, this case is of little practical interest. We can also produce a relationship of the cepstral matrices to the MA coefficients by expanding the matrix logarithm and matching powers of z:

$$\Omega_k = \sum_{\ell \geq 1} \frac{(-1)^{\ell}}{\ell} \left(\sum_{\lambda \vDash k : |\lambda| = \ell} \Pi_{j=1}^{\ell} \Psi_{i_j} \right).$$

This is typically of lesser interest in applications. For modeling, one posits values for the cepstral matrices, and determines the MA coefficients. Counting the numbers of partitions is laborious, because the total number of (ordered) partitions of an integer k is equal to 2^k. The first few MA coefficients are given by (recall that $\Psi_0 = I$)

$$\begin{aligned}
\Psi_1 &= \Omega_1, \\
\Psi_2 &= \Omega_2 + \frac{\Omega_1^2}{2}, \\
\Psi_3 &= \Omega_3 + \frac{\Omega_1 \Omega_2 + \Omega_2 \Omega_1}{2} + \frac{\Omega_1^3}{6}.
\end{aligned}$$

For higher MA coefficients, the number of terms quickly grows out of scope. Note that for Ψ_3 the non-Abelian nature of the cepstral matrices comes into play, since in general $\Omega_1 \Omega_2 \neq \Omega_2 \Omega_1$. However, a simpler method is available that allows the computer to implicitly determine the appropriate partitions. Let $\Upsilon(z) = \Omega(z)/z$, which is a well-defined power series in z. In the case that $\Omega(z)$ is a degree q matrix polynomial, then $\Upsilon(z)$ is a degree $q - 1$ matrix polynomial. Denote the j-th derivative of a polynomial with respect to z by the superscript (j).

Proposition 3. *Consider the MA power series $\Psi(z)$ such that, with respect to the Frobenius norm,* (2.1) *holds, with cesptral power series $\Omega(z)$. With $\Upsilon(z) = \Omega(z)/z$, the k-th MA coefficient can be computed by*

$$\Psi_k = \frac{1}{k!}\sum_{\ell=1}^{k}\binom{k}{\ell}\left[\Upsilon(z)^{\ell}\right]^{(k-\ell)}|_{z=0} = \sum_{\ell=1}^{k}\frac{1}{\ell!}\left[\Upsilon(z)^{\ell}\right]_{k-\ell}. \tag{2.4}$$

Letting $\Xi(z) = (\Psi(z) - I)/z$, the k-th cepstral coefficient can be computed by

$$\Omega_k = \frac{1}{k!}\sum_{\ell=1}^{k}\binom{k}{\ell}(-1)^{\ell}(\ell-1)!\left[\Xi(z)^{\ell}\right]^{(k-\ell)}|_{z=0} = -\sum_{\ell=1}^{k}\frac{(-1)^{\ell}}{\ell}\left[\Xi(z)^{\ell}\right]_{k-\ell}. \tag{2.5}$$

Remark 1. Evidently, $\Psi(B)$ is block upper triangular if and only if $\Omega(B)$ is block upper triangular; hence Granger non-causality (e.g., see Lütkepohl, 2007) can be parsed purely in terms of the cepstral power series. In particular, if the upper right block of each Ω_k is zero, then the second block of time series does not Granger-cause the first block of time series. Hence, tests of Granger non-causality can be easily constructed from Wald tests over the estimated coefficients in the upper triangular block of the cepstral power series.

From an algorithmic standpoint, one is required to generate powers of the matrix polynomial $\Upsilon(z)$ (or $\Xi(z)$) and read off the appropriate coefficients. The product of two matrix polynomials is easily encoded; the resulting matrix polynomial has coefficients given by the convolution of the coefficient matrices, respecting the order of the product. These programs have been coded in R (R Development Core Team, 2014) and are available upon request; consequently, the computations for the MA coefficients are straightforward.

3. VEXP Modeling of Vector Time Series

Suppose that we have an observed series of length T from a mean-zero m-variate time series $\{X_t\}$, that we wish to model via a VEXP(q) process. Whether estimation proceeds through maximum likelihood or through a Bayesian approach, typically, q is chosen via some model selection criteria (e.g., AIC, BIC, DIC, or Bayes factor), order selection (e.g., forward selection or backward deletion), or to minimize out-of-sample prediction. Then, given q, we postulate that the MA representation can be modeled via (2.2) such that the spectral density can be expressed as (2.3). To distinguish the model spectrum from the true spectral density of the process $\{X_t\}$, we refer to the latter spectrum as $\widetilde{f}$ and the former spectrum as f_{ϖ}, where $\varpi = \text{vec}\{\text{lotri}[\Omega_0], \Omega_1, \ldots, \Omega_q\}$. (Here lotri refers to the diagonal and lower triangular entries.) Apart from the mean of the series

(for the non-zero mean case), ϖ completely parametrizes the process. The case of a non-zero mean is readily handled; see Section 3.2.

