arXiv:1107.5239v1 [stat.ME] 26 Jul 2011

AUTOREGRESSIVE MODELS FOR VARIANCE MATRICES: STATIONARY INVERSE WISHART PROCESSES

BY EMILY B. FOX* AND MIKE WEST†

Duke University, Durham NC, USA

We introduce and explore a new class of stationary time series models for variance matrices based on a constructive definition exploiting inverse Wishart distribution theory. The main class of models explored is a novel class of stationary, first-order autoregressive (AR) processes on the cone of positive semi-definite matrices. Aspects of the theory and structure of these new models for multivariate "volatility" processes are described in detail and exemplified. We then develop approaches to model fitting via Bayesian simulation-based computations, creating a custom filtering method that relies on an efficient innovations sampler. An example is then provided in analysis of a multivariate electroencephalogram (EEG) time series in neurological studies. We conclude by discussing potential further developments of higher-order AR models and a number of connections with prior approaches.

1. Introduction. Modeling the temporal dependence structure in a sequence of variance matrices is of increasing interest in multi- and matrix-variate time series analysis, with motivating applications in fields as diverse as econometrics, neuroscience, epidemiology and spatial-temporal modeling. Some key interests and needs are in defining: (i) classes of stationary stochastic process models on the cone of symmetric, non-negative definite matrices that offer flexibility to model differing degrees of dependence structures as well as short-term predictive ability; (ii) models that are open to theoretical study and interpretation; and (iii) models generating some degree of analytic and computational tractability for model fitting and exploitation in applied work.

The context is a sequence of $q \times q$ variance matrices (i.e., symmetric, non-negative definite matrices) Σ_t in discrete time $t = 0, 1, \ldots,$ typically the variance matrices of components of more elaborate state-space models for an observable

*Research partially supported by a Mathematical Sciences Postdoctoral Research Fellowship from the National Science Foundation

†Research partially supported by the National Science Foundation under grants DMS-0342172 and DMS-1106516. Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

AMS 2000 subject classifications: Primary 62M10; secondary 37M10,62H99

Keywords and phrases: Autoregressive models, Bayesian forecasting, Innovations sampling, Matrix-Variate Autoregressions, Multiple time series analysis, Multivariate stochastic volatility, Time-varying variance matrix

time series. In econometrics and finance, variants of "observation-driven" multivariate ARCH (Engle, 2002) models and state-space or "parameter-driven" multivariate stochastic volatility models (Harvey et al., 1994; Quintana and West, 1987) are widely used. While the former directly specify "volatility" matrices Σ_t as functions of lagged values and past data, state-space approaches use formal stochastic process models that offer cleaner interpretation, access to theoretical understanding as well as potential to scale more easily with dimension; see Chib et al. (2009) for a survey of such approaches. The class of state-space models based on Bayesian discount methods (Quintana and West, 1987; Quintana, 1992; Quintana et al., 2003, 2010; Uhlig, 1994; West and Harrison, 1997), are also widely used in financial applications for local volatility estimation and smoothing. These methods are, however, restricted to local estimation due to the underlying non-stationary random-walk style model for Σ_t; see Prado and West (2010) for recent review and additional developments.

Two recent contributions explore constructions of AR(1) style models based on conditional Wishart transition distributions (Gouriéroux et al., 2009; Philipov and Glickman, 2006a,b). These aim to provide flexibility in modeling one-step dependencies balanced with parsimony in parameterization through properties of the Wishart distribution. These models tend to be rather intractable theoretically, hence somewhat difficult to understand and interpret, while model fitting is challenging and there are open questions of how useful potential higher-order variants might be. We discuss these approaches and issues further in Section 8.

The centrality of inverse Wishart theory to current Bayesian state-space approaches underlies the ideas for new model classes explored in this paper. We introduce a class of stationary, non-linear autoregressive (AR) models for variance matrices by exploiting the structure of conditional and marginal distributions in the inverse Wishart family. We denote the resulting models by AR or IW-AR for definiteness, and use AR(1) or IW-AR(1) to be more specific about first-order models when needed; most of the development of this paper is for first-order models. The new IW-AR models are open to some useful theoretical analysis of margins, stationarity, reversibility, and conditional moments, among other properties. Exploiting the state-space nature of the IW-AR(1) process, we develop an MCMC sampler based on forward filtering backward sampling (FFBS) proposals that results in tractable Bayesian computations. This operates locally on a matrix innovations process to ameliorate issues arising from global accept-rejects of the variance matrix process (e.g., exponential decrease in acceptance rates with increasing sequence length) albeit at increased computational cost.

Section 2 introduces the new models and some aspects of the theoretical structure are explored in Section 3. Posterior computations are developed in Section 5, building on a data augmentation idea discussed in Section 4. An example in EEG

time series analysis is given in Section 6 and Section 7 discusses extensions to higher-order AR dependencies. Section 8 discusses connections with other approaches, Section 9 provides summary comments and supporting technical material is appended.

For time ranges we use the concise notation $s : t$ to denote the sequence of time indices $s, s+1, \ldots, t$; e.g., $\Sigma_{0:T} = \{\Sigma_0, \ldots, \Sigma_T\}$ and $\Sigma_{t-1:t} = \{\Sigma_{t-1}, \Sigma_t\}$.

2. First-Order Inverse Wishart Autoregressive Processes.

2.1. *Construction.* As context, suppose we are to observe a series of $q \times 1$ vector observations x_t with

$$x_t|\Sigma_t \sim N(0, \Sigma_t), \qquad t = 1 : T, \tag{2.1}$$

where x_t is independent of $\{x_s, \Sigma_s;\ s < t\}$ conditional on Σ_t. We aim to capture the volatility dynamics with a stationary, first-order Markov model for the Σ_t sequence. The joint density for matrices over an arbitrary time period $t = 0 : T$ is

$$p(\Sigma_{0:T}) = \frac{\prod_{t=1}^{T} p(\Sigma_{t-1}, \Sigma_t)}{\prod_{t=2}^{T} p(\Sigma_{t-1})} \tag{2.2}$$

for some time invariant joint density $p(\Sigma_{t-1}, \Sigma_t)$ for consecutive matrices in the numerator terms; this joint density has common margins given by the time invariant $p(\Sigma_t)$ appearing in the denominator terms.

We take the defining joint density $p(\Sigma_{t-1}, \Sigma_t)$ as arising from an inverse Wishart on an augmented state-space. Specifically, introduce random matrices ϕ_t such that

$$\begin{pmatrix} \Sigma_{t-1} & \phi_t' \\ \phi_t & \Sigma_t \end{pmatrix} \sim \mathrm{IW}_{2q}\left(n+2, n\begin{pmatrix} S & SF' \\ FS & S \end{pmatrix}\right) \tag{2.3}$$

for some degree of freedom parameter $n > 0$, a $q \times q$ variance matrix parameter S and a $q \times q$ matrix parameter F such that the $2q \times 2q$ parameter matrix parameter of the distribution above is non-negative definite. This inverse Wishart has common margin for the diagonal blocks; for each t,

$$\Sigma_t \sim \mathrm{IW}_q(n+2, nS) \tag{2.4}$$

with $E[\Sigma_t] = S$. It is now clear that eqn. (2.2) defines stationary first-order process with eqn. (2.4) as the stationary (marginal) distribution. Transitions are governed by the conditional density $p(\Sigma_t|\Sigma_{t-1})$ implicitly defined by eqn. (2.3). This has no closed analytic form but is now explored theoretically.

2.2. *Innovations Process.* The joint distribution of $\{\Sigma_{t-1}, \Sigma_t, \phi_t\}$ defined in eqn. (2.3) can be reformulated in terms of $\{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$ where $\{\Upsilon_t, \Psi_t\}$ are marginally matrix normal, inverse Wishart distributed and independent of Σ_{t-1}. Specifically, standard inverse Wishart theory (e.g. Carvalho et al., 2007) implies that

$$\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'. \tag{2.5}$$

and $\phi_t = \Upsilon_t \Sigma_{t-1}$ where the $q \times q$ matrices $\{\Upsilon_t, \Psi_t\}$ follow

$$\begin{aligned} \Psi_t &\sim \text{IW}_q(n+q+2, nV) \\ \Upsilon_t \mid \Psi_t &\sim N(F, \Psi_t, (nS)^{-1}) \end{aligned} \tag{2.6}$$

with $V = S - FSF'$ and where $\{\Upsilon_t, \Psi_t\}$ are conditionally independent of $\Sigma_{t-1} \sim \text{IW}_q(n+2, nS)$. Eqn. (2.5) is an explicit AR(1) equation in which Υ_t acts as a random autoregressive coefficient matrix and Ψ_t an additive random disturbance. Since $\{\Upsilon_t, \Psi_t\}$ are independent at each t and drive the dynamics of this IW-AR process, we refer to them as latent *innovations*.

2.3. *Special Case of* $q = 1$. When $q = 1$, $\Sigma_t \equiv \sigma_t > 0$ and the IW-AR process reduces to an inverse gamma autoregressive process. Now $S \equiv s > 0$ and $F \equiv f \in (-1, 1)$ and the joint density of eqn. (2.3) is

$$\begin{pmatrix} \sigma_{t-1} & \phi_t \\ \phi_t & \sigma_t \end{pmatrix} \sim \text{IW}_2\left(n+2, ns\begin{pmatrix} 1 & f \\ f & 1 \end{pmatrix}\right), \tag{2.7}$$

such that

$$\sigma_t \sim IG\left(\frac{n+2}{2}, \frac{ns}{2}\right). \tag{2.8}$$

Equivalently, with scalar innovations $\{\Upsilon_t, \Psi_t\} \equiv \{\upsilon_t, \psi_t\}$,

$$\sigma_t = \psi_t + \upsilon_t^2 \sigma_{t-1} \tag{2.9}$$

where

$$\psi_t \sim IG\left(\frac{n+3}{2}, \frac{ns(1-f^2)}{2}\right) \quad \text{and} \quad \upsilon_t \mid \psi_t \sim N\left(f, \frac{\psi_t}{ns}\right). \tag{2.10}$$

We can see immediate analogies with the standard linear, Gaussian AR(1) process with a random AR coefficient. The marginal mean of υ_t^2 is $(nf^2+1)/(n+1)$ which plays the role of an average linear autoregressive coefficient. For $|f|$ close to 1, the model approaches the stationary/non-stationary boundary, and when n is large, the

mean AR(1) coefficient is close to f^2. Also, $E[\psi_t] = ns(1-f^2)/(n+1)$ so that for fixed n and s the additive innovation noise tends to be smaller as $|f|$ approaches unity. Parameters (n, f) also control dispersion of the additive innovations through, for example, $V(\psi_t) = 2[ns(1-f^2)]^2/[(n+1)^2(n-1)]$. Section 3 further explores this in the general multivariate setting as well as this special case of $q = 1$.

This inverse gamma autoregressive process is related to the formulation of Pitt et al. (2002). In that work, the authors construct a stationary autoregressive process $\{\sigma_t\}$ with inverse gamma marginals by harnessing a conditionally gamma distributed latent process $\{z_t\}$. The sequence $\{\sigma_t, z_t\}$ obtained by generating $\sigma_t \mid z_{t-1}$ and $z_t \mid \sigma_t$ from the respective closed form conditional distributions leads to a marginal process $\{\sigma_t\}$ with the desired autoregressive structure. Extensions to Bayesian nonparametric transition kernels is considered in Mena and Walker (2005) and to state-space volatility processes in Pitt and Walker (2005). Although related in spirit to this work, the proposed IW-AR process represents a novel construction. One attribute of the IW-AR approach, as explored in Section 3, is that our process need not be reversible depending upon the parameterization specified by f and s. Furthermore, our formulation allows straightforward higher-order extensions, discussed in Section 7. Additional discussion and other related approaches appears in Section 8.

3. Theoretical Properties.

3.1. *Marginal Processes for Submatrices and Univariate Elements.* Consider any partition of Σ_t into blocks $\Sigma_{t,ij}$, $(i = 1 : I, j = 1 : J)$, where i, j represent consecutive blocks of consecutive sets of row and column indices, respectively. As a special case this also defines scalar elements. Then the evolution of each submatrix $\Sigma_{t,ij}$ depends upon every element of Σ_{t-1} as follows:

$$\Sigma_{t,ij} = \Psi_{t,ij} + \begin{bmatrix} \Upsilon_{t,i1} & \dots & \Upsilon_{t,iJ} \end{bmatrix} \Sigma_{t-1} \begin{bmatrix} \Upsilon'_{t,1j} \\ \vdots \\ \Upsilon'_{t,Ij} \end{bmatrix}. \tag{3.1}$$

Here $\Psi_{t,ij} \sim \text{IW}(n+q+2, nV_{ij})$ and $\begin{bmatrix} \Upsilon_{t,i1} & \dots & \Upsilon_{t,iJ} \end{bmatrix}$ has a conditional matrix normal distribution induced from the joint distribution of eqn. (2.6).

3.2. *Stationarity.*

THEOREM 3.1. *The process defined via eqn. (2.3) is strictly stationary when the parameterization of the inverse Wishart of yields a valid distribution: that is, when S and $S - FSF'$ are positive definite.*

PROOF. This follows directly from the constructive definition using eqn. (2.3). For a valid model, the scale matrix must be positive definite. Equivalently, via Sylvester's criterion and the Schur complement, S and $S - FSF'$ must be positive definite. □

We note extensions to non-negative definite cases when the resulting Σ_t matrices are singular with singular inverse Wishart distributions, although these are of limited practical interest so we focus on non-singular cases throughout.

In simple cases of $F = \rho I_q$ the stationarity condition reduces to $|S| > 0$ and $|\rho| < 1$. Other special cases are those in which F, S share eigenvectors with eigendecompositions $F = ERE'$ and $S = EQE'$. Stationarity is assured when $||Q||_0 > 0$ and $||R||_0 < 1$.

3.3. *Reversibility.*

THEOREM 3.2. *The process is time-reversible if and only if $FS = SF'$.*

PROOF. The reverse-time process on the Σ_t is as follows. Eqn. (2.3) implies that

$$\begin{pmatrix} \Sigma_t & \tilde{\phi}_t' \\ \tilde{\phi}_t & \Sigma_{t-1} \end{pmatrix} \sim \text{IW}_{2q}\left(n+2, n\begin{pmatrix} S & FS \\ SF' & S \end{pmatrix}\right), \tag{3.2}$$

for some latent process $\tilde{\phi}_t$. Then, as in Section 2, we have

$$\Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Upsilon}_t \Sigma_t \tilde{\Upsilon}_t' \tag{3.3}$$

where

$$\tilde{\Psi}_t \sim \text{IW}_q(n+q+2, n\tilde{V}) \quad \text{and} \quad \tilde{\Upsilon}_t \mid \tilde{\Psi}_t \sim N(\tilde{F}, \tilde{\Psi}_t, (nS)^{-1}) \tag{3.4}$$

with $\tilde{V} = S - SF'S^{-1}FS$ and $\tilde{F} = SF'S^{-1}$. If $FS = SF'$, then $\tilde{V} = V$ and $\tilde{F} = F$ and the reverse-time process follows the same model as the forward-time process. Conversely, assume a reversible process (i.e., $\tilde{F} = F$ and $\tilde{V} = V$) with $FS \neq SF'$. Since $\tilde{F} = SF'S^{-1}$, a contradiction immediately arises. □

Examples of reversible IW-AR processes include cases when $F = \rho I_q$ or when $F = ERE'$ and $S = EQE'$. Note, however, that the process is irreversible when $F = \text{diag}(\rho_1, \ldots, \rho_q)$ with distinct elements and S is non-diagonal.

