AUTOREgressive Models for Variance Matrices: Stationary Inverse Wishart Processes

BY EMILY B. FOX*

University of Washington, Seattle WA, USA

AND

BY 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 primary class of models explored is a 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 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.

*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
The context is a sequence of $q \times q$ variance matrices (i.e., symmetric, non-negative definite matrices) $\Sigma_t$ in discrete time $t = 0, 1, \ldots$, typically the variance matrices of components of more elaborate state-space models for an observable 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, Ruiz and Shephard, 1994; Quintana and West, 1987) are widely used. While the former directly specify “volatility” matrices $\Sigma_t$ 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, Omori and Asai (2009) and ? for surveys of such approaches. The class of state-space models based on Bayesian discount methods (Quintana, 1992; Quintana and West, 1987; 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 $\Sigma_t$; see Prado and West (2010) for a recent review.

Two recent contributions explore constructions of AR(1) style models based on conditional Wishart transition distributions (Gouri´eroux, Jasiak and Sufana, 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. Employing a Wishart transition automatically ensures the symmetry and positivity of the variance matrices, and has origins in the continuous-time Wishart processes proposed by ?. We discuss these and related approaches further in Section 8.

Instead of defining Wishart transitions, we introduce a new class of AR(1) style processes that maintain inverse Wishart margins. The centrality of inverse Wishart theory to current Bayesian state-space approaches underlies the ideas for the new model classes explored in this paper. Specifically, the inverse Wishart distribution provides a conjugate prior to multivariate normal likelihoods with unknown covariance. Furthermore, the inverse Wishart distribution yields extensions to high-dimensional analysis of variance matrices constrained by specified graphical models via the hyper-inverse Wishart distribution (Carvalho, Massam and West, 2007; Dawid and Lauritzen, 1993). Just as motivated in Gouriéroux, Jasiak and Sufana (2009); Philipov and Glickman (2006a,b), the inverse Wishart distribution ensures the symmetry and non-negativity of the variance matrices.

The key challenge addressed in this paper is in defining a class of stationary autoregressive processes that maintain the desired inverse Wishart margins. To do this, we exploit the structure of conditional and marginal distributions in the inverse Wishart family. In addition to the computational advantages garnered via inverse
Wishart margins, two of the main attributes of the resulting autoregression are its (i) openness to theoretical analysis and (ii) interpretability. In particular, our proposed process yields a form analogous to a standard random coefficient autoregressive process; here, of course, constrained to evolve on the cone of symmetric, non-negative definite matrices. The induced conditional transition distribution also has an interpretation as a mixture of inverse Wisharts. We denote the resulting models by IW-AR, and use 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. In terms of theoretical analysis, the IW-AR readily enables the establishment of a number of properties that are fundamental to time series analysis in a range of application domains. First, stationarity is easily assessed by construction. Furthermore, we can examine time-reversibility, invariance to transformations, 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 result in tractable Bayesian computations. Our sampling 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 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. Proofs of theoretical results and extended figures from the EEG example analysis are included as Supplementary Material.

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

\[(2.1) \quad x_t | \Sigma_t \sim N(0, \Sigma_t), \quad t = 1 : T,\]

where $x_t$ is independent of $\{x_s, \Sigma_s; s < t\}$ conditional on $\Sigma_t$. We aim to capture the volatility dynamics with a stationary, first-order Markov model for the $\Sigma_t$ sequence. Any first-order Markov process for $\Sigma_t$ must satisfy

\[(2.2) \quad p(\Sigma_{0:T}) = p(\Sigma_0) \prod_{t=1}^{T} p(\Sigma_t | \Sigma_{t-1}),\]