3.1. Likelihood estimation

The Gaussian likelihood is an appealing objective function, because maximum likelihood estimates (MLEs) have the properties of consistency and efficiency (a precise discussion is provided in the Supplemental Appendix – see Theorems 2, 3, and 4 of Appendix S4). Writing $\underline{X} = \text{vec}\{X_1, X_2, \ldots, X_T\}$ and $\underline{\Gamma}_\varpi$ for the mT dimensional covariance matrix of the sample, the log Gaussian likelihood for a mean-zero sample, scaled by -2 (sometimes called the deviance) is

$$\mathcal{D}(\varpi; \underline{X}) = \log\det \underline{\Gamma}_\varpi + \underline{X}' \, \underline{\Gamma}_\varpi^{-1} \, \underline{X}, \tag{3.1}$$

which one seeks to minimize. Efficient computation of the quadratic form and log determinant in (3.1) can proceed utilizing the multivariate Durbin-Levinson algorithm described in Brockwell and Davis (1991); knowledge of the autocovariance function (acvf) – once this is computed from the model with parameter vector ϖ – determines $\underline{\Gamma}_\varpi$ and thereby the deviance. The entries of each cepstral matrix Ω_j are unconstrained (although Ω_0 is symmetric), being allowed to be any real number. If any estimated coefficient, or component of ϖ, is not significantly different from zero, the model could be re-estimated with all such coefficients (or some subset) constrained to be zero, in order to obtain a more parsimonious model. One could also attempt to refine the choice of q utilizing such a procedure.

In order to refine the model of order q or to impose additional sparsity, it is important to have the standard errors of the parameter estimates. Corollary 4 of Appendix S4 in the Supplement establishes efficiency of MLEs, with asymptotic covariance matrix estimated by T times the inverse of the numerical Hessian, i.e., the standard errors are obtained directly from the inverse Hessian, with no sample size correction. The corollary also discusses the log likelihood ratio test, which can be used to discern between nested VEXP models; setting any of the $\binom{m+1}{2} + m^2 q$ parameters of a VEXP(q) to zero yields a nested model, and the usual χ^2 distribution can be applied to the difference in deviances.

There may be interest in using other objective functions. In particular, because there is some computational cost associated with the inversion of $\underline{\Gamma}_\varpi$, an approximate version of the deviance, known as the Whittle likelihood, may be preferable for very large sample sizes. In this case, one replaces the inverse of $\underline{\Gamma}_\varpi$ by the covariance matrix corresponding to the inverse autocovariances. The inverse autocovariances are the autocovariances corresponding to $f_\varpi^{-1} = f_{-\varpi}$: we obtain the inverse of the VEXP spectrum by considering the alternative VEXP process where each coefficient is multiplied by negative one. This is true, because $[\Omega]_1^q(z)$

commutes with $-[\Omega]_1^q(z)$ for any value of z, and $\exp(A) \cdot \exp(B) = \exp(A+B)$ when matrices A and B commute. The expression for the Whittle likelihood given in Taniguchi and Kakizawa (2000) also replaces the log determinant term by the log of the determinant of the innovation variance matrix, $\log\det\exp\{\Omega_0\}$. By (S1.1) of Supplemental Appendix S1, this is equal to the trace of Ω_0. Therefore, for the mean-zero case, the deviance of the Whittle likelihood can be written as

$$\mathcal{W}(\varpi; \underline{X}) = \mathrm{tr}(\Omega_0) + T^{-1}\, \underline{X}'\, \underline{\Gamma}_{-\varpi}\, \underline{X}, \tag{3.2}$$

which is to be minimized with respect to ϖ. While for some time series models (such as unobserved components models) the inverse autocovariances are time-consuming to calculate, they are immediate in the case of a VEXP, given that we have already computed the autocovariances, the insertion of a minus sign in the algorithm is all that is needed. So (3.2) implies a speedier algorithm, as no matrix inversion is required.

Although mathematically equal, the form of the Whittle likelihood in Taniguchi and Kakizawa (2000) is slightly different from (3.2). This other expression involves the integral over all frequencies $\lambda \in [-\pi, \pi]$ of the trace of the periodogram multiplied by f_ϖ^{-1}; straightforward algebra yields that this integral is equal to the quadratic form $T^{-1}\underline{X}'\,\underline{\Gamma}_{-\varpi}\,\underline{X}$ in (3.2). The periodogram is defined to be

$$I_T(\lambda) = T^{-1} \left(\sum_{t=1}^{T} X_t e^{-i\lambda t} \right) \left(\sum_{t=1}^{T} X_t' e^{i\lambda t} \right),$$