3.4. *Conditional Mean.* The IW-AR yields a simple form for the conditional expectation of Σ_t given Σ_{t-1}.

THEOREM 3.3.

$$E[\Sigma_t \mid \Sigma_{t-1}] = F\Sigma_{t-1}F' + c_{n,q}\left[1 + \mathrm{tr}\{\Sigma_{t-1}(nS)^{-1}\}\right] V \tag{3.5}$$

with $c_{n,q} = n/(n+q)$.

PROOF. See Appendix. □

Theorem 3.3 illuminates the inherent matrix linearity of the model and the interpretation of F as a "square root" AR parameter matrix. The conditional mean regression form is $F\Sigma_{t-1}F'$ corrected by a term that reflects the skewness of the conditional distribution. For large n, the underlying inverse Wishart distributions are less skewed and this latter term is small; indeed

$$\lim_{n\to\infty} E[\Sigma_t \mid \Sigma_{t-1}] = S + F(\Sigma_{t-1} - S)F'. \tag{3.6}$$

3.5. *Principal Component IW-AR Processes.* Assume that F and S share eigenvectors so that the IW-AR model is reversible. There exists a principal component IW-AR process, as follows.

THEOREM 3.4. *Suppose that* $F = ERE'$ *and* $S = EQE'$ *where* E *is orthogonal and* $R = \mathrm{diag}(\rho_1, \ldots, \rho_q)$ *and* $Q = \mathrm{diag}(\xi_1, \ldots, \xi_q)$ *with positive elements. Then the sequence of matrices defined by* $\hat{\Sigma}_t = E'\Sigma_t E$ *follows an IW-AR(1) model with degrees of freedom* n, *scale matrix* Q *and AR matrix* R. *Specifically,*

$$\hat{\Sigma}_t = \hat{\Psi}_t + \hat{\Upsilon}_t\hat{\Sigma}_{t-1}\hat{\Upsilon}_t'$$

where $\hat{\Psi}_t = E'\Psi_t E$ *and* $\hat{\Upsilon}_t = E'\Upsilon_t E$ *are such that*

$$\hat{\Psi}_t \sim IW_q(n+q+2, nQ(I-R^2)) \quad \textit{and} \quad \hat{\Upsilon}_t \mid \hat{\Psi}_t \sim N(R, \hat{\Psi}_t, (nQ)^{-1})$$

and with marginal distribution $\hat{\Sigma}_t \sim IW_q(n+2, nQ)$.

The conditional moment is given by

$$E[\hat{\Sigma}_t \mid \hat{\Sigma}_{t-1}] = R\hat{\Sigma}_{t-1}R + c_{n,q}\left[1 + \sum_i (n\xi_i)^{-1}\hat{\Sigma}_{t-1,ii}\right] Q(I-R^2). \tag{3.7}$$

eqn. 3.7 implies that there exists a zero-mean noise process $v_{t,j}$ *such that*

$$\hat{\Sigma}_{t,jj} = \rho_j^2\hat{\Sigma}_{t-1,jj} + c_{n,q}\left[1 + \sum_i (n\xi_i)^{-1}\hat{\Sigma}_{t-1,ii}\right](1-\rho_j^2)\xi_j + v_{t,j}. \tag{3.8}$$

Autoregressive processes for the other terms of $\hat{\Sigma}_t$ *are similarly defined.*

PROOF. See Appendix. □

3.6. *Exponential Forgetting.* It is also interesting to examine properties of the IW-AR process as a function of the parameters F, S, and n and an initial value Σ_0. The mean of the IW-AR process forgets its initial condition exponentially fast under a wide range of conditions on the parameterization.

THEOREM 3.5. *Assuming that* $||S||_\infty$, $||S^{-1}||_\infty$, *and* $||F||_\infty$ *are each bounded by some finite* $\lambda \in \mathbb{R}$, *then*

$$E[\Sigma_t \mid \Sigma_0] = F^t \Sigma_0 F'^t + c_{n,q}(S - F^t S F'^t) + O(n^{-1} 1 \cdot 1') \tag{3.9}$$

where $c_{n,q} = n/(n+q)$ *and* 1 *denotes a column vector of ones. Further,*

$$\lim_{n\to\infty} E[\Sigma_t \mid \Sigma_0] = S + F^t(\Sigma_0 - S)F'^t. \tag{3.10}$$

PROOF. See Appendix. □

From eqn. (3.10), based on fixed S and in the limit as $n \to \infty$, we can directly analyze the effects of the elements of F on the conditional mean.

Unitary Bounded Spectral Radius of F. If we assume that F has spectral radius $\rho(F) < 1$ (i.e., the magnitude of the largest eigenvalue of F is less than 1), Theorem 3.5 implies that for $n \to \infty$ the conditional mean goes exponentially fast to the marginal mean S, with a rate proportional $\rho(F)$.

Univariate Process ($q = 1$). In the univariate case we can analytically examine the conditional mean $E[\sigma_t \mid \sigma_0]$ without relying on the limit of $n \to \infty$. Using the notation of Section 2.3 and recursing on the form of $E[\sigma_t \mid \sigma_{t-1}]$ specified in Theorem 3.3,

$$E[\sigma_t \mid \sigma_0] = \left(\frac{nf^2+1}{n+1}\right)^t \sigma_0 + \frac{ns(1-f^2)}{n+1} \sum_{\tau=0}^{t-1} \left(\frac{nf^2+1}{n+1}\right)^\tau. \tag{3.11}$$

Since we assume $|f| < 1$ this becomes

$$E[\sigma_t \mid \sigma_0] = s + \left(\frac{nf^2+1}{n+1}\right)^t (\sigma_0 - s). \tag{3.12}$$

This has the form of a linear AR(1) model with AR parameter $(nf^2+1)/(n+1)$. Then $\lim_{t\to\infty} E[\sigma_t \mid \sigma_0] = s$ when $|f| < 1$, and does so exponentially fast regardless of n. However, the overall rate of this exponential forgetting is governed by the AR parameter $(nf^2+1)/(n+1)$.

Shared Eigenvectors Between F, S and Σ_0. In Theorem 3.4, we examined a principal component IW-AR process based on the shared eigenvectors of F and S. If we further assume that Σ_0 shares eigenvectors with F and S, the following theorem shows that $\hat{\Sigma}_t = E'\Sigma_t E$ remains diagonal in expectation and has a closed form mean recursion.

THEOREM 3.6. *Under the conditions of Theorem 3.4 assume $\Sigma_0 = E\Theta_0 E'$ where Θ_0 is diagonal, $\Theta_0 = \text{diag}(\theta_0)$ for some $q-$vector θ_0 with positive elements. Let $\theta_{t|0}$ denote the $q-$vector of the eigenvalues of $E[\Sigma_t \mid \Sigma_0]$, $\xi = \text{diag}(Q)$ and $\xi_{-1} = \text{diag}\left(Q^{-1}\right)$. Then $E[\Sigma_t \mid \Sigma_0]$ has the form of a first-order, non-diagonal autoregression on $\theta_{t|0}$*

$$\theta_{t|0} = c_{n,q}(I - R^2)\xi + \left[n^{-1}c_{n,q}(I - R^2)\xi\xi'_{-1} + R^2\right]\theta_{t-1|0}, \tag{3.13}$$

or

$$\theta_{t|0} = B^t\theta_0 + \sum_{\tau=0}^{t-1} B^\tau \alpha \tag{3.14}$$

where $\alpha = c_{n,q}(I - R^2)\xi$ and $B = n^{-1}c_{n,q}(I - R^2)\xi\xi'_{-1} + R^2$. Assuming a stationary process such that $||R||_0 < 1$,

$$\theta_{t|0} = B^t\theta_0 + (I - B)^{-1}(I - B^t)\alpha, \tag{3.15}$$

and

$$\lim_{t\to\infty} \theta_{t|0} = \xi. \tag{3.16}$$

That is, the eigenvalues of the limiting conditional mean are exactly those of the marginal mean S.

PROOF. See Appendix. □

Recalling that $\theta_{t|0}$ fully determines $E[\Sigma_t \mid \Sigma_0]$ in the case of shared eigenvectors, we once again conclude that the conditional mean of the process forgets the initial condition Σ_0 exponentially fast—this occurs irregardless of the value of n.

4. Data Augmentation. Augmentation of the observation model eqn. (2.1) provides interpretation of the latent innovations process $\{\Upsilon_t, \Psi_t\}$ as well as forming central and critical theoretical development for posterior computations as detailed in Section 5. Conditional on Σ_0 and the innovations sequence, the observation model can be regarded as arising by marginalization over an inherent latent $q-$vector process z_t, $(t = 1, \ldots)$, where

$$x_t \mid z_t \sim N(\Upsilon_t z_t, \Psi_t) \quad \text{and} \quad z_t \sim N(0, \Sigma_{t-1}) \tag{4.1}$$

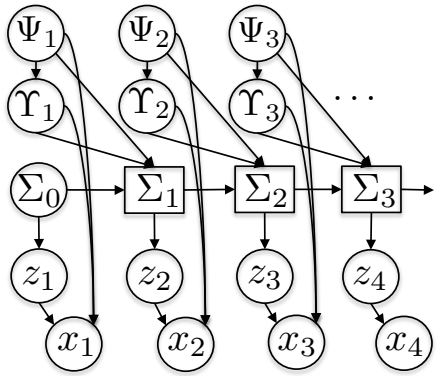

FIG 1. *Representation of the graphical model of the IW-AR(1) process under augmentation by the latent z_t. Circles indicate random variables, arrows imply probabilistic conditional relationships while squares represent quantities that are deterministic based on an instantiation of the variables in their parents nodes.*

independently over time. That is, the observations x_t are from a conditionally linear model with latent covariate vectors z_t and regression parameters $\{\Upsilon_t, \Psi_t\}$. The normal-inverse Wishart prior for $\{\Upsilon_t, \Psi_t\}$ provides a conjugate prior in this standard multivariate regression framework. See Figure 1 for a graphical model representation of this process.

Let $y_t = [z_t' \; x_t']'$ and $\Delta_t = \{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$. Then

$$
\begin{aligned}
p(y_{1:T} \mid \Delta_{1:T}) &= \prod_{t=1}^{T} N(x_t \mid \Upsilon_t z_t, \Psi_t) N(z_t \mid 0, \Sigma_{t-1}) \\
p(\Delta_{1:T}) &= p(\Sigma_0 \mid n, S) \prod_{t=1}^{T} NIW_q(\Upsilon_t, \Psi_t \mid F, n, S)
\end{aligned}
\tag{4.2}
$$

where NIW denotes the matrix normal, inverse Wishart prior on $\{\Upsilon_t, \Psi_t\}$ of eqn. (2.6). We omit the dependency of the left hand side on the hyperparameters n, F, and S for notational simplicity. Figure 2 displays the resulting graphical model, clearly illustrating the simplified conditional independence structure that enables computation as developed below. Note that Δ_t plays the role of an augmented state and the evolution to time t defines Σ_t as a deterministic function of this state.

5. Model Fitting via MCMC. For model fitting, we develop a Markov chain Monte Carlo (MCMC) sampler that harnesses the simplified state-space structure of the augmented model comprised of Gaussian observations with an IW-AR process. This structure (Figure 2) immediately suggests a natural MCMC sampler that iterates between the following steps:

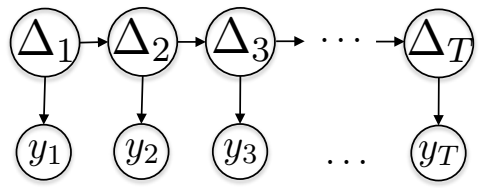

FIG 2. *Representation of the graphical model for the augmented data* $y_t = [z_t' \; x_t']'$ *and latent states* $\Delta_t = \{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$.

Step 1. Impute the latent process by sampling each z_t from

$$\begin{aligned} p(z_t \mid x_t, \Upsilon_t, \Psi_t, \Sigma_{t-1}) &\propto N(z_t \mid 0, \Sigma_{t-1}) N(x_t \mid \Upsilon_t z_t, \Psi_t) \\ &= N(z_t \mid (\Sigma_{t-1}^{-1} + \Upsilon_t' \Psi_t^{-1} \Upsilon_t)^{-1} \Upsilon_t' \Psi_t^{-1} x_t, (\Sigma_{t-1}^{-1} + \Upsilon_t' \Psi_t^{-1} \Upsilon_t)^{-1}). \end{aligned} \tag{5.1}$$

Step 2. Update the hyperparameters S and F conditioned on $x_{1:T}$ and $z_{1:T}$ by sampling steps defined in Section 5.2 below.

Step 3. Impute the augmented variance matrix states using a Metropolis-Hastings approach targeting the conditional posterior

$$p(\Delta_{1:T} \mid x_{1:T}, z_{1:T}, F, S). \tag{5.2}$$

We do this using an approximate forward filtering, backward sampling (FFBS) algorithm to define proposal distributions; see Section 5.1 below.

Note that *Step 2* and *Step 3* comprise a block-sampling of the IW-AR hyperparameters $\{F, S\}$ and the augmented process $\Delta_{1:T}$ conditioned on $x_{1:T}$ and $z_{1:T}$. This greatly improves efficiency relative to a sampler that iterates between (i) sampling $\Delta_{1:T}$ given F, S, $x_{1:T}$. and $z_{1:T}$ and (ii) sampling $\{F, S\}$ given $\Delta_{1:T}$ (which is then conditionally independent of $x_{1:T}$. and $z_{1:T}$).

5.1. *Forward Filtering, Backward Sampling.* We utilize the fact that there is a deterministic mapping from Δ_t to the augmented matrix

$$\Delta_t \to \begin{pmatrix} \Sigma_{t-1} & \Sigma_{t-1}\Upsilon_t' \\ \Upsilon_t \Sigma_{t-1} & \Sigma_t \end{pmatrix}, \tag{5.3}$$

and thus use the two interchangeably. Our goal is to develop an approximate forward filtering algorithm that produces an approximation to $p(\Delta_{1:T} \mid y_{1:T})$, which can then be used in backward-sampling a posterior sequence $\Sigma_{0:T}$. We examine the filtering and sampling stages in turn.