...
assuming that the sequence of conditional densities $p(\Sigma_t \mid \Sigma_{t-1})$ exists. Instead of explicitly defining the conditional distribution, as in Gouriéroux, Jasiak and Sufana (2009); Philipov and Glickman (2006b), one can induce the transition distribution by specifying pairwise joint distributions with a common margins (i.e., $p(\Sigma_{t-1}, \Sigma_t)$ and $p(\Sigma_t, \Sigma_{t+1})$ both consistent on margin $p(\Sigma_t)$). The joint density of eqn. (2.2) is then equivalently represented as

$$p(\Sigma_0; T) = \frac{\prod_{t=1}^{T} p(\Sigma_{t-1}, \Sigma_t)}{\prod_{t=2}^{T} p(\Sigma_{t-1})}. \tag{2.3}$$

Such an approach is appealing when there is a target form for the marginal distribution, and is somewhat related in concept to transitions built from copulas (cf. ?), but more closely to decomposable graphical models such as junction trees (cf. ?). Likewise, stationarity of the resulting autoregressive process is trivially satisfied by taking $p(\Sigma_{t-1}, \Sigma_t)$ to be time invariant. Section 3.2 includes further analysis of stationarity.

To construct an autoregressive process with inverse Wishart margins, 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 $\phi_t$ such that

$$\left( \begin{array}{c} \Sigma_{t-1} \\ \phi_t \\ \Sigma_t \end{array} \right) \sim \text{IW}_{2q} \left( n + 2, n \left( \begin{array}{cc} S & SF' \\ FS & S \end{array} \right) \right) \tag{2.4}$$

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 $2q \times 2q$ scale matrix parameter of the distribution above is non-negative definite. See Section 3.2 for conditions on $S$ and $F$, and for example specifications. Details on the inverse Wishart density are in the Supplementary Material. The inverse Wishart of eqn. (2.4) has common margin for the diagonal blocks; for each $t$,

$$\Sigma_t \sim \text{IW}_q(n + 2, nS) \tag{2.5}$$

with $E[\Sigma_t] = S$. From eqn. (2.3), it is now clear that the specified inverse Wishart form defines a stationary first-order process with eqn. (2.5) as the stationary (marginal) distribution. See Theorem 3.1. The transitions are governed by the conditional density $p(\Sigma_t \mid \Sigma_{t-1})$ implicitly defined by eqn. (2.4); 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. (2.4) 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 $\Sigma_{t-1}$.
Specifically, standard inverse Wishart theory (cf. Carvalho, Massam and West, 2007) implies that

\[ \Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t' \]

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

\[ \Psi_t \sim \text{IW}_q(n + q + 2, nV) \]

\[ \Upsilon_t \mid \Psi_t \sim \mathcal{N}(F, \Psi_t, (nS)^{-1}). \]

Here, \( V = S - FSF' \) and \{\( \Upsilon_t, \Psi_t \}\) are conditionally independent of \( \Sigma_{t-1} \sim \text{IW}_q(n + 2, nS) \). Further details on the matrix normal distribution are provided in the Supplemental Material. Eqn. (2.6) is an explicit AR(1) equation in which \( \Upsilon_t \) acts as a random autoregressive coefficient matrix and \( \Psi_t \) 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.

Another interpretation of the Markov transitions is as a mixture of inverse Wisharts. In particular,

\[ p(\Sigma_t \mid \Sigma_{t-1}) = \int p(\Sigma_t \mid \Sigma_{t-1}, \phi_t) p(\phi_t \mid \Sigma_{t-1}) d\phi_t, \]

where, as derived in the Supplementary Material,

\[ \Sigma_t \mid \Sigma_{t-1}, \phi_t \sim \text{IW}_q(n + q + 2, nV + (n + q)\phi_t \Sigma_{t-1} \phi_t'). \]

The mixing distribution is defined via \( p(\phi_t' \mid \Sigma_{t-1}) \) being a matrix T distribution.