which is a rank one matrix. Some statisticians write the Whittle likelihood in terms of the periodogram only being evaluated at Fourier frequencies, which amounts to discretizing the integral in the exact Whittle likelihood by a Riemann approximation. The advantage of doing this further approximation is that the objective function is then expressed purely in terms of the periodogram and the model spectral density, and no calculation of inverse autocovariances is required at all. This approximate Whittle likelihood can be written

$$\mathcal{W}_T(\varpi; \underline{X}) = \mathrm{tr}(\Omega_0) + \frac{1}{2T} \sum_{j=-T}^{T} \mathrm{tr}\{I_T(\tfrac{\pi j}{T})\, f_{-\varpi}(\tfrac{\pi j}{T})\}. \tag{3.3}$$

Once the periodogram is computed, evaluation of (3.3) is extremely fast: one only needs to evaluate $[-\Omega]_1^q(z)$ for z corresponding to the Fourier frequencies, and determine the matrix exponential (for example, via Taylor series directly) and construct $f_{-\varpi}$ via (2.3). If computation of the autocovariances is prohibitively expensive (due to large m and/or q) then the approximate Whittle likelihood may

be preferable. Asymptotic results for Whittle estimates (WLEs) are discussed in Theorem 2 and Corollary 2 of Appendix S4 in the Supplement.

3.2. Bayesian estimation

For an exact Bayesian analysis, the formal procedure for a Gaussian VEXP(q) model requires an exact expression for the likelihood. Although it is possible to pose an approximate Bayesian procedure based on the Whittle (or approximate Whittle) likelihood formulation, in moderate sample sizes our preference is for an exact Bayesian approach. Nevertheless, in large sample sizes, and/or analyses consisting of a large number of time series, an approximate Bayesian procedure may be preferred due to the fact that implementation using an exact or approximate Whittle specification is extremely computationally efficient.

Depending on the desired goals of a particular analysis, it is often advantageous to treat the elements of the cepstral matrices as nuisance parameters and average over different model specifications using stochastic search variable selection (SSVS) (George and McCulloch (1993, 1997); George, Sun and Ni (2008)). This type of Bayesian model averaging (Hoeting et al. (1999)) implicitly weights the elements of the cepstral matrices through the MCMC sampling algorithm. This strategy is extremely effective in the context of forecasting (Holan et al. (2012)), where interest resides in a target other than the cepstral matrix elements. If, instead, the main goal is inferential, then a model corresponding to the posterior mode for each cepstral matrix, from the SSVS algorithm, could be re-estimated. Alternatively, models could be considered without SSVS (e.g., with order selection proceeding through Bayes factor or DIC).

To implement the SSVS algorithm we begin by assuming that the likelihood of $\underline{Y} = (\underline{X} - \mu)'$ is specified as

$$L(\delta, \varpi|\cdot) \propto |\Gamma_\varpi|^{-1/2} \exp\left(-\frac{1}{2}\underline{Y}'\Gamma_\varpi^{-1}\underline{Y}\right),$$

where $\delta = (\mu_1, \mu_2, \ldots, \mu_m)'$ and $\mu = 1_T \otimes \delta$. Non-constant μ could also be considered through straightforward modification of the Markov chain Monte Carlo (MCMC) algorithm (e.g., μ can be specified in terms of covariates). We further assume that the elements of δ and the diagonal elements Ω_0 are in the model with probability one. For the other elements of Ω_0 and Ω_j ($j = 1, \ldots, q$), we specify a SSVS prior based on a mixture of normal distributions.

Let γ_i, $i = 1, \ldots, p = m_0 + qm^2$, denote a latent zero-one random variable, where $m_0 = m(m-1)/2$. Further, let $V_0 = (V_{01}, \ldots, V_{0m_0})$ be the vector of length m_0 containing the unique off-diagonal elements of Ω_0, $V_j = \text{vec}(\Omega_j)$ ($j = 1, \ldots, q$), and $V = (V_0', V_1', \ldots, V_q')'$. Then, $V = (V_{01}, \ldots, V_{0m_0}, V_{11}, V_{12}, \ldots, V_{1m^2}, \ldots, V_{qm^2})'$ $= (v_1, \ldots, v_p)'$ and we have

$$v_i|\gamma_i \sim (1 - \gamma_i)N(0, \tau_i^2) + \gamma_i N(0, c_i^2\tau_i^2),$$

with $P(\gamma_i = 1) = 1 - P(\gamma_i = 0) = \pi_i$. Thus, $V|\gamma \sim N(\mathbf{0}, D_\gamma R D_\gamma)$, where R is the prior correlation matrix – which in our case we assume to be the identity matrix – and $D_\gamma \equiv \text{diag}(a_1\tau_1, \ldots, a_p\tau_p)$ with $p = m_0 + qm^2$. In this case, for $i = 1, \ldots, p$, $\gamma_i \overset{iid}{\sim} \text{Bern}(\pi_i)$, with $\pi_i \equiv 1/2$ and $a_i = 1$ if $\gamma_i = 0$, and $a_i = c_i$ if $\gamma_i = 1$.