Approximate Forward Filtering. An exact forward filtering would involve recursively *updating* $p(\Delta_t \mid y_{1:t-1})$ to $p(\Delta_t \mid y_{1:t})$ and *propagating* $p(\Delta_t \mid y_{1:t})$ to $p(\Delta_{t+1} \mid y_{1:t})$. However, as examined in the Appendix, this filter is analytically intractable for the IW-AR so we use an approximate filtering procedure based on moment-matching in order to maintain inverse Wishart approximations to each propagate and update step. Specifically, let $g_{t-1|t-1}(\Delta_{t-1} \mid y_{1:t-1})$ denote the approximation to the posterior $p(\Delta_{t-1} \mid y_{1:t-1})$ at time $t-1$. We then approximate the predictive distribution $p(\Delta_t \mid y_{1:t-1})$ by

$$g_{t|t-1}(\Delta_t \mid y_{1:t-1}) = \text{IW}\left(r_t, (r_t-2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}]\right). \tag{5.4}$$

Here, r_t is a specified degree of freedom to use in the approximation at time t. The subsequent update step of incorporating observation y_t is exact based on the approximations made so far. Namely,

$$g_{t|t}(\Delta_t \mid y_{1:t}) = \text{IW}\left(r_t+1, (r_t-2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] + y_t y_t'\right). \tag{5.5}$$

The required expectation here is easily seen to be

$$\begin{aligned} &E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] \\ &\qquad = \begin{pmatrix} S_{t-1} & S_{t-1}F' \\ FS_{t-1} & FS_{t-1}F' + c_{n,q}(1+\text{tr}(S_{t-1}(nS)^{-1})V) \end{pmatrix} \end{aligned} \tag{5.6}$$

where the S_t sequence is updated using the identity

$$(r_t-1)S_t = (r_t-2)\left\{FS_{t-1}F' + c_{n,q}(1+\text{tr}(S_{t-1}(nS)^{-1}))V\right\} + x_t x_t' \tag{5.7}$$

as further detailed in the Appendix.

In summary, the approximate forward filtering is defined by a recursion in which the inverse Wishart distribution for Δ_{t-1} is converted into a predictive distribution for Δ_t. This predictive distribution is also taken to be inverse Wishart degree of freedom r_t and mean $E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}]$ – i.e., the predictive mean under the distribution $g_{t-1|t-1}$ and the dynamics specified by the IW-AR prior. The inverse Wishart predictive distribution is then directly updated to the resulting inverse Wishart distribution for Δ_t.

Backward Sampling. Running these forward filtering computations to time $t = T$ yields $g_{T|T}(\Delta_T \mid y_{1:T})$ approximating the true posterior $p(\Delta_T \mid y_{1:T})$. We use the sequence of approximations $g_{t|t}(\Delta_t \mid y_{1:t})$ in deriving a backwards sampling stage, which we show is exact based on the approximations made in the forward filtering. At time $t = T$, we sample Σ_T from the implied approximate posterior margin

$$\Sigma_T \sim g_{T|T}(\Sigma_T|y_{1:T}) = \text{IW}(r_T+1, (r_T-2)E_{g_{T-1|T-1}}[\Sigma_T|y_{1:T-1}] + x_T x_T'). \tag{5.8}$$

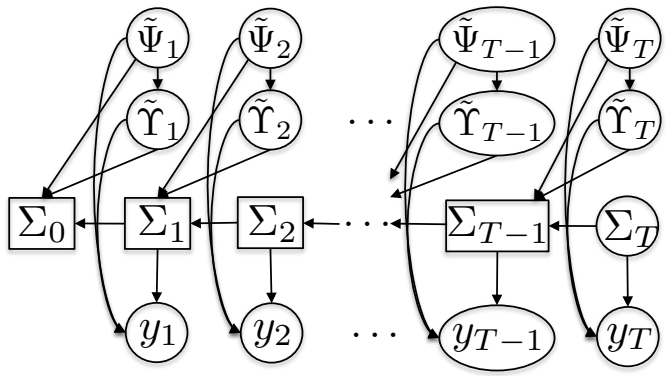

FIG 3. *Graphical model of the reverse time inverse Wishart autoregressive process where* $\tilde{\Psi}_t \sim IW(n+q+2, n\tilde{V})$ *and* $\tilde{\Upsilon}_t \mid \tilde{\Psi}_t \sim N(\tilde{F}, \tilde{\Psi}_t, (nS)^{-1})$ *with* $\tilde{F} = SF'S^{-1}$ *and* $\tilde{V} = S - SF'S^{-1}FS$*. The reverse time transitions are defined by* $\Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Upsilon}_t \Sigma_t \tilde{\Upsilon}_t'$*. Since one can deterministically compute* $\{\Upsilon_t, \Psi_t\}$ *from* $\tilde{\Delta}_t = \{\tilde{\Upsilon}_t, \tilde{\Psi}_t, \Sigma_t\}$ *(see eqn. (5.12)), the augmented observation* y_t *are conditionally independent given* $\tilde{\Delta}_{1:T}$*.*

We then harness the reverse time process, depicted in Figure 3. As we iterate backwards in time, we condition on the previously sampled Σ_t to sample Σ_{t-1} as follows. We first sample

$$\begin{aligned} &\tilde{\Psi}_t \sim \mathrm{IW}(r_t + 1 + q, G_t^{11} - G_t^{21'}(G_t^{22})^{-1}G_t^{21}) \\ &\tilde{\Upsilon}_t \mid \tilde{\Psi}_t \sim N(G_t^{21'}(G_t^{22})^{-1}, \tilde{\Psi}_t, (G_t^{22})^{-1}), \end{aligned} \tag{5.9}$$

and then set

$$\Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Upsilon}_t \Sigma_t \tilde{\Upsilon}_t'. \tag{5.10}$$

Here, $G_t = (r_t - 2)E_{g_{t-1|t-1}}[\Delta_t | y_{1:t-1}] + y_t y_t'$, with G_t^{11}, G_t^{21}, G_t^{22} denoting the three unique $q \times q$ sub-blocks ($G_t^{12} = G_t^{21'}$). These terms, which can be regarded as sufficient statistics of the forward filtering procedure, can be written as

$$\begin{aligned} G_t^{11} &= (r_t - 2)S_{t-1} + z_t z_t' \\ G_t^{21} &= (r_t - 2)FS_{t-1} + x_t z_t' \\ G_t^{22} &= (r_t - 1)S_t, \end{aligned} \tag{5.11}$$

with S_t recursively defined as in (5.7). In practice, conditioned on $\{n, S, F\}$, the sequence $S_{1:T}$ is precomputed and simply accessed in the backward sampling of $\Sigma_{0:T}$.

Note that if we wish to impute $\{\Upsilon_t, \Psi_t\}$, we can deterministically compute them based on the sampled $\{\Sigma_{t-1}, \Sigma_t, \tilde{\Upsilon}_t, \tilde{\Psi}_t\}$; that is,

$$\Upsilon_t = \Sigma_t \tilde{\Upsilon}_t' \Sigma_{t-1}^{-1} \quad \text{and} \quad \Psi_t = \Sigma_t - \Upsilon_t \Sigma_{t-1} \Upsilon_t'. \tag{5.12}$$

Accept-Reject Calculation. We use the proposed approximate FFBS scheme as a proposal distribution for a Metropolis Hastings stage. Let $q(\cdot \mid \cdot)$ represent the proposal distribution for $\{\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T}\}$ implied by the sequence of forward filtering approximations $g_{t|t}(\Delta_t \mid y_{1:t})$. For every proposed $\Delta^*_{1:T}$, we compare the ratio

$$r(\Delta^*_{1:T}) = \frac{p(\Sigma^*_T, \tilde{\Upsilon}^*_{1:T}, \tilde{\Psi}^*_{1:T} \mid y_{1:T})}{q(\Sigma^*_T, \tilde{\Upsilon}^*_{1:T}, \tilde{\Psi}^*_{1:T} \mid y_{1:T})} \tag{5.13}$$

to $r(\Delta_{1:T})$, where $\Delta_{1:T}$ is the previous sample of the augmented sequence. If $r(\Delta^*_{1:T}) > r(\Delta_{1:T})$, we accept the proposed sequence. Otherwise, we accept the sequence with probability $r(\Delta^*_{1:T})/r(\Delta_{1:T})$.

The accept-reject ratio is calculated as follows. Noting that there is a one-to-one mapping between $\{\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T}\}$ and $\{\Sigma_0, \Upsilon_{1:T}, \Psi_{1:T}\}$,

$$\frac{p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T} \mid y_{1:T})}{q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T} \mid y_{1:T})} \propto \frac{p(y_{1:T} \mid \Sigma_0, \Upsilon_{1:T}, \Psi_{1:T}) p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T})}{q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T} \mid y_{1:T})}. \tag{5.14}$$

The augmented data likelihood is given by

$$p(y_{1:T} \mid \Sigma_0, \Upsilon_{1:T}, \Psi_{1:T}) = \prod_{t=1}^{T} N(x_t \mid \Upsilon_t z_t, \Psi_t) N(z_t \mid 0, \Sigma_{t-1}). \tag{5.15}$$

As specified in eqn. (3.4), the prior of the reverse time process is given by

$$\begin{aligned} p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T}) &= \mathrm{IW}(\Sigma_T \mid n+2, nS) \\ &\quad \times \prod_{t=1}^{T} \mathrm{IW}(\tilde{\Psi}_t \mid n+q+2, n\tilde{V}) N(\tilde{\Upsilon}_t \mid \tilde{F}, \tilde{\Psi}_t, (nS)^{-1}). \end{aligned} \tag{5.16}$$

Similarly, the proposal density decomposes as

$$q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T} \mid y_{1:T}) = \mathrm{IW}(\Sigma_T \mid r_T + 1, G^{22}_T) \prod_{t=1}^{T} f_t(\tilde{\Upsilon}_t) h_t(\tilde{\Psi}_t \mid \tilde{\Upsilon}_t) \tag{5.17}$$

where

$$f_t(\tilde{\Upsilon}_t) = \mathrm{IW}(\tilde{\Psi}_t \mid r_t + q + 1, G^{11}_t - G^{21'}_t (G^{22}_t)^{-1} G^{21}_t)$$

and

$$h_t(\tilde{\Psi}_t \mid \tilde{\Upsilon}_t) = N(\tilde{\Upsilon}_t \mid G^{21'}_t (G^{22}_t)^{-1}, \tilde{\Psi}_t, (G^{22}_t)^{-1}).$$

FFBS Computations. One important note is that the acceptance rate decreases exponentially fast with the length of the time series, as with all Metropolis-based samplers for sequences of states in hidden Markov models. Recall that the proposed $\Sigma^*_{0:T}$ sequence is based on a sample Σ^*_T from $g_{T|T}$ and a collection of T independent samples $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ from distributions based on $g_{t|t}$. If the final approximate filtered distribution $g_{T|T}$ is a poor approximation to the true distribution, then the collection of T independent proposed innovations $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ are unlikely to result in a $\Sigma^*_{0:T}$ that explains the data well. The accuracy of the approximation $g_{T|T}$ decreases with T. Furthermore, even if $g_{T|T}$ is a good approximation to the true posterior, a single poor innovations sample $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ can be detrimental since the effects propagate in defining $\Sigma^*_{0:t-1}$. The chance of obtaining an unlikely sample $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ for some t increases with T.

Since the distributions contributing to the accept-reject ratio $r(\Delta_{1:T})$ factor over t, one can sequentially compute and monitor this ratio based on the samples of Σ^*_T and $\{\tilde{\Upsilon}^*_{t:T}, \tilde{\Psi}^*_{t:T}\}$. One can then imagine harnessing ideas from randomness recycling (Fill and Huber, 2001) to improve efficiency by rejecting locally instead of rejecting an entire sample path from $t = 0, \ldots, T$. Additionally, one could develop adaptive methods in which samples $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ leading to drastic declines in the acceptance-ratio-to-t were rejected and $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ was then resampled, but only for some finite period of adaptation. These ideas all focus on the ability to accept or reject entire sub-sequences $\{\Sigma^*_T, \tilde{\Upsilon}^*_{t:T}, \tilde{\Psi}^*_{t:T}\}$, and require theoretical analysis to justify convergence to the correct stationary distribution. Alternatively, we develop below an *innovations-based* sampling approach in which we fix $\{\Sigma_T, \tilde{\Upsilon}_{1:t-1,t+1:T}, \tilde{\Psi}_{1:t-1,t+1:T}\}$ and simply consider accepting or rejecting $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ at a single time step t.

An Innovations-Based Sampler. We propose an *innovations* sampler that accepts or rejects $\{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$ for every t instead of accepting or rejecting the entire chain $\Sigma^*_{0:T}$ induced from the collection of these backwards innovations samples (and Σ^*_T). Specifically, let $\theta_t = \{\tilde{\Upsilon}_t, \tilde{\Psi}_t\}$ and $\mathcal{S}_t = \{\theta_1, \theta_2, \ldots, \theta_T, \Sigma_T\}$. For each t we propose

$$\mathcal{S}^*_t = \{\theta_1, \theta_2, \ldots, \theta_{t-1}, \theta^*_t, \theta_{t+1}, \ldots, \theta_T, \Sigma_T\}, \tag{5.18}$$

with $\theta^*_t = \{\tilde{\Upsilon}^*_t, \tilde{\Psi}^*_t\}$. The accept-reject ratio based on eqn. (5.13) simplifies to

$$\begin{aligned}
\frac{r(\Delta^*_{1:T})}{r(\Delta_{1:T})} &= \frac{p(\mathcal{S}^*_t \mid y_{1:T})}{q(\mathcal{S}^*_t \mid y_{1:T})} \frac{q(\mathcal{S}_t \mid y_{1:T})}{p(\mathcal{S}_t \mid y_{1:T})} \\
&= \frac{p(y_{1:T} \mid \Sigma^*_0, \Upsilon^*_{1:t}, \Psi^*_{1:t}, \Upsilon_{t+1:T}, \Psi_{t+1:T})}{p(y_{1:T} \mid \Sigma_0, \Upsilon_{1:T}, \Psi_{1:T})} \frac{p(\mathcal{S}^*_t)}{p(\mathcal{S}_t)} \frac{q(\mathcal{S}_t \mid y_{1:T})}{q(\mathcal{S}^*_t \mid y_{1:T})} \\
&= \frac{\prod_{\tau=1}^t N(x_\tau \mid \Upsilon^*_\tau z_\tau, \Psi^*_\tau) N(z_\tau \mid \Sigma^*_{\tau-1})}{\prod_{\tau=1}^t N(x_\tau \mid \Upsilon_\tau z_\tau, \Psi_\tau) N(z_\tau \mid \Sigma_{\tau-1})} \frac{p_t(\theta^*_t)}{p_t(\theta_t)} \frac{q_t(\theta_t \mid y_{1:T})}{q_t(\theta^*_t \mid y_{1:T})}.
\end{aligned} \tag{5.19}$$

Here $p_t(\theta_t)$ denotes the matrix normal, inverse Wishart prior on the backwards innovations $\theta_t = \{\tilde{\Upsilon}_t, \tilde{\Psi}_t\}$ and $q_t(\theta_t \mid y_{1:T})$ the corresponding distribution under the forward-filtering based proposal. That is, $p_t(\cdot)$ and $q_t(\cdot \mid y_{1:T})$ represent the time t components of eqns. (5.16) and (5.17), respectively. We utilize the fact that the prior, proposal and likelihood terms all factor over t. For the prior and proposal, the only terms that differ between the proposed sequence $\mathcal{S}_t^*$ and the previous $\mathcal{S}_t$ are the backwards innovations at time t (i.e, θ_t). For the likelihood, the effects of the change in *backwards* innovations at time t propagate to the *forward* parameters $\{\Sigma_0^*, \Upsilon_{1:t}^*, \Psi_{1:t}^*\}$ while leaving $\{\Upsilon_{t+1:T}, \Psi_{t+1:T}\}$ unchanged. See Figure 3.