2.3. Special Case of \( q = 1 \). When \( q = 1 \), \( \Sigma_t \equiv \sigma_t^2 > 0 \) and the IW-AR process reduces to an inverse gamma autoregressive process. Now \( S \equiv s > 0 \) and \( F \equiv f \in (-1, 1) \) ensures the positivity of the \( 2 \times 2 \) scale matrix of the joint density of eqn. (2.4), which simplifies to

\[ \left( \begin{array}{cc} \sigma_{t-1}^2 & \phi_t \\ \phi_t & \sigma_t^2 \end{array} \right) \sim \text{IW}_2(n + 2, ns \left( \begin{array}{cc} 1 & f \\ f & 1 \end{array} \right)), \]

such that

\[ \sigma_t^2 \sim IG\left(\frac{n + 2}{2}, \frac{ns}{2}\right). \]

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

\[ \sigma_t^2 = \psi_t + \upsilon_t^2 \sigma_{t-1}^2 \]
where
\begin{equation}
\psi_t \sim IG \left( \frac{n + 3}{2}, \frac{ns(1 - f^2)}{2} \right) \quad \text{and} \quad v_t | \psi_t \sim N \left( f, \frac{\psi_t}{ns} \right).
\end{equation}

We can see immediate analogies with the standard linear, Gaussian AR(1) process with a random AR coefficient. The marginal mean of $v^2_t$ 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, Chatfield and Walker (2002). In that work, the authors construct a stationary autoregressive process $\{\sigma^2_t\}$ with inverse gamma marginals by harnessing a conditionally gamma distributed latent process $\{z_t\}$. The sequence $\{\sigma^2_t, z_t\}$ obtained by generating $\sigma^2_t | z_{t-1}$ and $z_t | \sigma^2_t$ from the respective closed-form conditional distributions leads to a marginal process $\{\sigma^2_t\}$ 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 of this and other related approaches appears in Section 8.

We now examine some theoretical properties of the IW-AR in Section 3, including stationarity, time-reversibility, invariance to transformations, and the conditional mean for finite and infinite horizons. Subsequent sections focus on inference of a latent IW-AR based on observations $x_t$, $(t = 1, 2, \ldots)$: Section 4 presents a data augmentation technique that is pivotal for the MCMC-based posterior computations outlined in Section 5, and Section 6 presents an analysis of EEG data.

### 3. Theoretical Properties.


Consider any partition of $\Sigma_t$ 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
INVERSE WISHART PROCESSES

submatrix \( \Sigma_{t,ij} \) depends upon every element of \( \Sigma_{t-1} \) as follows:

\[
\Sigma_{t,ij} = \Psi_{t,ij} + \begin{bmatrix} \Upsilon_{t,i1} & \ldots & \Upsilon_{t,iJ} \end{bmatrix} \Sigma_{t-1} \begin{bmatrix} \Upsilon_{t,j1} & \ldots & \Upsilon_{t,jI} \end{bmatrix}'.
\]

(3.1)

Here \( \Psi_{t,ij} \sim \text{IW}(n+q+2, nV_{ij}) \) and \( \begin{bmatrix} \Upsilon_{t,i1} & \ldots & \Upsilon_{t,iJ} \end{bmatrix} \) has a conditional matrix normal distribution induced from the joint distribution of eqn. (2.7).

3.2. Stationarity.

**Theorem 3.1.** The process defined via eqn. (2.4) is strictly stationary when the parameterization of the inverse Wishart of eqn. (2.4) yields a valid distribution: that is, when \( S \) and \( S - FSF' \) are positive definite.

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

In the case of \( F = \rho I_q \), the stationarity condition reduces to \( S \) positive definite and \( |\rho| < 1 \). Another special case is when \( F, S \) share eigenvectors with eigen-decompositions \( F = ERE' \) and \( S = EQE' \). Stationarity is assured when \( Q = \text{diag}(\xi_1, \ldots, \xi_q) \) has positive elements and \( R = \text{diag}(\rho_1, \ldots, \rho_q) \) has each \( |\rho_i| < 1 \).