Here π_i can be viewed as the prior probability that the i-th element of V should be included in the model. Therefore, $\gamma_i = 1$ indicates that the i-th variable is included in the model. Now, in general, c_i, τ_i, and π_i are fixed hyperparameters; George and McCulloch (1993, 1997) describe various alternatives for their specification. They suggest that one would like τ_i to be small so that when $\gamma_i = 0$ it is reasonable to specify an effective prior for the i-th element of V that is near zero. Additionally, one typically wants c_i to be large (greater than 1) so that if $\gamma_i = 1$, our prior would favor a non-zero value for the i-th element of V.

To complete the Bayesian model, we need to specify prior distributions for the remaining parameters. In terms of the mean, we assume that $\delta \sim N(\delta_0, \Sigma_\delta)$, with $\delta_0 = (u_1, \ldots, u_m)'$ and $\Sigma_\delta = \text{diag}(\sigma^2_{\mu_1}, \sigma^2_{\mu_2}, \ldots, \sigma^2_{\mu_m})$. Recall, the diagonal elements of Ω_0 are assumed to be in the model with probability one and, thus, we assume that $\Omega_0 \sim N(0, \Sigma_{\Omega_0})$, where $\Sigma_{\Omega_0} = \text{diag}(\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_m)$. Lastly, for $(j = 1, \ldots, m)$, $\sigma^2_{\mu_j} \sim \text{IG}(A_{\mu_j}, B_{\mu_j})$ and $\sigma^2_j \sim \text{IG}(A_j, B_j)$.

In high-dimensional cases (e.g., $m \geq 10$ and/or large q) direct implementation of the SSVS algorithm may be slow and may not search the entire model space. In these cases, it may be advantageous to do some form of dimension reduction prior to specifying the VEXP model and carrying out Bayesian variable selection (e.g., see Wikle and Holan (2011); Holan et al. (2012); Yang et al. (2013), among others). Similarly, in high-dimensional settings, doing some form of initial dimension reduction may help facilitate prior selection in the SSVS algorithm.

In some cases it may be of interest to estimate the VEXP model without conducting Bayesian model averaging through SSVS, as is the case in the example we present involving squared coherence estimation (see Section 4.2). Under this scenario, a prior distribution for the elements of V_0 and V_j $(j = 1, \ldots, q)$ needs to be specified. Here, for $k' = 1, \ldots, m_0$, we assume that $V_{0k'} \sim N(0, \sigma^2_{0k'})$ and, for $j = 1, \ldots, q$ and $k = 1, \ldots, m$, we assume that $V_{jk} \sim N(0, \sigma^2_{jk})$.

In general, regardless of whether a SSVS prior is implemented, the full conditional distributions are not of standard form, with the only exceptions being δ and the elements of Σ_δ and Σ_{Ω_0}. Consequently, all of the parameters aside from δ and the elements of Σ_δ and Σ_{Ω_0} can be sampled using a random walk Metroplis-Hastings within Gibbs MCMC sampling algorithm. Sampling of δ and the elements of Σ_δ and Σ_{Ω_0} proceeds directly using a Gibbs step, as the full conditionals distribution have a closed form.

4. VEXP Modeling Illustrations

To demonstrate the versatility and overall utility of the VEXP modeling framework, we present two data examples. The first considers multi-step ahead forecasts for a bivariate macroeconomic time series and uses Bayesian model averaging through SSVS as a means of obtaining superior forecasts. The second example examines the squared coherence between monthly sunspots and critical radio frequencies and does not make use of SSVS. Instead, the goal of this analysis is to demonstrate the VEXP approach to multivariate spectral (squared coherence) estimation.

4.1. Multi-step ahead forecasting

Multi-step ahead forecasting is an area of considerable interest in such disciplines as atmospheric science and macroeconomics, among others. One paramount concern when constructing long-lead forecasts is to ensure the model specification is not explosive. In the context of VAR modeling (or VARMA) this can be facilitated through imposing restrictions on the coefficient matrices to ensure that certain determinants are nonzero. In contrast, our method provides an extremely convenient approach to model specification that does not require us to impose any constraints a priori, making estimation exceedingly straightforward within the Bayesian paradigm.