Note that the proposed innovations sampler is quite computationally intensive since the accept-reject ratio calculation for each proposed $\{\tilde{\Upsilon}_t^*, \tilde{\Psi}_t^*\}$ requires re-computing $\Sigma_{0:t-1}^*$. In practice, we employ an approximate sampler that harnesses the fact that the effects of a given $\{\tilde{\Upsilon}_t^*, \tilde{\Psi}_t^*\}$ on $\Sigma_{t-\tau}^*$ decreases in expectation as τ increases. That is, $\{\tilde{\Upsilon}_t^*, \tilde{\Psi}_t^*\}$ represents a stochastic input whose effect is propagated through a stable dynamical system (assuming the F has spectral norm less than 1). In particular, we only calculate $\Sigma_{t-\tau}^*$ for τ increasing until the Frobenius norm $||\Sigma_{t-\tau}^* - \Sigma_{t-\tau}||_2 < \epsilon$ for some pre-specified, small value $\epsilon > 0$. That is, we propagate the effects of $\{\tilde{\Upsilon}_t^*, \tilde{\Psi}_t^*\}$ until the value of $\Sigma_{t-\tau}^*$ becomes nearly indistinguishable numerically from the previous $\Sigma_{t-\tau}$. Alternatively, in order to maintain a constant per-sample computational complexity, one can specify a fixed lag τ based on F and n since these hyperparameters determine (in expectation) the rate at which the effects of the proposed $\{\tilde{\Upsilon}_t^*, \tilde{\Psi}_t^*\}$ decay.

In contrast to sequential sampling of $\{\tilde{\Upsilon}_t, \tilde{\Psi}_t\}$ for $t = T, \ldots, 1$, one can imagine focusing on regions where the current Σ_t is "poor", where "poor" is determined by some specified metric (e.g., the likelihood function $N(x_t \mid 0, \Sigma_t)$). As long as there is still positive probability of considering any $t \in \{1, \ldots, T\}$, the resulting sampler will converge to the correct stationary distribution.

Finally, instead of always running an approximate forward filter and performing backward sampling (where the "backward sampling" can occur in any order based on the innovations representation), one could run a *backward filter* and perform *forward sampling* (BFFS), exploiting the theory of the reverse-time IW-AR process. By interchanging FFBS with BFFS, the errors aggregated during filtering and the uncertainty inherent at the filter's starting point alternate from $t = 0$ to $t = T$, thus producing samples closer to the values that would be obtained if smoothing were analytically feasible.

5.2. *Hyperparameters.* In sampling the IW-AR hyperparameters F and S, we need to ensure that $V = S - FSF'$ remains positive definite. Section 3.2 explored two cases in which simple constraints on F imply V positive definite for S positive definite: (i) $F = \rho I_q$ or (ii) $F = ERE'$ with E the eigenvectors of S. A simple

framework for sampling the hyperparameters in this case is to propose S from a Wishart distribution, thus ensuring its positive-definiteness, and the eigenvalues of F from a beta proposal, thus ensuring spectral radius bounded by 1. The induced V will then be positive definite. One can also assume Wishart and beta priors.

Both of the above specifications of F and S lead to reversible IW-AR processes. For a non-reversible IW-AR process (assuming S non-diagonal), we can take $F = \text{diag}(\rho)$, implying $V = (1 \cdot 1' - \rho \cdot \rho') \circ S$ where $\circ$ denotes the Hadamard product. Note that even with $|\rho_i| < 1$ and S positive definite, V need not be positive definite. However, for V positive definite and $|\rho_i| < 1$, S will be positive definite and has elements simply defined by $S_{ij} = V_{ij}/(1 - \rho_i \rho_j)$. Thus, in the case of F diagonal we sample F and V and then compute S from these values. Once again, we employ a beta prior on ρ_i and a Wishart prior now on V. The details of the posterior computations for the case of F diagonal are outlined below. The case of $F = ERE'$ and $S = EQE'$ follows similarly.

Let $\text{W}(v_0, V_0)$ denote a Wishart prior for V and $\text{Beta}(c\rho_{0,i}, c(1 - \rho_{0,i}))$ a beta prior for ρ_i. We use an independence chain sampler in which V^* is proposed from a Wishart proposal $\text{W}(v_1, V_1)$ and ρ_i^* from a beta proposal $\text{Beta}(d\rho_{1,i}, d(1 - \rho_{1,i}))$. The accept-reject ratio is then calculated based on the ratio

(5.20)
$$\frac{r(V^*, F^*)}{r(V, F)} = \frac{p(x_{1:T} \mid z_{1:T}, F^*, V^*)p(z_{1:T} \mid F^*, V^*)p(V^*)p(F^*)q(V)q(F)}{p(x_{1:T} \mid z_{1:T}, F, V)p(z_{1:T} \mid F, V)p(V)p(F)q(V^*)q(F^*)}.$$

Here, $p(\cdot)$ and $q(\cdot)$ denote the prior and proposal for the specified argument, respectively. The conditional likelihood $p(x_{1:T} \mid z_{1:T}, F, V)$ and marginal likelihood $p(z_{1:T} \mid F, V)$ are derived below. We interchange $\{F, V\}$ and $\{F, S\}$ since there is a bijective mapping between the two when F is diagonal.

Conditional Likelihood $p(x_{1:T} \mid z_{1:T}, F, S)$. Since $x_t \sim N(\Upsilon_t z_t, \Psi_t)$ with $\Upsilon_t \mid \Psi_t \sim N(F, \Psi_t, (nS)^{-1})$ and $\Psi_t \sim IW(n + q + 2, nV)$, marginalizing Υ_t and Ψ_t yields a multivariate t distribution as detailed in the Appendix. This results in

$$p(x_{1:T} \mid z_{1:T}, F, S) = \prod_{t=1}^{T} t_{n+q+2}\left(x_t \mid Fz_t, c_{n,q+2}\{1 + z_t'(nS)^{-1}z_t\}V\right). \tag{5.21}$$

Marginal Likelihood $p(z_{1:T} \mid F, S)$. Computing the marginal likelihood requires the evaluation of the analytically intractable integral

$$p(z_{1:T} \mid F, S) = \int \prod_t p(z_t \mid \Sigma_{t-1})p(\Sigma_{0:T} \mid F, S)d\Sigma_{0:T}. \tag{5.22}$$

However, we can approximate the marginal likelihood by employing an approximate filter in a manner analogous to that of Section 5.1. In particular, if we had

an exact filter that produced the *predicted* distribution $p(\Sigma_{t-1} \mid z_{1:t-1})$ and the *updated* distribution $p(\Sigma_{t-1} \mid z_{1:t})$, we could recursively compute the marginal likelihood as

$$p(z_{1:t}) = \frac{p(z_t \mid \Sigma_{t-1})p(\Sigma_{t-1} \mid z_{1:t-1})}{p(\Sigma_{t-1} \mid z_{1:t})} p(z_{1:t-1}) \tag{5.23}$$

(here F and S are omitted for notational simplicity). Recall that $p(z_t \mid \Sigma_{t-1}) = N(z_t \mid 0, \Sigma_{t-1})$. Since exact filtering is not possible, we propose an approximate moment-matched filter using ideas parallel to those used for the FFBS approximation to $p(\Delta_t \mid y_{1:t})$. Specifically,

$$g_{t|t}(\Sigma_t \mid z_{1:t}) = \text{IW}\left(r_t, (r_t - 2)\bar{\Sigma}_{t|t}\right) \tag{5.24}$$

$$g_{t|t+1}(\Sigma_t \mid z_{1:t+1}) = \text{IW}\left(r_t + 1, (r_t - 2)\bar{\Sigma}_{t|t} + z_{t+1}z'_{t+1}\right), \tag{5.25}$$

with $\bar{\Sigma}_{t|t} = E_{g_{t-1|t}}[\Sigma_t \mid z_{1:t}]$. The matched-means are recursively computed using

$$(r_{t-1} - 1)\bar{\Sigma}_{t-1|t} = (r_{t-1} - 2)\bar{\Sigma}_{t-1|t-1} + z_t z'_t \tag{5.26}$$

$$\bar{\Sigma}_{t|t} = F\bar{\Sigma}_{t-1|t}F' + c_{n,q}\left[1 + \text{tr}\{\bar{\Sigma}_{t-1|t}(nS)^{-1}\}\right], \tag{5.27}$$

with initial condition $\bar{\Sigma}_{0|0} = S$.

Using this approximate filter in eqn. (5.23) for the marginal likelihood and canceling terms yields

$$p(z_{1:t} \mid F, S) = \frac{1}{\pi^{Tq/2}} \prod_{t=1}^{T} \frac{|(r_{t-1} - 2)\bar{\Sigma}_{t-1|t-1}|^{(r_{t-1}+q-1)/2}}{|(r_{t-1} - 2)\bar{\Sigma}_{t-1|t-1} + z_t z'_t|^{(r_{t-1}+q)/2}} \frac{\Gamma\left(\frac{r_{t-1}+q}{2}\right)}{\Gamma\left(\frac{r_{t-1}}{2}\right)}. \tag{5.28}$$

Note that we could have employed our filter for Δ_t based on observations $y_t = [z'_t \; x'_t]'$ to produce an approximation to $p(x_{1:T}, z_{1:T} \mid F, S)$. However, since we have an exact form for $p(x_{1:T} \mid z_{1:T}, F, S)$ we choose to reduce the impact of our approximation by simply using it to compute $p(z_{1:T} \mid F, S)$.

We note further that it is straightforward to analyze $p(F, S \mid \Sigma_0, \Upsilon_{1:T}, \Psi_{1:T})$, suggesting that the MCMC use Gibbs sampler components for F, S that would avoid approximations. However, in practice we found use of this leads to extremely slow mixing rates relative to our proposed strategy above.

6. Stochastic Volatility in Time Series. In this section, we consider a full analysis in which actual *observations* ξ_t are from a VAR(r) model whose *innovations* x_t have IW-AR(1) volatility matrices. That is, we observe $q-$vector data ξ_t such that

$$\xi_t = \sum_{i=1}^{r} A_i \xi_{t-i} + x_t, \qquad x_t \sim N(0, \Sigma_t), \tag{6.1}$$

where A_i is the $q \times q$ autoregressive parameter matrix at lag i and we now assume that Σ_t is an IW-AR(1) process. Define $A = \begin{bmatrix} A_1 & \cdots & A_r \end{bmatrix}$.

We modify the MCMC of Section 5 as follows. In place of *Step 1* that previously sampled $z_{1:T}$ given $\Delta_{1:T}$ and $x_{1:T}$, we now block sample $\{A, z_{1:T}\}$ given $\Delta_{1:T}$ and $\xi_{1-r:T}$. That is, we first sample A given $\Delta_{1:T}$ and $\xi_{1-r:T}$ and then $z_{1:T}$ given A, $\Delta_{1:T}$ and $\xi_{1-r:T}$. Noting that $x_{1:T}$ is a deterministic function of $\xi_{1-r:T}$ and the autoregressive matrix A, the latter step follows exactly as before. *Step 2* and *Step 3* remain unchanged. Thus, the only modification to the sampler is the insertion of a *Step 0* to sample A given $\Delta_{1:T}$ and $\xi_{1-r:T}$. See the Appendix for further details.

6.1. *Example in Analysis of EEG Time Series.* In multi-channel electroencephalogram (EEG) studies, multiple probes on the scalp of a patient undergoing an induced brain seizure generate electrical potential fluctuation signals that represent the spatially localized read-outs of the underlying signal (Krystal et al., 1999a). Much of prior work with clinically relevant data sets has been on the evaluation of time:frequency structure in such series (Freyermuth et al., 2010; Krystal et al., 2000; Ombao et al., 2005) and time-varying parameter vector autoregressions are key tools in this applied context, as in others (Prado and West, 1997, 2010; West et al., 1999). Existing models represent some aspects of cross-series structure in this inherently spatially distributed multiple time series context (Krystal et al., 1999b; Prado et al., 2001), but past studies have shown substantial residual dependencies among estimated innovations processes across EEG probe locations and the implications for estimation of such structure in models that ignore significant patterns of time-varying cross-series correlations are largely unexplored. Hence it is of interest to explore models that use IW-AR models for multivariate volatility processes of innovations driving vector autoregressions as an obvious first-step.

We explore one initial example using the model of eqn. (6.1). Define $\nu_t = \Sigma_t^{-1/2}\xi_t$ so that

$$\nu_t = \sum_{i=1}^{r} A_{t,i}\nu_{t-i} + w_t, \qquad w_t \sim N(0, I) \tag{6.2}$$

where the $A_{t,i} = \Sigma_t^{-1/2} A_i \Sigma_{t-i}$ are structured, time-varying AR parameter matrices for the transformed process. We can fit this model in the original form of eqn. (6.1) and this transformed series is then of interest as defining underlying independent component series.

An example data analysis uses $q = 10$ channels of a sub-sampled series of 1000 time points, taken from the larger data set of West et al. (1999). The original series were collected at a rate of 256/second and these are down-sampled by a factor of 2 here to yield $T = 1000$ observations over roughly 8 seconds. The data were first

standardized over a significantly longer time window, and the selected 8 second section of data corresponds to a recording period containing abnormal neuronal activity and thus increased changes in volatility. The example sets $r = 8$ and uses underlying diagonal autoregressive matrices $A_i = \text{diag}(a_i)$ with independent and relatively diffuse priors $a_i \sim N(0, 10I_q)$.