The proof of Theorem 3.1, detailed in the Supplementary Material, follows straightforwardly from considering eqn. (2.3) and (2.4) on an arbitrary time window \( t : t + \tau, \tau > 0 \). For approaches based on specification of a time-invariant transition distribution, such as in Gouriéroux, Jasiak and Sufana (2009), we see from eqn. (2.2) that proving the time-invariance of \( p(\Sigma_t) \) is necessary and sufficient for proving stationarity. Typically, analysis of the stationary margin relies on asymptotic arguments. In contrast our proof of stationarity follows directly from specifying time-invariant pairwise joints with common (time-invariant) margins.

3.3. Reversibility. The IW-AR readily yields analysis of the time-reversibility of the defined process as a simple function of the hyperparameters \( F \) and \( S \). Such analysis is important to the generality of the IW-AR in a variety of application domains; e.g., many financial time series are assumed to be time-irreversible (?). We include the proof of reversibility for Theorem 3.2 because the details will feature in the Bayesian model fitting of Section 5.

**Theorem 3.2.** The process is time-reversible if and only if \( FS = SF' \).

**Proof.** The reverse-time process on the \( \Sigma_t \) is derived as follows. Eqn. (2.4) implies that

\[
\left( \begin{array}{c} \Sigma_t \\ \hat{\phi}_t \\ \Sigma_{t-1} \end{array} \right) \sim \text{IW}_{2q} \left( n + 2, n \left( \begin{array}{cc} S & FS \\ SF' & S \end{array} \right) \right),
\]

(3.2)
for some latent process $\tilde{\phi}_t$. Then, as in Section 2, we have

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

where

$$\tilde{\Psi}_t \sim \text{IW}_q(n + q + 2, n \tilde{V}) \quad \text{and} \quad \tilde{\Upsilon}_t | \tilde{\Psi}_t \sim N(\tilde{F}, \tilde{\Psi}_t, (nS)^{-1})$$ (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. A key benefit of the IW-AR is the ability to switch between reversible or irreversible stationary processes simply by changing the hyperparameter specification.

### 3.4. Transformed IW-AR Processes

Theorem 3.3 shows that the IW-AR process is invariant under invertible quadratic transformations.

**Theorem 3.3.** Let $\Sigma_t$ follow an IW-AR(1) process with parameter $\{n, F, S\}$ and take $A$ to be any invertible matrix. Then, $\bar{\Sigma}_t = A^t \Sigma_t A$ follows an IW-AR(1) process defined by parameters $\{n, \bar{F} = A^t F (A')^{-1}, \bar{S} = A^t S A\}$.

Assuming observations $x_t$ with variance matrices $\Sigma_t$ following an IW-AR(1), as in eqn. (2.1), Theorem 3.3 implies that $y_t = A^t x_t$ has variance matrices $A^t \Sigma_t A$ that also follow a transformed IW-AR(1). Invariance to such transformations of the data is useful in a wide range of applications, including financial time series. For example, taking $x_t$ to represent $q$ base assets and $y_t$ a collection of $q$ portfolios with various allocations determined by $A$, we see that the IW-AR is invariant to portfolio allocation. Likewise, in electroencephalogram (EEG) studies based on a collection of $q$ spatially distributed electrodes as examined in Section 6, the IW-AR is invariant to permutation or rotational transformations of electrodes.

**Whitened IW-AR Processes.** As provided by Corollary 3.1, for any $\Sigma_t$ following an IW-AR(1), there exists a whitened IW-AR process $\hat{\Sigma}_t$ such that $E[\hat{\Sigma}_t] = I$ where $\hat{\Sigma}_t$ is a quadratic transformation of $\Sigma_t$. One can view this IW-AR as a canonical representation of the process.