The macroeconomic time series we consider are regression adjusted ("Holiday" and "Trading Day" effects are removed) monthly retail sales time series from the U.S. Census Bureau. Specifically, we consider a bivariate analysis of (entire) "Retail Trade Sector" (RTS) and "Automotive Parts, Accessories, and Tire Stores" (APATS) from January 1992 through December 2007, $T = 192$ (Figure 1). McElroy and McCracken (2014) describe multi-step forecasting for nonstationary vector time series with a general MA form. From that work, the forecast filter (from an infinite past) for h-step ahead forecasting for a stationary process with invertible MA power series $\Psi(z)$ is given by $\Pi(z) = z^{-h}[\Psi]_h^{\infty}(z)\,\Psi^{-1}(z)$.

The bivariate time series considered here are annual-differenced (the operator $(1 - B^{12})$ is applied to the data) prior to estimation using the VEXP model. The cross-correlation function of the differenced series indicates that they are cross-correlated (Figure 2). From the acf (not shown), it appears that an AR(1) or low-order ARMA model may be reasonable for each series; although some serial correlation at lag 12 is indicated (which one would expect, given the crudeness of seasonal adjustment achieved by the action of $1 - B^{12}$), most of the dependence is concentrated in the low lags. Therefore, we focus upon low order VEXP models for our forecasting application; while we do not claim that our VEXP(5) specification is the best possible for the data, our subsequent forecasting results indicate that this specification is nonetheless adequate for our purposes.

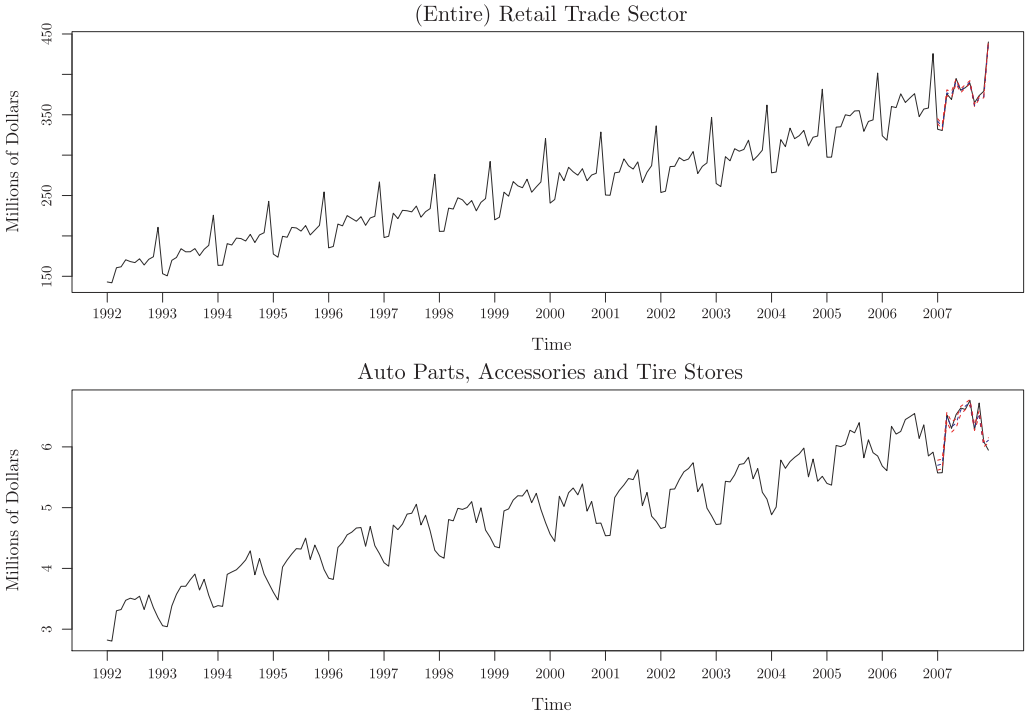


Figure 1. Pointwise multi-step ahead (12-steps ahead) forecast plots for the (entire) "Retail Trade Sector" (RTS) and "Automotive Parts, Accessories, and Tire Stores" (APATS) series (Section 4.1). The dashed lines and the additional solid line (beginning January 2007) denote the pointwise 95% credible intervals and the posterior mean, respectively, for the distribution of forecasts at each time.

Prior specification was identical to that of Simulated Example I (Supplemental Appendix S3), except in this example we conducted a factorial experiment over various combinations of the SSVS hyperparameters and chose the combination of c and τ that minimized the out-of-sample mean squared prediction error (MSPE) over the last 12 values of the series. Based on previous forecasting analyses (Holan et al. (2012)), for $q = 2, \ldots, 5$, the SSVS hyperparameters considered for this experiment were $\pi_i \equiv \pi = 0.5$ and $(\tau, c) = (0.001, 10), (0.001, 100), (0.01, 10), (0.01, 100), (0.1, 10), (0.1, 100)$. Judging from the overall MSPE, the parameters $q = 5$, $\tau = 0.1$, and $c = 10$ gave the best performance among the values considered. Although these hyperparameters could be tuned further, our experience is that such tuning leads to minimal gains in forecasting monthly retail sales. The SSVS sampler results were based on 60,000 iterations with a 40,000 iteration burn-in. Convergence was assessed through visual inspection of the sample chains, with no lack of convergence detected. When conducting out-of-sample