For the IW-AR model component, we assume the rather general, irreversible IW-AR process with a diagonal $F = \text{diag}(\rho)$ and set $n = 6$. We specify priors on ρ and S based on an exploratory analysis of an earlier held-out section of the time series, $\xi^0_{1:T}$, also of length 1000. Specifically, this was based on estimating innovations $x^0_{1:T}$ from q separate, univariate TVAR models as in Krystal et al. (1999b). Treating these constructed zero-mean series as raw data, the standard variance matrix discounting method (Prado and West, 2010) was applied using an initial 20 degrees of freedom and a discount factor $\beta = 0.95$ to generate 100 independent posterior samples of the series of $q \times q$ variance matrices, say $U_{0:T}$, across this prior, hold-out period. We then applied individual univariate IW-AR(1) models – the inverse gamma processes of Section 2.3 – to each of the diagonal data sets $U_{ii,t}$. From these, we extracted summary information on which to base the priors for the real data analysis, as follows. First, we take $\rho_i \sim \text{Beta}(100\rho_{0,i}, 100(1 - \rho_{0,i}))$, independently, where $\rho^2_{0,i} = (a_i(n+1) - 1)/n$ and a_i is the approximate posterior mean of the IW-AR autoregressive parameter from the hold-out data analysis of $U_{ii,t}$; second, we set $\nu_0 = q + 2$ and $V_0 = (1 \cdot 1' - \rho_0 \cdot \rho_0') \circ S_0$, where S_0 is the sample mean of all of the the $U_{1:T}$.

Although centered around a held-out-data-informed mean, the chosen Wishart prior for S is quite diffuse and the beta priors for the ρ_i are weakly informative relative to the number of observations $T = 1000$. Our use of initial hold-out data to specify priors is coherent and consistent with common practice in other areas of Bayesian forecasting and dynamic modeling such as in using factor models; Aguilar and West (2000), for example, adopt such an approach and give useful discussion of the importance of centering hyperprior support around "reasonable" values for these types of dynamic models.

From an identical analysis on the batch of test data, we infer values ρ_1 and V_1 that are used in specifying the $\text{Beta}(d\rho_{1,i}, d(1 - \rho_{1,i}))$ proposal for ρ_i and $W(v_1, V_1)$ proposal for V used in our MCMC algorithm. After some experimentation, this used tuning parameters $d = 750$ and $v_1 = 40$. The FFBS proposals also rely on defining the moment-matched IW degree of freedom parameters $r_{0:T}$ for which we set $r_0 = n + 2$, which matches the prior specification, and then discount as $r_t = 0.98r_{t-1} + 1$. Also, in employing the approximate innovations-based sampler described in Section 5.1, the analysis monitors based on $||\Sigma^*_{t-\tau} - \Sigma_{t-\tau}||_2 < \epsilon$ and uses $\epsilon = 1e - 4$.

Some summaries of analysis are based on running 5 separate MCMC chains for

5000 iterations, discarding the first 1000 samples of each and thinning by examining every 10th MCMC iteration. Note that we count one full sweep of side-by-side innovations based FFBS of $\Sigma_{0:T}$ as one step in an iteration. The sampler was initialized with F and S based on the mean of their respective proposal distributions and the residuals $x_{1:T}$ computed from q separate univariate TVAR analyses. The sequences $\Sigma_{0:T}$, $\Upsilon_{1:T}$, and $\Psi_{1:T}$ were initialized by directly accepting the first proposal from one step of the FFBS algorithm.

Figure 4 displays volatility trajectories for each of the 10 examined EEG channels showing clear changes in volatility over the 8 seconds of data, while related temporal structure in cross-series covariances is evident in Figure 5. These changes are also captured in Figure 6, which display the time-varying correlations between the EEG channel AR innovations. For the model parameters, Figure 7 shows clear evidence of learning via changes from prior to posterior summaries for the ρ_i and S_{ii} elements; this figure also highlights the high dependence in the IW-AR(1) model and heterogeneity across EEG channels.

7. IW-AR(2) and Higher Order Models. The constructive approach for IW-AR(1) models extends to higher orders. This can be done in a number of ways, as follows.

7.1. *Direct Extension.* For any order $p \geq 1$, transition distributions of IW-AR(p) processes can be defined by the conditionals of $q \times q$ diagonal blocks of underlying inverse Wishart distributions for $(p+1)q \times (p+1)q$ matrices. This involves a direct extension of the basic idea underlying the IW-AR(1) model construction. We develop this here for the case of $p = 2$.

With $p = 2$, begin with

$$\begin{pmatrix} \Delta_{t-1} & \Delta_{t-1}\Gamma_t' \\ \Gamma_t\Delta_{t-1} & \Sigma_t \end{pmatrix} = \begin{pmatrix} \Sigma_{t-2} & \phi_{t-1}' & \gamma_t' \\ \phi_{t-1} & \Sigma_{t-1} & \phi_t' \\ \gamma_t & \phi_t & \Sigma_t \end{pmatrix} \sim IW_{3q}(n+2, nS_3), \tag{7.1}$$

where

$$S_3 = \begin{pmatrix} S & G' & H' \\ G & S & G' \\ H & G & S \end{pmatrix}. \tag{7.2}$$

Then, for all t,

$$\Delta_t \sim IW_{2q}(n+2, nS_2) \quad \text{with} \quad S_2 = \begin{pmatrix} S & G' \\ G & S \end{pmatrix}, \tag{7.3}$$

$$\Sigma_t \sim IW_q(n+2, nS). \tag{7.4}$$

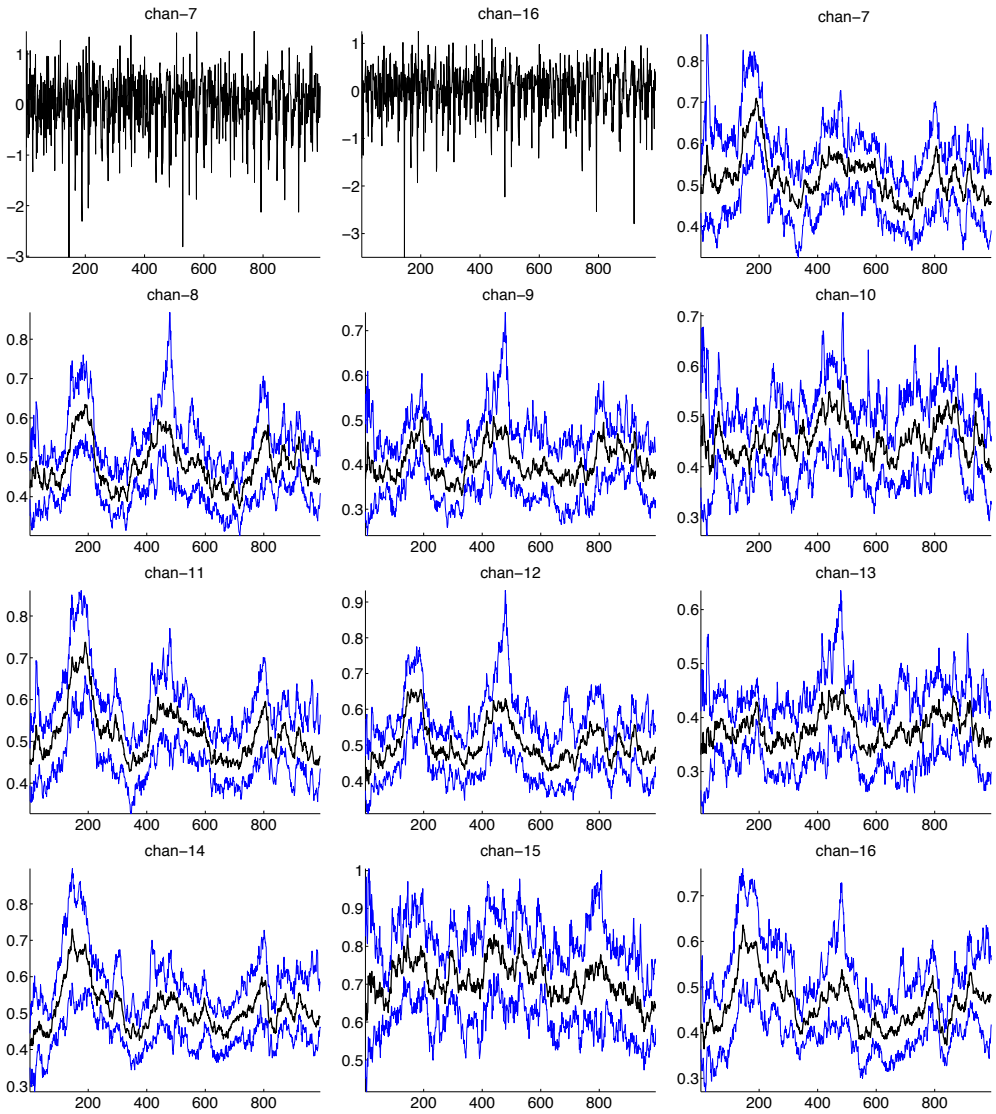

FIG 4. *EEG time series from 2 of the 10 channels (first two frames) followed by estimated trajectories of volatilities* $\Sigma_{ii,t}^{1/2}$ *for each of* $i = 1 : 10$ *time series representing EEG channels 7-16 in the original data set. The IW-AR posterior mean, computed based on averaging over 5 chains from iterations* $[1000 : 10 : 5000]$*, is shown in black. The point-wise 95% highest posterior density intervals are indicated in blue.*

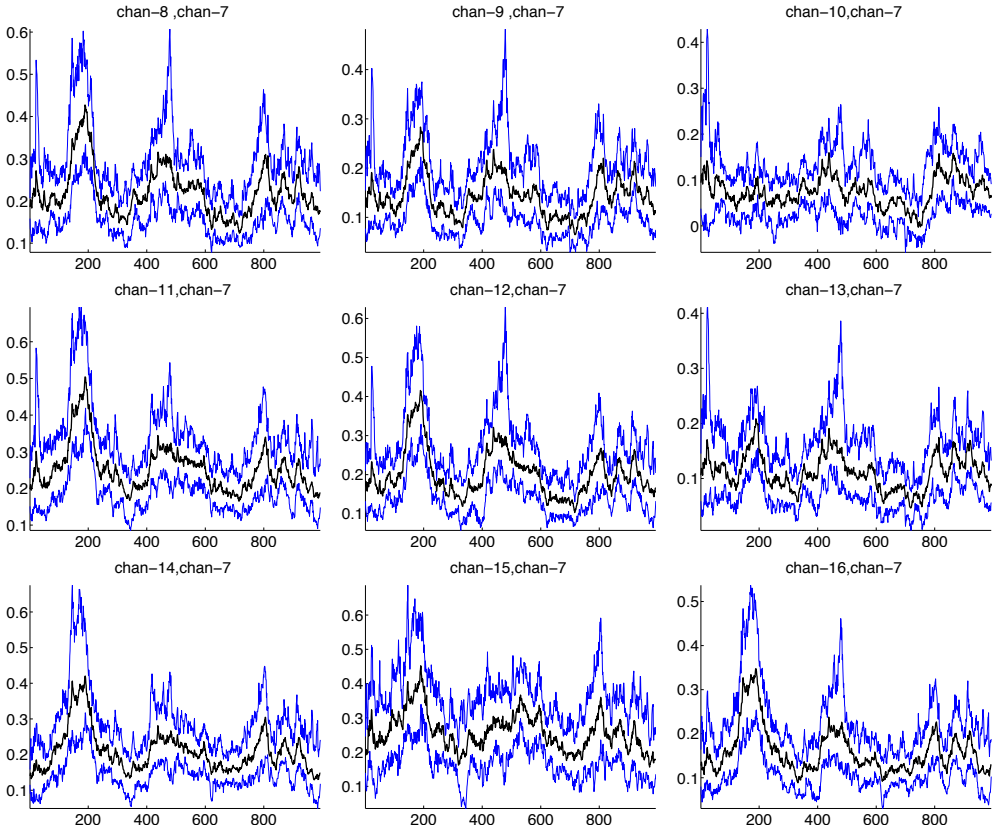

FIG 5. *Estimated trajectories of covariance terms* $\Sigma_{ij,t}$ *for* $i \neq j$ *for* $j = 1$ *(EEG channel 7) colored as in Figure 4.*

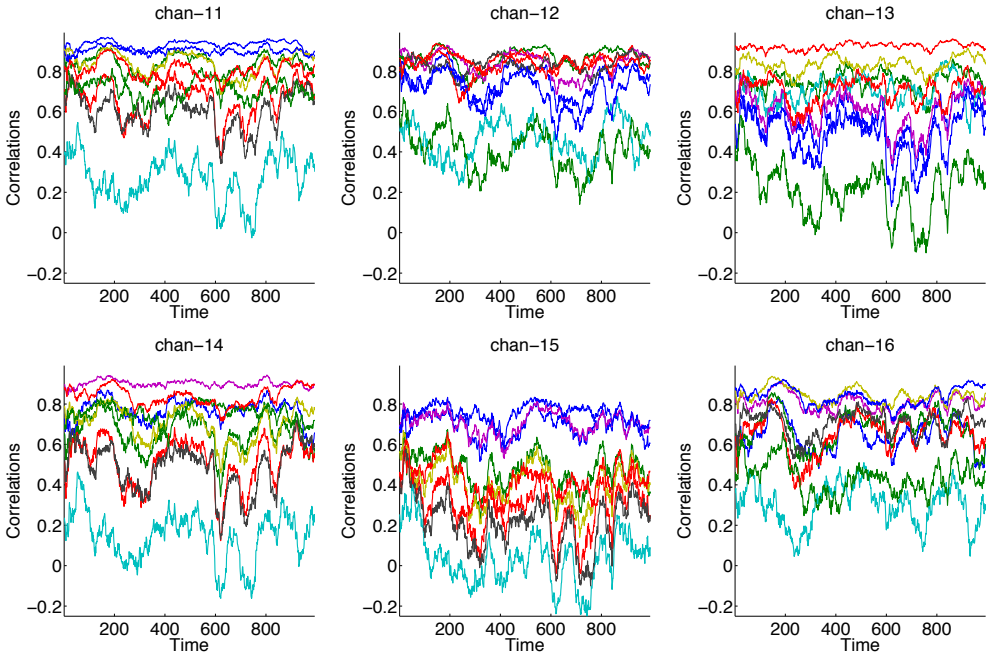

FIG 6. *Estimated trajectories of correlations between each of 6 channels and all other channels as a function of time. The correlations are computed based on posterior means of* Σ_t *using MCMC samples* $[1000 : 10 : 5000]$ *from 5 chains.*

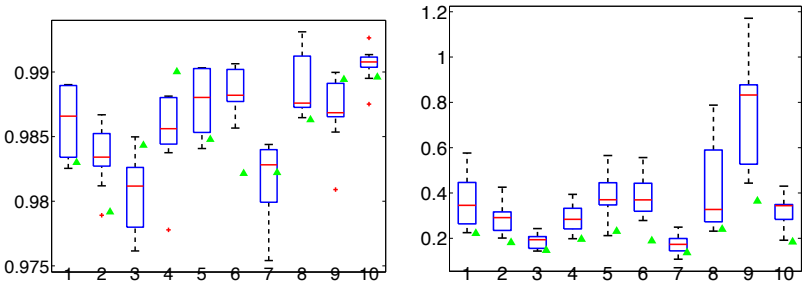

FIG 7. *Box plots of the posterior samples of* ρ_i *and* S_{ii}*, respectively, based on the same MCMC samples as in Figure 4. The prior means are marked by green triangles.*

This then constructively defines a stationary order 2 process with common bivariate and univariate margins. In contrast with the IW-AR(1) construction of Section 2, $\{\phi_t, \gamma_t\}$ in eqn. (7.1) are not independent over time. Rather, if $\Gamma_t = [\Gamma_{1,t}, \Gamma_{2,t}]$, then we have defined an autoregressive process on the augmented variance elements:

$$\gamma_t = \Gamma_{t1}\Sigma_{t-2} + \Gamma_{t2}\phi_{t-1} \tag{7.5}$$

$$\phi_t = \Gamma_{t2}\Sigma_{t-1} + \Gamma_{t1}\phi'_{t-1}. \tag{7.6}$$

The "memory" induced by these off-diagonal elements is evident as the full conditional distribution for Σ_t is $p(\Sigma_t \mid \Delta_{t-1})$, whereas the IW-AR(2) observation model is $p(\Sigma_t \mid \Sigma_{t-1:t-2})$ which involves marginalization over the relevant conditional for the off-diagonal matrices.