**Corollary 3.1.** Let $\Sigma_t$ follows an IW-AR(1) with parameters $\{n, F, S\}$ and let $S^{1/2}$ be any matrix square root of $S$. Then, $\hat{\Sigma}_t = S^{-1/2} \Sigma_t (S^{-1/2})'$ follows an IW-AR(1) with parameters $\{n, \bar{F} = S^{-1/2} FS^{1/2}, \bar{S} = I\}$. Thus, marginally $\hat{\Sigma}_t \sim \text{IW}_q(n + 2, nI)$. 

3.5. Conditional Mean. The IW-AR yields a simple form for the conditional expectation of $\Sigma_t$ given $\Sigma_{t-1}$.

**Theorem 3.4.**

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

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

Theorem 3.4 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

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

Intuitively, here the conditional mean is centered about the marginal mean $S$ plus the propagation of the previous deviation from the marginal mean through the mean autoregressive dynamic defined by $F$.

Using iterated expectations, one can straightforwardly compute $E[\Sigma_{t+h} \mid \Sigma_t]$ for some finite horizon $h$. Theorem 3.5 shows that under mild conditions on $F$ and $S$, this conditional mean forecast has a form analogous to that for standard autoregressions, which is especially clear in the univariate ($q = 1$) case. Section 3.6 explores the behavior of $E[\Sigma_t \mid \Sigma_0]$ as the horizon increases relative to a fixed initial variance $\Sigma_0$. The special case of $E[\Sigma_{t+h} \mid \Sigma_t]$ follows directly with a simple change of time indices.

3.6. Exponential Forgetting. In the following, we show that the mean of the IW-AR process forgets its initial condition exponentially fast under a wide range of conditions on the parameterization $\{n, F, S\}$. The form of the condition mean forecast $E[\Sigma_t \mid \Sigma_0]$ is provided in Theorem 3.5.

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

\[
E[\Sigma_t \mid \Sigma_0] = F^t\Sigma_0F't + c_{n,q}(S - F^tS)F't + O(n^{-1}1 \cdot 1')
\]

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

\[
\lim_{n \to \infty} E[\Sigma_t \mid \Sigma_0] = S + F^t(S_0 - S)F't.
\] (3.8)
Unitary Bounded Spectral Radius of $F$. If we further 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 $E[\sigma_t^2 | \sigma_0^2]$ without relying on the limit of $n \to \infty$. Using the notation of Section 2.3 and recursing on the form of $E[\sigma_t^2 | \sigma_{t-1}^2]$ specified in Theorem 3.4,

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

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

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

Eqn. (3.10) has the form of a linear AR(1) model with AR parameter $(nf^2 + 1)/(n + 1)$. Then $\lim_{t \to \infty} E[\sigma_t^2 | \sigma_0^2] = s$ when $|f| < 1$, and does so exponentially fast regardless of $n$. The overall rate of this exponential forgetting is governed by the AR parameter $(nf^2 + 1)/(n + 1)$.

Shared Eigenvectors Between $F$, $S$. Assume a reversible IW-AR process based on a specification with $F$ and $S$ having shared eigenvectors, $E$. In Corollary 3.2, we show that this process forgets its initial condition $\Sigma_0$ in expectation, and does so exponentially fast. Here again we do not rely on examining limits of $n$.

The proof harnesses a transformed IW-AR process $\tilde{\Sigma}_t = E\Sigma_tE$, in which the random autoregressive coefficients $\{\tilde{\Psi}_t, \tilde{\Upsilon}_t\}$ are diagonal in expectation. Some properties of this process are presented in Theorem 3.6. Due to the one-to-one mapping between $\Sigma_t$ and $\tilde{\Sigma}_t$, showing that $\tilde{\Sigma}_t$ forgets $\Sigma_0$ in the mean allows us to conclude the same for the original $\Sigma_t$ process.