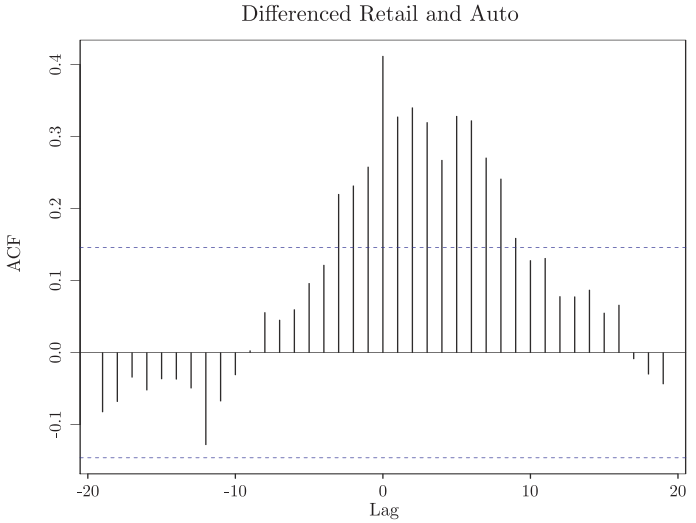


Figure 2. Cross-correlation function plot for the regression-adjusted and annual-differenced (entire) "Retail Trade Sector" (RTS) and "Automotive Parts, Accessories, and Tire Stores" (APATS) time series (Section 4.1).

forecasting, our h-step-ahead forecasts were based on the posterior mean forecast h steps ahead over all iterations of the SSVS MCMC run. Similar to Holan et al. (2012), this forecasting represents a "model averaging" over all possible elements of the cepstral matrices, and accounts for their relative importance through the SSVS procedure. Specifically, our approach produces a posterior distribution of forecasts rather than associating the forecasts with a single estimated model, as would be the case in OLS estimation (or maximum likelihood estimation). This can similarly be done using a Bayesian VAR(p) (or VARMA) formulation. However, in the Bayesian VAR case problems of identifying the stationary region persist and may lead to an unstable representation.

Using the VEXP model, the MSPE for this example was 21.59 and 0.0112 for the RTS and APATS series, respectively. In contrast, conducting the same experiment with a VAR(1) model, estimated through OLS, yielded a MSPE of 22.95 and 0.0244 for the RTS and APATS series respectively. Therefore, relative to the OLS VAR(1) model, the Bayesian VEXP(5) model provides roughly a 6% decrease in MSPE for the RTS series and a 55% decrease for the APATS series. Figure 1 displays the forecasted values along with their pointwise 95% CIs and clearly demonstrates the effectiveness of our approach. These results are not unexpected. In general, model-averaged forecasts typically out-perform forecasts from a single model. In contrast, the single VAR(1) model is easy to fit, but likely misspecified. Ultimately, this example illustrates the ease with which

Bayesian model averaging can be done using the VEXP model specification.

4.2. Modeling of squared coherence

This example considers bivariate spectral estimation and, in particular, estimation of the squared coherence

$$\rho^2_{X_1 \cdot X_2}(\lambda) = \frac{|f_{X_1X_2}(\lambda)|^2}{f_{X_1X_1}(\lambda) f_{X_2X_2}(\lambda)}.$$

Specifically, we consider a monthly bivariate time series of critical radio frequencies and sunspots (Newton (1988)). The first series consists of the monthly median noon-hour value of the critical radio frequencies (the highest radio frequency that can be used for broadcasting) in Washington D.C. for the period of May 1934 through April 1954. The second series consists of the total number of monthly sunspots over the same period.

For this illustration, we considered a VEXP(5) model without SSVS. Prior specification was identical to that of Simulated Example II (Supplemental Appendix S3). The MCMC sampling algorithm consisted of 60,000 iterations with a 40,000 iteration burn-in. Convergence was assessed through visual inspection of the sample chains, with no lack of convergence detected. The spectral estimates and squared coherence were obtained by taking the pointwise (by frequency) posterior mean and 95% CIs.

The coherence between the two series is illustrated through a plot of the squared coherence (Figure 3a). From this plot, we see that there is strong coherence at low frequencies (i.e., $\lambda \approx \pi/66$) corresponding to the so-called sunspot cycle (≈ 11 years) Additionally, there is also relatively strong coherence at higher frequencies (i.e., $\lambda \approx 2.5$), which may correspond to some sort of (approximately quarterly) seasonal relationship. These relationships are corroborated through an empirical plot of the squared coherence using a modified Daniell window in R (using kernel ("modified.daniell", c(8,8,8)) with taper=.2 in the function spec.pgram); see Figure 3b. Discussion regarding these series in the univariate setting can be found in Newton (1988, p. 194).