As in the case of the IW-AR(1), the construction of eqn. (7.1) implies that

$$\Sigma_t = \Omega_t + \Gamma_t \Delta_{t-1} \Gamma'_t, \tag{7.7}$$

with time t innovation matrices Γ_t $(q \times 2q)$ and Ω_t $(q \times q)$ independent of Δ_{t-1} and distributed as

$$\Omega_t \sim IW_q\left(n+2+2q, n\left(S - \begin{bmatrix} H & G \end{bmatrix} S_2^{-1} \begin{bmatrix} H & G \end{bmatrix}'\right)\right) \tag{7.8}$$

$$\Gamma_t \mid \Omega_t \sim N\left(\begin{bmatrix} H & G \end{bmatrix} S_2^{-1}, \Omega_t, (nS_2)^{-1}\right). \tag{7.9}$$

If we assume that $H = GS^{-1}G$ such that

$$S_3 = \begin{pmatrix} S & G' & G'S^{-1}G' \\ G & S & G' \\ GS^{-1}G & G & S \end{pmatrix} = \begin{pmatrix} S_2 & S_2 \begin{bmatrix} 0 & GS^{-1} \end{bmatrix}' \\ \begin{bmatrix} 0 & GS^{-1} \end{bmatrix} S_2 & S \end{pmatrix}, \tag{7.10}$$

then

$$\Omega_t \sim IW_q\left(n+2+2q, n\left(S - GS^{-1}G'\right)\right) \tag{7.11}$$

$$\Gamma_t \mid \Omega_t \sim N\left(\begin{bmatrix} 0 & GS^{-1} \end{bmatrix}, \Omega_t, (nS_2)^{-1}\right). \tag{7.12}$$

Furthermore, taking $G = FS$ leads to

$$S_3 = \begin{pmatrix} S & SF' & SF^{2'} \\ FS & S & SF' \\ F^2S & FS & S \end{pmatrix} = \begin{pmatrix} S_2 & S_2 \begin{bmatrix} 0 & F \end{bmatrix}' \\ \begin{bmatrix} 0 & F \end{bmatrix} S_2 & S \end{pmatrix} \tag{7.13}$$

and

$$\Omega_t \sim IW_q\left(n+2+2q, n\left(S - FSF'\right)\right) \tag{7.14}$$

$$\Gamma_t \mid \Omega_t \sim N\left(\begin{bmatrix} 0 & F \end{bmatrix}, \Omega_t, (nS_2)^{-1}\right). \tag{7.15}$$

Note that for the specified IW_{3q} to be a valid distribution, we need S_3 positive definite. As before, this is equivalent to S_2 and the Schur complement of S_3 being positive definite. Taking $G = FS$ and $H = F^2S$, the Schur complement is simply $S - FSF'$ and

$$|S_3| = |S_2||S - FSF'| = |S||S - FSF'|^2. \tag{7.16}$$

So, just as in the IW-AR(1), we require S and $V = S - FSF'$ to be positive definite; the conditions for a valid process and stationarity have not changed in this extension to the IW-AR(2) process based on the chosen parameterization.

More general IW-AR(p) follow from the obvious extension of this constructive approach. Note that the ancillary off-diagonal blocks of the extended $IW_{(p+1)q}$ matrix defining the IW-AR(p) transition distributions are latent variables that will feature in Bayesian fitting.

7.2. *A Second Constructive Approach to Higher-Order Models.* A related, alternative and novel approach is defined by coupling AR components to generate higher order AR structures. Specifically, take $\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'$ with $p(\Upsilon_t, \Psi_t)$ having a marginal matrix normal, inverse Wishart form as in the IW-AR(1) model of Section 2. Denote the hyperparameters of this IW-AR(1) by $\mu_1 = \{n, F, S\}$.

Now introduce Markovian dependence into the Ψ_t sequence while maintaining the same conditional independence of $(\Upsilon_t|\Psi_t)$ on the history of the Σ_t process. Specifically, take an IW-AR(1) model for Ψ_t so that for each t

$$\Psi_t = \Xi_t + \Phi_t \Psi_{t-1} \Phi_t', \tag{7.17}$$

with time t innovations $\{\Phi_t, \Xi_t\}$ having independent matrix normal, inverse Wishart distributions with defining parameters $\mu_2 = \{n+q, H, V\}$, where $V = S - FSF'$.

This induces a second-order Markov model

$$\Sigma_t = \Upsilon_t \Sigma_{t-1} \Upsilon_t' + \Phi_t \Sigma_{t-1} \Phi_t' - \Phi_t \Upsilon_{t-1} \Sigma_{t-2} \Upsilon_{t-1}' \Phi_t' + \Xi_t. \tag{7.18}$$

Stationarity of this IW-AR(2) process is implied by simply ensuring the stationarity of the IW-AR(1) Σ_t process and the embedded IW-AR(1) Ψ_t process: each of S, V and now $W = V - HVH'$ must be positive definite.

Conditional on Δ_{t-1} earlier defined, we have

$$\begin{aligned} E[\Sigma_t \mid \Delta_{t-1}] &= F\Sigma_{t-1}F' + \\ &\left\{H\Psi_{t-1}H' + c_{n,2q}W + c_{n,2q}W \mathrm{tr}(\Psi_{t-1}(nV)^{-1})\right\}\left(1 + \mathrm{tr}(\Sigma_{t-1}(nS)^{-1})\right), \end{aligned} \tag{7.19}$$

where $c_{n,2q} = n/(n+2q)$ and $\Psi_{t-1} = \Sigma_{t-1} - \Upsilon_{t-1}\Sigma_{t-2}\Upsilon'_{t-1}$ is a deterministic function of the elements of Δ_{t-1}. In the limit as $n \to \infty$,

$$\begin{aligned} E[\Sigma_t \mid \Delta_{t-1}] &= W + F\Sigma_{t-1}F' + H\Psi_{t-1}H' \\ &= S + F(\Sigma_{t-1} - S)F' + \\ &\qquad H((\Sigma_{t-1} - S) - (\Upsilon_{t-1}\Sigma_{t-2}\Upsilon'_{t-1} - FSF'))H'. \end{aligned}$$

Derivations of these conditional moments are provided in the Appendix.

The new structure of joint distributions of the innovations $\{\Upsilon_t, \Psi_t\}$ is to be explored, as are extensions of the MCMC for model fitting.

8. Related Models.

A Markov Latent Variable Construction. As discussed in Section 2.3, our IW-AR(1) model in $q = 1$ dimensions relates closely to the univariate model arising via a latent variable construction introduced by Pitt and Walker (2005); Pitt et al. (2002). We can extend the univariate model of that reference to the multivariate case, as follows. The $\Sigma_{1:T}$ process is coupled with a latent $q \times q$ variance matrix process $\Lambda_{1:T}$ via time t conditionals: $(\Sigma_t|\Lambda_t) \sim IW_q(n+2+a, nS+aB\Lambda_t B')$ with $(\Lambda_t|\Sigma_{t-1}) \sim W_q(a, A\Sigma_{t-1}A'/a)$ a Wishart conditional for some $a > 0$ and non-singular $q\times q$ matrix $A = B^{-1}$. It can be shown that this latent variable construction defines a valid joint distribution with margin $\Sigma_t \sim IW_q(n+2, nS)$ for all t. This leads to an AR(1) transition model $p(\Sigma_t|\Sigma_{t-1})$ in closed form and appears to be the most general AR(1) construction based on the latent variable/process idea of Pitt and Walker (2005).

Although producing identical margins to the IW-AR(1), the proposed multivariate extension of the Pitt and Walker (2005) construction is limited. Such models are always reversible. Most critically, the construction implies $E(\Sigma_t|\Sigma_{t-1}) = S + w(\Sigma_{t-1} - S)$, where $w = a/(n+a)$ is scalar. So, in contrast to the F matrix of the IW-AR(1), there is no notion of multiple autoregressive coefficients for flexible autocorrelation structures on the elements of Σ_t. Finally, it is not clear how to extend to higher-order autoregressive models.

Direct Specification of Transition Distributions. The interesting class of models of Philipov and Glickman (2006a) specifies the *transition* distribution for Σ_t given Σ_{t-1} as inverse Wishart, discounting information from the previous matrix. Specifically, the conditional mean is given by $cF\Sigma_{t-1}^d F'$ for some $c, d > 0$ and matrix F. The Markov construction generates models with stationary structure. Scaling to higher dimensions, the authors apply the proposed stationary Wishart models to the variance matrix of a lower-dimensional latent factor in a latent factor volatility model (Philipov and Glickman, 2006b), extending prior approaches based on dynamic latent factor models (Aguilar and West, 2000; Aguilar et al., 1999). Based on

the specification of Wishart Markov transition kernels, the proposed models do not yield a clear marginal structure and extensions to higher dimensions appear challenging. Furthermore, the proposed sampling-based model fitting strategy yields low acceptance rates in moderate dimensions (e.g., $q = 12$).

Related approaches in Gouriéroux et al. (2009) define Wishart (and non-central Wishart) processes via functions of sample variance matrices of a collection of latent vector autoregressive processes. Specifically, when the degree of freedom n is integer, $\Sigma_t = \sum_{k=1}^{n} x_{kt} x'_{kt}$, with each x_k independently defined via $x_{k,t} = Mx_{k,t-1} + e_{k,t}$ and $e_{k,t} \sim N(0, \Sigma_0)$. For stationary autoregressions, one can analyze the marginal distribution of Σ_t. Extensions to higher order processes are also presented. For model fitting, the authors rely on a (non-asymptotically efficient) method of moments assuming that a sequence of observed *volatility/co-volatility matrices* are available. Extensions to embedding the proposed Wishart autoregressive process within a standard stochastic volatility framework is computationally complicated: a mean model can be estimated based on nonlinear filtering approximations of latent volatilities. Within Bayesian analysis of such a setup, the non-central Wishart does not yield an analytic posterior distribution and is challenging to sample. One might be able to exploit latent process constructions, but the analysis is not straightforward.

AR models for Cholesky elements. Several recent works use linear, normal AR(1) models for off-diagonal elements of the Cholesky of Σ_t and for the log-diagonal elements (Cogley and Sargent, 2005; Lopes et al., 2010b; Nakajima and West, 2011; Primiceri, 2005), building on the Cholesky-based heteroscedastic model of Pourahmadi (1999), and a natural parallel of Bayesian factor models for multivariate volatility (Aguilar and West, 2000; Aguilar et al., 1999). However, each autoregression has an interpretation as the time-varying regression parameters in a model in which the *ordering* of the elements of the observation vector is required and plays a key role in model formulation. For models in which this is not the case, the parameters employed in the autoregressions are less interpretable. We can cast our IW-AR within a similar framework. The inverse Wishart margins for Σ_t and Σ_{t-1} translate to Wishart margins for the precision matrices Σ_t^{-1} and Σ_{t-1}^{-1}. Since each $q \times q$ Wishart matrix can be equivalently described via an outer product of a collection of $q \times q$ identically distributed normal random variables, our IW-AR implicitly arises from a first-order Markov process on the normal random variables and thus defines a Gaussian autoregression, though possibly of a nonlinear form. Note that there are a few key differences between the IW-AR induced element-wise autoregressions and the Cholesky component AR models: (i) the IW-AR autoregressions are on elements of the *precision* matrix and (ii) these elements comprise a matrix square root, but not the Cholesky square root. The issue of implicitly defining an ordering of observations when using a Cholesky decomposition is not present in

the matrix square root considered in the IW-AR case.

9. Final Comments. The structure of the proposed IW-AR processes immediately open possibilities for examining alternative computational methods and extensions to parsimonious modeling of higher-dimensional time series.

The inherent state-space structure of the IW-AR also suggests opportunity to develop more effective computational methods using some variant of particle filtering and particle learning (Carvalho et al., 2010; Lopes et al., 2010a). Among the main challenges here is that of including the fixed parameters – or expanded state variables that include approximate sufficient statistics for these parameters – in particulate representations of filtering distributions (Liu and West, 2001). One possible approach is to harness ideas from particle MCMC (Andrieu et al., 2010). Otherwise, the new IW-AR model class is inherently well-suited to the most effective *reweight/resample* strategies of particle learning for sequential Monte Carlo.

The inverse Wishart distribution also has extensions to *hyper-inverse Wishart* (HIW) distributions for variance matrices constrained by specified graphical models (Carvalho et al., 2007; Dawid and Lauritzen, 1993). Graphical models provide scalable structuring for higher-dimensional problems, and it would be interesting to consider extensions of the IW-AR to *HIW-AR* processes that evolve maintaining the sparsity structure (of the precision matrix) specified by a graphical model.

APPENDIX A: PROOFS OF THEORETICAL PROPERTIES

A.1. Proof of Theorem 3.3. Let υ_{t_k} denote the kth row of Υ_t and f_k the kth column of F. Then, for the IW-AR(1) we can write the (i,j) element of Σ_t as

$$\Sigma_{t,ij} = \Psi_{t,ij} + \upsilon_{t_i}\Sigma_{t-1}\upsilon'_{t_j}. \tag{A.1}$$

Taking the expectation conditioned on Σ_{t-1},

$$\begin{aligned} E\left[\Sigma_{t,ij} \mid \Sigma_{t-1}\right] &= \frac{nV_{ij}}{n+q} + \sum_{k=1}^{q}\sum_{\ell=1}^{q}\Sigma_{t-1,k\ell}E[\upsilon_{t_{ik}}\upsilon_{t_{j\ell}}] \\ &= \frac{nV_{ij}}{n+q} + \sum_{k=1}^{q}\sum_{\ell=1}^{q}\Sigma_{t-1,k\ell}\left\{\left[\frac{nV_{ij}}{n+q}(nS)^{-1}\right]_{k\ell} + F_{ik}F_{j\ell}\right\} \\ &= \frac{nV_{ij}}{n+q} + \mathrm{tr}\left(\Sigma_{t-1}\frac{nV_{ij}}{n+q}(nS)^{-1}\right) + f_i\Sigma_{t-1}f'_j \end{aligned} \tag{A.2}$$

where we have used the fact that $E[\Psi_t] = nV/(n+q)$. In matrix form, we have

$$E[\Sigma_t \mid \Sigma_{t-1}] = \frac{(1+\mathrm{tr}(\Sigma_{t-1}(nS)^{-1}))}{n+q}nV + F\Sigma_{t-1}F'. \tag{A.3}$$

A.2. Proof of Theorem 3.4. For the case of $F = ERE'$ and $S = EQE'$, we can write eqn. (2.3) as

(A.4)
$$\begin{pmatrix} \Sigma_{t-1} & \Sigma_{t-1}\Upsilon_t' \\ \Upsilon_t\Sigma_{t-1} & \Sigma_t \end{pmatrix} \sim \text{IW}_{2q}\left(n+2, n\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix}\begin{pmatrix} Q & QR \\ QR & Q \end{pmatrix}\begin{pmatrix} E' & 0 \\ 0 & E' \end{pmatrix}\right).$$

Standard theory implies that

(A.5)
$$\begin{pmatrix} E' & 0 \\ 0 & E' \end{pmatrix}\begin{pmatrix} \Sigma_{t-1} & \Sigma_{t-1}\Upsilon_t' \\ \Upsilon_t\Sigma_{t-1} & \Sigma_t \end{pmatrix}\begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \sim \text{IW}_{2q}\left(n+2, n\begin{pmatrix} Q & QR \\ QR & Q \end{pmatrix}\right).$$

The derivation of the conditional mean is exactly as in the general IW-AR case, noting that $\text{tr}(\hat{\Sigma}_{t-1}(nQ)^{-1}) = \sum_i \hat{\Sigma}_{t-1,ii}/(n\xi_i)$.

A.3. Proof of Theorem 3.5. Assume that $||S||_\infty \leq \lambda$, $||S^{-1}||_\infty \leq \lambda$, $||F||_\infty \leq \lambda$, $||\Sigma_0||_\infty \leq \lambda$, and for some $t-1$

(A.6)
$$E[\Sigma_{t-1} \mid \Sigma_0] = F^{t-1}\Sigma_0 F^{t-1\prime} + \frac{n}{n+q}\left(S - F^{t-1}SF^{t-1\prime}\right) + O\left(\frac{1\cdot 1'}{n}\right).$$

To prove that eqn. (A.6) holds for general t, we apply iterated expectations to the conditional expectation of eqn. (3.5):

(A.7)
$$E[\Sigma_t \mid \Sigma_0] = FE[\Sigma_{t-1} \mid \Sigma_0]F' + \frac{n}{n+q}V + \frac{V}{n+q}\text{tr}\left(E[\Sigma_{t-1} \mid \Sigma_0]S^{-1}\right)$$

(A.8)
$$= F^t\Sigma_0 F'^t + \frac{n}{n+q}\left(S - F^tSF'^t\right) + \frac{V}{n+q}\text{tr}\left(E[\Sigma_{t-1} \mid \Sigma_0]S^{-1}\right),$$

where we have used the definition $V = S - FSF'$. Since

(A.9)
$$\frac{V}{n+q}tr\left(E[\Sigma_{t-1} \mid \Sigma_0]S^{-1}\right)$$
$$\leq \frac{V}{n+q}\text{tr}\left\{\left(\lambda^4 1\cdot 1' + \frac{n}{n+q}(1+\lambda^4)1\cdot 1' + \lambda O\left(\frac{1\cdot 1'}{n}\right)\right)1\cdot 1'\right\} = O\left(\frac{1\cdot 1'}{n}\right),$$

we conclude that, indeed,

(A.10)
$$E[\Sigma_t \mid \Sigma_0] = F^t\Sigma_0 F'^t + \frac{n}{n+q}\left(S - F^tSF'^t\right) + O\left(\frac{1\cdot 1'}{n}\right).$$

Then also

(A.11)
$$\lim_{n\to\infty} E[\Sigma_t \mid \Sigma_0] = S + F^t(\Sigma_0 - S)F'^t.$$

A.4. Proof of Theorem 3.6. Let $\Lambda = Q(I - R^2)$ such that $V = E\Lambda E'$. According to eqn. (3.5), $E[\Sigma_1 \mid \Sigma_0]$ then shares the same eigenspace as Σ_0 (i.e., the eigenvectors are given by E), and by induction, so does $E[\Sigma_t \mid \Sigma_0]$ for all t. Let $\Theta_{t|0}$ denote the diagonal matrix of eigenvalues of $E[\Sigma_t \mid \Sigma_0]$. Since $\text{tr}(\Sigma S^{-1}) = \text{tr}(\Theta Q^{-1})$, eqn. (3.5) can be rewritten solely in terms of the eigenvalues:

$$\Theta_{t|0} = \frac{\Lambda}{n+q}\text{tr}\left(\Theta_{t-1|0}Q^{-1}\right) + \frac{n}{n+q}\Lambda + R\Theta_{t-1|0}R'. \tag{A.12}$$

In terms of the vectors of eigenvalues $\theta_{t|0} = \text{diag}\left(\Theta_{t|0}\right)$, $\xi = \text{diag}(Q)$ and $\xi_{-1} = \text{diag}\left(Q^{-1}\right)$ we have

$$\theta_{t|0} = \frac{n}{n+q}(I - R^2)\xi + \left[\frac{1}{n+q}(I - R^2)\xi\xi'_{-1} + R^2\right]\theta_{t-1|0}. \tag{A.13}$$

Letting $\alpha = \frac{n}{n+q}(I - R^2)\xi$ and $B = \frac{1}{n+q}(I - R^2)\xi\xi'_{-1} + R^2$, we conclude that

$$\theta_{t|0} = B^t\theta_0 + \sum_{\tau=0}^{t-1} B^\tau\alpha. \tag{A.14}$$

Since B represents a matrix (strictly) convex combination of $\frac{\xi\xi'_{-1}}{n+q}$ and I, the maximum eigenvalue of B is bounded by

$$\left|\left|(I - R^2)\max\left\{\text{eig}\left(\frac{\xi\xi'_{-1}}{n+q}\right)\right\}\cdot 1 + R^2\cdot 1\right|\right|_0. \tag{A.15}$$

Here, $\max\{\text{eig}(A)\}$ denotes the maximum eigenvalue of A. The term $\xi\xi'_{-1}$ is a rank 1 matrix implying that the only non-zero eigenvalue is equal to $\text{tr}(\xi\xi'_{-1}) = q$. Thus, regardless of n, B has eigenvalues with modulus strictly less than 1 since $\frac{\xi\xi'_{-1}}{n+q}$ has $q-1$ eigenvalues equal to 0 and one equal to $\frac{q}{n+q} < 1$. This implies that the conditional mean of the process forgets the initial condition Σ_0 exponentially fast regardless of n. Furthermore, since the eigenvalues of B have modulus less than 1,

$$\theta_{t|0} = B^t\theta_0 + (I - B)^{-1}(I - B^t)\alpha, \tag{A.16}$$

implying that, as expected, the eigenvalues of the limiting conditional mean are exactly those of the marginal mean S:

$$\lim_{t\to\infty} \theta_{t|0} = (I-B)^{-1}\alpha \tag{A.17}$$

$$= \left[\left(I - \frac{\xi\xi'_{-1}}{n+q}\right)^{-1}(I-R^2)^{-1}\right]\frac{n}{n+q}(I-R^2)\xi \tag{A.18}$$

$$= \frac{n}{n+q}\left(I - \frac{\xi\xi'_{-1}}{n+q}\right)^{-1}\xi \tag{A.19}$$

$$= \xi. \tag{A.20}$$

The last equality follows from matrix inversion and the fact that $\xi'_{-1}\xi = q$.

APPENDIX B: DERIVATION OF FORWARD FILTERING BACKWARD SAMPLING ALGORITHM

B.1. Approximate Forward Filtering. The inverse Wishart prior on Δ_1 can be analytically updated to an inverse Wishart posterior conditioned on y_1:

$$p(\Delta_1 \mid y_1) = \mathrm{IW}_{2q}\left(n+3, n\begin{pmatrix} S & SF' \\ FS & S \end{pmatrix} + y_1 y_1'\right). \tag{B.1}$$

To propagate to $t = 2$, we use the Chapman-Kolmogorov equation, integrating over Δ_1:

$$\begin{aligned} p(\Delta_2 \mid y_1) &\propto \int p(\Delta_2 \mid \Delta_1)p(\Delta_1 \mid y_1)d\Delta_1 \\ &\propto \int \delta_{\Sigma_1 = \Psi_1 + \Upsilon_1\Sigma_0\Upsilon_1'} p(\Psi_2)p(\Upsilon_2 \mid \Psi_2)p(\Delta_1 \mid y_1)d\Upsilon_1 d\Psi_1 d\Sigma_0 \\ &\propto \mathrm{IW}_q(\Psi_2 \mid n+q+2, n\tilde{V})N(\Upsilon_2 \mid \tilde{F}, \tilde{\Psi}_2, (nS)^{-1}) \\ &\qquad \mathrm{IW}_q(\Sigma_1 \mid n+3, nS + x_1x_1'). \end{aligned} \tag{B.2}$$

Here, we have used the fact that the transition kernel $p(\Delta_2 \mid \Delta_1)$ simply involves independent innovations $\{\Upsilon_2, \Psi_2\}$ and deterministically computing Σ_1. Integrating the elements used to compute Σ_1 (a component of Δ_2), the marginal posterior can be derived from the joint posterior of the augmented variance matrix at time t given in eqn. (B.1). Although an independent normal-inverse Wishart set of random variables can be combined with a q-dimensional inverse Wishart matrix to form a $2q$-dimensional inverse Wishart, as discussed in Section 2, there are restrictions on the parameterizations of these respective distributions. The set of distributions specified in eqn. (B.2) do not satisfy these constraints, and thus do not combine to form a $2q$-dimensional inverse Wishart distribution on Δ_2. Namely, the Ψ_2 prior

and Σ_1 posterior degrees of freedom do not match, nor do the Σ_1 posterior scale matrix and the Υ_2 prior variance term $(nS)^{-1}$.

Since exact, analytic forward filtering is not possible, we instead approximate the propagate step with a moment-matched inverse Wishart distribution. That is,

$$\begin{aligned} p(\Delta_2 \mid y_1) &\approx \text{IW}(n+2, nE[\Delta_2 \mid y_1]) \\ &\triangleq g_{2|1}(\Delta_2 \mid y_1). \end{aligned} \tag{B.3}$$

Based on the approximations made in propagating, the subsequent update step is exact due to the conjugacy of the Gaussian observation and inverse Wishart predictive distribution.

In general, we can choose an arbitrary degree of freedom parameter in our approximate forward filtering. Assume that at time t we use r_t degrees of freedom for the moment-matched approximation $g_{t|t-1}(\Delta_t \mid y_{1:t-1})$ to the predictive distribution $p(\Delta_t \mid y_{1:t-1})$. We use $g_{t|t}(\Delta_t \mid y_{1:t})$ to denote the resulting approximation to the updated posterior $p(\Delta_t \mid y_{1:t})$.

We initialize at $t = 1$ with $r_1 = n + 2$ and

$$\begin{aligned} g_{1|0}(\Delta_1) = p(\Delta_1) = \text{IW}(r_1, (r_1-2)E[\Delta_1]), \qquad E[\Delta_1] = \begin{pmatrix} S & SF' \\ FS & S \end{pmatrix}, \\ g_{1|1}(\Delta_1 \mid y_1) = p(\Delta_1 \mid y_1) = \text{IW}(r_1+1, (r_1-2)E[\Delta_1] + y_1 y_1'). \end{aligned} \tag{B.4}$$

Propagating forward,

$$\begin{aligned} g_{2|1}(\Delta_2 \mid y_1) &= \text{IW}(r_2, (r_2-2)E_{g_{1|1}}[\Delta_2 \mid y_1]), \\ g_{2|2}(\Delta_2 \mid y_{1:2}) &= \text{IW}(r_2+1, (r_2-2)E_{g_{1|1}}[\Delta_2 \mid y_1] + y_2 y_2'). \end{aligned} \tag{B.5}$$

Here, the predictive mean is derived as

$$E_{g_{1|1}}[\Delta_2 \mid y_1] = \begin{pmatrix} S_1 & S_1 F' \\ FS_1 & FS_1F' + \frac{nV}{n+q}(1 + \text{tr}(S_1(nS)^{-1})) \end{pmatrix}, \tag{B.6}$$

with $S_1 = E_{g_{1|1}}[\Sigma_1 \mid y_1] = \frac{(r_1-2)S + x_1 x_1'}{r_1 - 1}$. The term $E_{g_{1|1}}[\Sigma_2 \mid y_1]$ is derived via iterated expectations, namely $E_{g_{1|1}}[E[\Sigma_2 \mid \Sigma_1, y_1]]$, and employing eqn. (3.5) noting that Σ_2 is conditionally independent of y_1 given Σ_1.

The forward filter then recursively defines

$$\begin{aligned} g_{t|t-1}(\Delta_t \mid y_{1:t-1}) &= \text{IW}\left(r_t, (r_t-2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}]\right), \\ g_{t|t}(\Delta_t \mid y_{1:t}) &= \text{IW}\left(r_t+1, (r_t-2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] + y_t y_t'\right), \end{aligned} \tag{B.7}$$

with

$$E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] = \begin{pmatrix} S_{t-1} & S_{t-1}F' \\ FS_{t-1} & FS_{t-1}F' + \frac{nV}{n+q}(1 + \text{tr}(S_{t-1}(nS)^{-1})) \end{pmatrix} \tag{B.8}$$

and

$$S_t = \frac{(r_t - 2)\left(FS_{t-1}F' + \frac{nV}{n+q}(1 + \text{tr}(S_{t-1}(nS)^{-1}))\right) + x_t x_t'}{r_t - 1}. \tag{B.9}$$

B.2. Backward Sampling. The density required for backward sampling is the posterior of $\{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$ conditioned on Σ_t and $y_{1:T}$, which can be written as

$$p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid \Sigma_t, y_{1:T}) = p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid \Sigma_t, y_{1:t}) \tag{B.10}$$

$$\propto p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid y_{1:t})\delta_{\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'} \tag{B.11}$$

$$\propto g_{t|t}(\Delta_t \mid y_{1:t})\delta_{\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'}. \tag{B.12}$$

Thus, sampling $\{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$ from this conditional posterior is equivalent to fixing Σ_t in the Δ_t matrix and sampling a valid $\{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$ conditioned on Σ_t based on the forward filtering distribution $g_{t|t}(\Delta_t \mid y_{1:t})$. By *valid*, we mean a value that corresponds to $\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t'$.