5. Conclusion

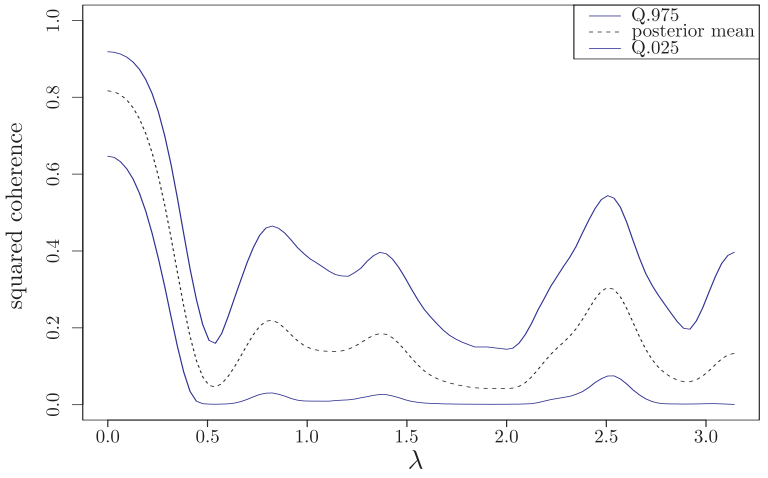
We propose a new class of cepstral models for multivariate time series – the VEXP. Conveniently, this model is cast in the frequency domain and has an unrestricted parameter space. This is in stark contrast to the VARMA modeling paradigm where one must impose restrictions on the coefficient matrices to ensure that certain determinants are nonzero in order to ensure the model is stationary (or nonexplosive).

We provide theoretical justification for this new class of models and show that this model is dense in the class of short memory time series. Additionally, for $q < \infty$, we show that the VEXP(q) process is always stable, invertible, and identifiable. Importantly, we derive the necessary computational formulas for efficient model implementation and discuss several approaches to estimation, including maximum likelihood and Bayesian estimation. In fact, one of the primary strengths of the VEXP class of models is that a precise Bayesian treatment proceeds naturally.

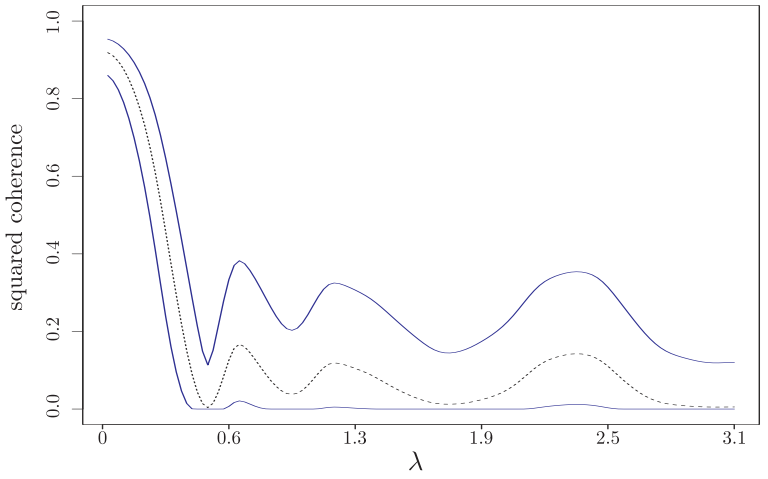
Similar to other multivariate time series models, issues regarding the dimension of the parameter space remain an area of concern. In cases where the model order and/or the number of series is substantially large, the number of parameters in the model causes difficulty in estimation. In these cases, further dimension reduction of the cepstral matrices is advantageous and can be achieved through low-rank methods or scientifically motivated parameterizations (e.g., see Cressie and Wikle (2011)). Another practical consideration concerns the number of MA coefficients, M, used for estimation, which needs to be chosen by the practitioner. In order to guarantee a sufficient approximation, this choice would depend on the underlying dependence structure. In the short memory cases considered in Section 4 and Supplemental Appendix S3, we have taken $M = 15$, and found that to be sufficient for our intended purpose.

The methodology is illustrated through multi-step ahead forecasting of bivariate retail trade series from the U.S. Census Bureau and through estimation of squared coherence for a bivariate time series of monthly sunspots and critical radio frequencies. The forecasting example uses SSVS and thus provides an implicit model averaging. We demonstrate the superiority of our approach, in terms of MSPE for multi-step-ahead forecasting, relative to an OLS VAR(1) model. Figures 1 and 2, respectively, illustrated the effectiveness of our approach and the need for a multivariate model. In contrast, the squared coherence example does not impose SSVS and, instead, illustrates various aspects concerning spectral estimation. Specifically, Figure 3 highlights our ability to capture cross-dependence in the spectral domain for correlated bivariate time series. Our results corroborate those of previous analyses, while providing a straightforward path to parametric squared coherence estimation.

Many other applications of the VEXP model exist. For example, multivariate long memory modeling and multivariate unobserved component models using the VEXP framework are areas of open research. In summary, any multivariate short memory time series application can be posed using the VEXP framework, thereby providing a rich class of models for multivariate time series.



(a)



(b)

Figure 3. (a) Pointwise posterior mean squared coherence plot between the critical-radio frequency - sunspots time series corresponding to the period May 1934 through April 1954. (b) Empirical squared coherence and pointwise 95% confidence intervals using modified Daniell window; see Section 4.2.

Supplementary Materials

The Supplemental Appendix contains proofs of technical results, a description of applications of the VEXP model, results of two simulated examples, and asymptotic theory for parameter estimates.

Acknowledgements

We thank the Editor, an associate editor, and two anonymous referees for providing valuable comments that have helped strengthen this manuscript. This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program. This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

References

Artin, M. (1991). *Algebra*. Prentice Hall, Englewood Cliffs, New Jersey.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217-226.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.

Chiu, T. Y., Leonard, T. and Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *J. Amer. Statist. Assoc.* **91**, 198-210.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, New Jersey.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica* **55**, 251-276.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88**, 881-889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7**, 339-373.

George, E. I., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *J. Econom.* **142**, 553-580.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.

Hannan, E.J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. SIAM, Pennsylvania.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14**, 382-401.

Holan, S. H. (2004). Time series exponential models: Theory and methods. Ph.D. Thesis, Texas A&M University.

Holan, S. H., McElroy, T. and Chakraborty, S. (2009). A Bayesian approach to estimating the long memory parameter. *Bayesian Analysis* **4**, 159-190.

Holan, S. H., McElroy, T. S. and Wu, G. (2014). The cepstral model for multivariate time series: The vector exponential model. arXiv preprint, arXiv:1406.0801.

Holan, S. H., Yang, W.-H., Matteson, D. S. and Wikle, C. K. (2012). An approach for identifying and predicting economic recessions in real-time using time–frequency functional models. *Appl. Stochastic Models in Business and Industry* **28**, 485-499.

Hurvich, C. M. (2002). Multistep forecasting of long memory series using fractional exponential models. *Internat. J. Forecasting* **18**, 167-179.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

McElroy, T. S. and Holan, S. H. (2012). On the computation of autocovariances for generalized Gegenbauer processes. *Statist. Sinica* **22**, 1661-1687.

McElroy, T. S. and Holan, S. H. (2014). Asymptotic theory of cepstral random fields. *Ann. Statist.* **42**, 64-86.

McElroy, T. S. and McCracken, M. W. (2014). Multi-step ahead forecasting of vector time series. *Econometric Reviews*. Published Online.

McElroy, T. and Trimbur, T. (2015). Signal extraction for non-stationary multivariate time series with illustrations for trend inflation. *J. Time Series Anal.* **36**, 209-227.

Newton, H. J. (1988). *Timeslab: A Time Series Analysis Laboratory*. Wadsworth & Brooks/Cole Publishing, Pacific Grove, CA.

Nyblom, J. and Harvey, A. (2000). Tests of common stochastic trends. *Econom. Theory* **16**, 176-199.

Pourahmadi, M. (1984). Taylor expansion of $\exp\{\sum_{k=0}^{\infty} a_k z^k\}$ and some applications. *Amer. Math. Monthly* **91**, 303-307.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Stanley, R. P. (1997). *Enumerative Combinatorics*. Cambridge University Press, Cambridge.

Stock, J.H. and Watson, M.W. (1988). Testing for common trends. *J. Amer. Statist. Assoc.* **83**, 1097-1107.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Taniguchi, M., Puri, M. L. and Kondo, M. (1996). Nonparametric approach for non- Gaussian vector stationary processes. *J. Multivariate Anal.* **56**, 259-283.

Wikle, C. K. and Holan, S. H. (2011). Polynomial nonlinear spatio-temporal integro-difference equation models. *J. Time Series Anal.* **32**, 339-350.

Yang, W.-H., Wikle, C. K., Holan, S. H. and Wildhaber, M. L. (2013). Ecological pre- diction with nonlinear multivariate time-frequency functional data models. *J. Agricultural, Biological, and Environmental Statistics* **18**, 450-474.

Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, USA.

E-mail: holans@missouri.edu

Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, USA.

E-mail: tucker.s.mcelroy@census.gov

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA.

E-mail: raywu2014@gmail.com

(Received August 2014; accepted December 2015)