Based on eqn. (5.5),

$$g_{t|t}\left(\left.\begin{pmatrix} \Sigma_{t-1} & \Sigma_{t-1}\Upsilon_t' \\ \Upsilon_t \Sigma_{t-1} & \Sigma_t \end{pmatrix}\right| y_{1:t}\right) = \text{IW}\left(r_t + 1, \begin{pmatrix} G_t^{11} & G_t^{21'} \\ G_t^{21} & G_t^{22} \end{pmatrix}\right), \tag{B.13}$$

implying that

$$g_{t|t}\left(\left.\begin{pmatrix} \Sigma_t & \Upsilon_t \Sigma_{t-1} \\ \Sigma_{t-1}\Upsilon_t' & \Sigma_{t-1} \end{pmatrix}\right| y_{1:t}\right) = \text{IW}\left(r_t + 1, \begin{pmatrix} G_t^{22} & G_t^{21} \\ G_t^{21'} & G_t^{11} \end{pmatrix}\right). \tag{B.14}$$

Here, G_t is the forward filtering term defined in eqn. (5.11), with G_t^{11}, G_t^{21}, G_t^{22} denoting the three unique $q \times q$ sub-blocks ($G_t^{12} = G_t^{21'}$). The form of eqn. (B.14) allows us to use the previously discussed properties of the inverse Wishart distribution to sample $\{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$ conditioned on Σ_t and $y_{1:t}$. Specifically, as discussed in Section 2, there exists a $\{\tilde{\Upsilon}_t, \tilde{\Psi}_t\}$ such that $\tilde{\Upsilon}_t \Sigma_t = \Sigma_{t-1}\Upsilon_t'$ and

$$\begin{aligned} \tilde{\Psi}_t \mid y_{1:t} &\sim \text{IW}(r_t + 1 + q, G_t^{11} - G_t^{21'}(G_t^{22})^{-1}G_t^{21}), \\ \tilde{\Upsilon}_t \mid \tilde{\Psi}_t, y_{1:t} &\sim N(G_t^{21'}(G_t^{22})^{-1}, \tilde{\Psi}_t, (G_t^{22})^{-1}) \end{aligned} \tag{B.15}$$

with

$$\Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Upsilon}_t \Sigma_t \tilde{\Upsilon}_t'. \tag{B.16}$$

Thus, to sample Σ_{t-1} conditioned on Σ_t from the approximation to $p(\Sigma_{t-1} \mid \Sigma_t, y_{1:t})$, we first sample $\{\tilde{\Upsilon}_t, \tilde{\Psi}_t\}$ as specified in eqn. (B.15) and then compute Σ_{t-1} based on eqn. (B.16).

APPENDIX C: MULTIVARIATE T DISTRIBUTION

The q-dimensional multivariate t distribution with ν degrees of freedom and parameters μ and Σ has density

$$t_\nu(x \mid \mu, \Sigma/\nu) = a_{\nu,q} |\Sigma|^{-1/2} \left(1 + \frac{1}{\nu}(x-\mu)'\Sigma^{-1}(x-\mu)\right)^{-(\nu+q)/2} \tag{C.1}$$

where

$$a_{q,\nu} = \frac{\Gamma((\nu+q)/2)}{\Gamma(\nu/2)\nu^{q/2}\pi^{q/2}}.$$

For the proposed IW-AR, since $\Upsilon_t \mid \Psi_t \sim N(F, \Psi_t, (nS)^{-1})$ standard theory gives $\Upsilon_t z_t \mid \Psi_t, z_t \sim N(Fz_t, \Psi_t(z_t'(nS)^{-1}z_t))$. Marginalizing over Υ_t in eqn. (4.1), we have

$$p(x_t \mid z_t, \Psi_t) = N(Fz_t, \Psi_t(1 + z_t'(nS)^{-1}z_t)). \tag{C.2}$$

Now, $\Psi_t \sim IW(n+q+2, nV)$ implies $\Psi_t(1 + z_t'(nS)^{-1}z_t) \mid z_t \sim IW(n+q+2, nV(1 + z_t'(nS)^{-1}z_t))$ Marginalizing Ψ_t from the distribution of eqn. (C.2) yields the t distribution with density

$$p(x_t \mid z_t) = t_{n+q+2}\left(Fz_t, (1 + z_t'(nS)^{-1}z_t)\frac{nV}{n+q+2}\right). \tag{C.3}$$

The marginal likelihood of F and S given $z_{1:T}$ follows immediately.

APPENDIX D: CONDITIONAL MEAN OF IW-AR(2) MODEL

For the IW-AR(2) process in Section 7.2 we have

$$E[\Psi_t \mid \Psi_{t-1}] = H\Psi_{t-1}H' + \frac{nW}{n+2q}\left(1 + \mathrm{tr}(\Psi_{t-1}(nV)^{-1})\right). \tag{D.1}$$

Since $E[\Sigma_t \mid \Delta_{t-1}] = E[\Psi_t \mid \Delta_{t-1}] + E[\Upsilon_t \Sigma_{t-1} \Upsilon_t' \mid \Delta_{t-1}]$, in deriving the conditional mean of Σ_t given Δ_{t-1} we first need

(D.2)
$$\begin{aligned} E[[\Upsilon_t \Sigma_{t-1} \Upsilon_t']_{ij} \mid \Delta_{t-1}] &= E[E[[\Upsilon_t \Sigma_{t-1} \Upsilon_t']_{ij} \mid \Psi_t] \mid \Delta_{t-1}] \\ &= E[\sum_{k,\ell} \Sigma_{t-1,k\ell} E[\Upsilon_{t,ik} \Upsilon_{t,j\ell} \mid \Psi_t] \mid \Delta_{t-1}] \\ &= E[\sum_{k,\ell} \Sigma_{t-1,k\ell} \left([\Psi_{t,ij}(nS)^{-1}]_{k\ell} + F_{ik} F_{j\ell}\right) \mid \Delta_{t-1}] \\ &= E[\mathrm{tr}(\Sigma_{t-1}(nS)^{-1}) \Psi_{t,ij} + F_{i\cdot} \Sigma_{t-1} F_{j\cdot}' \mid \Delta_{t-1}] \\ &= \mathrm{tr}(\Sigma_{t-1}(nS)^{-1}) E[\Psi_{t,ij} \mid \Delta_{t-1}] + F_{i\cdot} \Sigma_{t-1} F_{j\cdot}', \end{aligned}$$

implying

$$E[\Upsilon_t \Sigma_{t-1} \Upsilon_t' \mid \Delta_{t-1}] = \mathrm{tr}(\Sigma_{t-1}(nS)^{-1}) E[\Psi_t \mid \Delta_{t-1}] + F \Sigma_{t-1} F'. \tag{D.3}$$

Noting that $E[\Psi_t \mid \Delta_{t-1}] = E[E[\Psi_t \mid \Psi_{t-1}] \mid \Delta_{t-1}]$ and $E[\Psi_{t-1} \mid \Delta_{t-1}] = \Psi_{t-1}$ since Ψ_{t-1} is a deterministic function of the elements of Δ_{t-1}, eqn. (7.19) follows directly. That is,

$$E[\Sigma_t \mid \Delta_{t-1}] = F \Sigma_{t-1} F' + E[\Psi_t \mid \Psi_{t-1}](1 + \mathrm{tr}(\Sigma_{t-1}(nS)^{-1}))$$

with $E[\Psi_t \mid \Psi_{t-1}]$ as in eqn. (D.1).

In the limit as $n \to \infty$, we have

$$\begin{aligned} E[\Sigma_t \mid \Delta_{t-1}] &= W + F \Sigma_{t-1} F' + H \Psi_{t-1} H' \\ &= V - HVH' + F \Sigma_{t-1} F' + H(\Sigma_{t-1} - H \Upsilon_{t-1} \Sigma_{t-2} \Upsilon_{t-1}') H' \\ &= S - FSF' + F \Sigma_{t-1} F' \\ &\qquad + H((\Sigma_{t-1} - \Upsilon_{t-1} \Sigma_{t-2} \Upsilon_{t-1}') - (S - FSF'))H' \\ &= S + F(\Sigma_{t-1} - S)F' \\ &\qquad + H((\Sigma_{t-1} - S) - (\Upsilon_{t-1} \Sigma_{t-2} \Upsilon_{t-1}' - FSF'))H'. \end{aligned}$$

APPENDIX E: SAMPLING FOR STOCHASTIC VOLATILITY IN TIME SERIES MODELS WITH IW-AR(1) COMPONENTS

In eqn. (6.1), the conditional posterior of the autoregressive parameters A (i.e., *Step 0* of the sampler) is given as follows.

Step 0. Sample the observation autoregressive parameter A given $\Delta_{1:T}$ and $\xi_{1-r:T}$. Assume A_i diagonal defined by the q-vector $a_i = \mathrm{diag}(A_i)$. The autoregressive process of eqn. (6.1) can be equivalently represented as

$$\xi_t = \begin{bmatrix} \mathrm{diag}(\xi_{t-1}) & \cdots & \mathrm{diag}(\xi_{t-r}) \end{bmatrix} \begin{bmatrix} a_1' & \cdots & a_r' \end{bmatrix}' + x_t. \tag{E.1}$$

Under a prior $\begin{bmatrix} a_1' & \cdots & a_r' \end{bmatrix}' \sim N(\mu_a, \Sigma_a)$, standard theory yields the conditional for $\begin{bmatrix} a_1' & \cdots & a_r' \end{bmatrix}' \mid \Sigma_{0:T}, \xi_{1-r:T}$ as multivariate normal with easily computed moments. For $t = 1, \ldots, T$, set $x_t = \xi_t - \sum_{i=1}^{r} A_i \xi_{t-i}$.

REFERENCES

O. Aguilar and M. West. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics*, 18:338–357, 2000.

O. Aguilar, R. Prado, G. Huerta, and M. West. Bayesian inference on latent structure in time series (with discussion). In J.O. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 6*, pages 3–26. Oxford University Press, 1999.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte carlo methods. *Journal of the Royal Statistical Society, Series B*, 72(3):269–342, 2010.

C. M. Carvalho, M. Johannes, H. F. Lopes, and N. G. Polson. Particle learning and smoothing. *Statistical Science*, 25:88–106, 2010.

C.M. Carvalho, H. Massam, and M. West. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659, 2007.

S. Chib, Y. Omori, and M. Asai. Multivariate stochastic volatility. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*, pages 365–400. Springer-Verlag, Berlin Heidelberg, 2009.

T. Cogley and T. J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics*, 8:262–302, 2005.

A.P. Dawid and S.L. Lauritzen. Hyper-Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 3:1272–1317, 1993.

R. Engle. New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446, 2002.

J.A. Fill and M.L. Huber. The randomness recycler: A new technique for perfect sampling. In *Proc. 41th Annual IEEE Symposium on the Foundations of Computer Science*, pages 503–511, 2001.

J.M. Freyermuth, H. Ombao, and R. von Sachs. Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *Journal of the American Statistical Association*, 105 (490):634–646, 2010.

C. Gouriéroux, J. Jasiak, and R. Sufana. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181, 2009.

A.C. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

A. D. Krystal, S. Zoldi, R. Prado, H. S. Greenside, and M. West. The spatiotemporal dynamics of generalized tonic-clonic seizure EEG data: Relevance to the climinal practice of electroconvulsive therapy. In N. Pradhan, P.E. Rapp, and R. Sreenivasan, editors, *Nonlinear Dynamics and Brain Functioning*. New York: Novascience, 1999a.

A. D. Krystal, M. West, R Prado, H.S. Greenside, S. Zoldi, and R.D. Weiner. The EEG effects of ECT: Implications for rTMS. *Depress Anxiety*, 12:157–165, 2000.

A.D. Krystal, R. Prado, and M. West. New methods of time series analysis for non-stationary EEG data: Eigenstructure decompositions of time varying autoregressions. *Clinical Neurophysiology*, 110(12):2197–2206, 1999b.

J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In A. Doucet, J.F.G. De Freitas, and N.J. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 197–217. New York: Springer-Verlag, 2001.

H. F. Lopes, C. M. Carvalho, M. Johannes, and N. G. Polson. Particle learning for sequential Bayesian computation (with discussion). In J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid,

D. Heckerman, A.F.M. Smith, and M. West, editors, *Bayesian Statistics 9*, pages 317 – 360. Oxford University Press, 2010a.

H. F. Lopes, R. E. McCulloch, and R. Tsay. Cholesky stochastic volatility, booth chicago. Technical report, 2010b.

R.H. Mena and S.G. Walker. Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, 26(6):789–805, 2005.

J. Nakajima and M. West. Bayesian dynamic factor models: Latent threshold approach, duke statistical science. Technical report, 2011.

H. Ombao, R. von Sachs, and W. Guo. SLEX analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531, 2005.

A. Philipov and M.E. Glickman. Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics*, 24(3):313–328, 2006a.

A. Philipov and M.E. Glickman. Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews*, 25(2-3):311–334, 2006b.

M.K. Pitt and S.G. Walker. Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, 100(470):554–564, 2005.

M.K. Pitt, C. Chatfield, and S.G. Walker. Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29(4):657–663, 2002.

M. Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.

R. Prado and M. West. Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions. In T. Gregoire, editor, *Modelling Longitudinal and Spatially Correlated Data*, pages 349–362. Springer-Verlag, 1997.

R. Prado and M. West. *Time Series: Modelling, Computation and Inference*. Chapman & Hall/CRC Press, The Taylor Francis Group, 2010. ISBN 978-1-420093360.

R. Prado, M. West, and A. D. Krystal. Multi-channel EEG analyses via dynamic regression models with time-varying lag/lead structure. *Journal of the Royal Statistical Society (Ser. C)*, 50:95–110, 2001.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72:821–852, 2005.

J. M. Quintana and M. West. An analysis of international exchange rates using multivariate DLMs. *The Statistician*, 36:275–281, 1987.

J.M. Quintana. Optimal portfolios of forward currency contracts. In J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics IV*, pages 753–762. Oxford University Press, 1992.

J.M. Quintana, V. Lourdes, O. Aguilar, and J. Liu. Global gambling. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics VII*, pages 349–368. Oxford University Press, 2003.

J.M. Quintana, C. M. Carvalho, J. Scott, and T. Costigliola. Futures markets, Bayesian forecasting and risk modelling. In A. O'Hagan and M. West, editors, *The Handbook of Applied Bayesian Analysis*, pages 343–365. Oxford University Press, 2010.

H. Uhlig. On singular Wishart and singular multivariate beta distributions. *Annals of Statistics*, 22: 395–405, 1994.

M. West and P.J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 1997.

M. West, R. Prado, and A. D. Krystal. Evaluation and comparison of EEG traces: Latent structure in non-stationary time series. *Journal of the American Statistical Association*, 94:1083–1095, 1999.

EMILY FOX AND MIKE WEST
DEPARTMENT OF STATISTICAL SCIENCE
DUKE BOX #90251
DURHAM NC 27708-0251
URL: http://www.stat.duke.edu
E-MAIL: (fox,mw)@stat.duke.edu