THEOREM 3.6. Suppose that $\Sigma_t$ follows an IW-AR(1) with parameters $\{n, F = ERE', S = EQE'\}$ where $E$ is orthogonal, $R = \text{diag}(\rho_1, \ldots, \rho_q)$, and $Q = \text{diag}(\xi_1, \ldots, \xi_q)$ with positive elements. Then $\tilde{\Sigma}_t = E\Sigma_tE$ follows an IW-AR(1) model with parameters $\{n, R, Q\}$. Specifically, letting $\Sigma_t$ be defined as in eqn. (2.6),

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

where $\tilde{\Psi}_t = E\Psi_tE$ and $\tilde{\Upsilon}_t = E\Upsilon_tE$ are such that

$$\tilde{\Psi}_t \sim IW_q(n + q + 2, nQ(I - R^2)) \quad \text{and} \quad \tilde{\Upsilon}_t \mid \tilde{\Psi}_t \sim N(R, \tilde{\Psi}_t, (nQ)^{-1})$$
and with marginal distribution \( \hat{\Sigma}_t \sim IW_q(n + 2, nQ) \). The conditional mean is

\[
E[\hat{\Sigma}_t | \hat{\Sigma}_{t-1}] = R\hat{\Sigma}_{t-1}R + c_{n,q} \left[ 1 + \sum \left( n\xi_i \right)^{-1} \hat{\Sigma}_{t-1,ii} \right] Q(I - R^2),
\]

where \( c_{n,q} = n/(n + q) \). Assuming a stationary process such that each \( |\rho_i| < 1 \),

\[
\lim_{t \to \infty} E[\hat{\Sigma}_t | \hat{\Sigma}_0] = Q.
\]

Convergence is exponentially fast.

**Corollary 3.2.** Assuming that \( \Sigma_t \) follows a stationary IW-AR(1) with parameters \( \{n, F = ERE', S = EQE'\} \) as in Theorem 3.6,

\[
\lim_{t \to \infty} E[\Sigma_t | \Sigma_0] = S,
\]

and convergence is exponentially fast.

### 4. Data Augmentation.

We now turn our attention to inference of a latent IW-AR process based on observations following eqn. (2.1). 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 \( \Sigma_0 \) 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

\[
x_t | z_t \sim N(\Upsilon_t z_t, \Psi_t) \quad \text{and} \quad z_t \sim N(0, \Sigma_{t-1})
\]

independently over time. Marginalizing \( z_t \) and recalling that \( \Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Upsilon_t' \) yields the original observation model of eqn. (2.1). In the framework of eqn. (4.1), we interpret the observations \( x_t \) as 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

\[
p(y_{1:T} | \Delta_{1:T}) = \prod_{t=1}^{T} N(x_t | \Upsilon_t z_t, \Psi_t) N(z_t | 0, \Sigma_{t-1})
\]

\[
p(\Delta_{1:T}) = p(\Sigma_0 | n, S) \prod_{t=1}^{T} NIW_q(\Upsilon_t, \Psi_t | F, n, S)
\]
where $NIW$ denotes the matrix normal, inverse Wishart prior on $\{\Upsilon_t, \Psi_t\}$ of eqn. (2.7). We omit the dependency of the left hand side on the hyperparameters $n$, $F$, and $S$ for notational simplicity. The right panel of Figure 1 displays the resulting graphical model, clearly illustrating the simplified conditional independence structure that enables computation as developed below. Note that $\Delta_t$ plays the role of an augmented state and the evolution to time $t$ defines $\Sigma_t$ 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 for the sequence of variance matrices. This structure (Figure 1 (right)) immediately suggests a natural MCMC sampler that iterates between the following steps:

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

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

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

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

**Fig. 1.** Representation of the graphical model of (left) the IW-AR(1) process under augmentation by the latent $z_t$ and (right) the augmented data $y_t = [z_t' \ x_t']'$ and latent states $\Delta_t = \{\Sigma_{t-1}, \Upsilon_t, \Psi_t\}$. 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.
Our proposal distributions are defined using an approximate forward filtering, backwards sampling (FFBS) algorithm; 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 \( \Delta_t \) to the augmented matrix

\[
\Delta_t \rightarrow \left( \begin{array}{cc} \Sigma_{t-1} & \Sigma_{t-1} \Gamma_t' \\ \Gamma_t \Sigma_{t-1} & \Sigma_t \end{array} \right),
\]

and thus use the two interchangeably. Our goal is to develop a forward filtering algorithm that produces an approximation to \( p(\Delta_t \mid y_{1:t}) \), \( t = 1, \ldots, T \), which can then be used in backward-sampling an approximate posterior sequence \( \Sigma_{0:T} \). This approximate FFBS provides our Metropolis-Hastings proposal of \( \Sigma_{0:T} \) that is then accepted or rejected to produce a sample from the target posterior \( p(\Sigma_{0:T} \mid y_{1:T}) \).

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 Supplementary Material, 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 \( g_{t-1|t-1}^{\Delta_{t-1}}(\Delta_{t-1} \mid y_{1:t-1}) \) 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

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

Here, \( r_t \) is a specified degree of freedom to use in the approximation at time \( t \). The mean is simply \( E_{g_{t|t-1}[\Delta_{t-1}]}[\Delta_t \mid y_{1:t-1}] \) – i.e., the predictive mean under the distribution \( g_{t-1|t-1}^{\Delta_{t-1}} \) and the dynamics specified by the IW-AR prior. Note that the one-step propagation based on an inverse Wishart posterior is nearly an inverse Wishart form, so the approximation made at each step is relatively small. The required expectation here is easily seen to be

\[
E_{g_{t-1|t-1}[\Delta_{t-1}]}[\Delta_t \mid y_{1:t-1}] = \left( \begin{array}{c} S_{t-1} \\ FS_{t-1} \end{array} \right) \frac{S_{t-1} F'}{FS_{t-1} F' + c_{n,q} (1 + \text{tr}(S_{t-1}(nS)^{-1})V)}
\]

where

\[
E_{\Delta_{t-1}}[\Delta_t \mid y_{1:t-1}] = \left( \begin{array}{c} S_{t-1} \\ FS_{t-1} \end{array} \right) \frac{FS_{t-1} F'}{FS_{t-1} F' + c_{n,q} (1 + \text{tr}(S_{t-1}(nS)^{-1})V)}
\]
can deterministically compute \( \{IW_y\} \) where the sequence yields \( g \) backwards in time, we condition on the previously sampled \( \Sigma \). We then harness the reverse time process, depicted in Figure 2. As we iterate backwards in time, we condition on the previously sampled \( \Sigma \) to sample \( \Sigma_{t-1} \) as follows. Using the same ideas as in the proof of Theorem 3.2, we manipulate \( \Delta_t \) to sample the reverse time innovations as

\[
\begin{align*}
\Psi_t &\sim IW(r_t + 1 + q, G_t^{11} - G_t^{21} (G_t^{22})^{-1} G_t^{21}) \\
\tilde{\Psi}_t | \Psi_t &\sim N(G_t^{21'} (G_t^{22})^{-1}, \Psi_t, (G_t^{22})^{-1}).
\end{align*}
\]

where the \( S_t \) sequence is updated using the identity

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

with \( S_0 = S \). Further details are in the Supplementary Material. The subsequent update step of incorporating observation \( y_t \) is exact based on the approximations made so far. Namely,

\[
g_{t|t}(\Delta_t | y_{1:t}) = IW \left( r_t + 1, (r_t - 2)E_{g_{t-1|t-1}}[\Delta_t | y_{1:t-1}] + y_t y'_t \right).
\]

The forward filter thus boils down to recursively computing \( S_t \) as in eqn. (5.6).

**Backward Sampling.** Running these forward filtering computations to time \( t = T \) yields \( g_{T|T}(\Delta_T | y_{1:T}) \) approximating the true posterior \( p(\Delta_T | y_{1:T}) \). We use the sequence of approximations \( g_{t|t}(\Delta_t | y_{1:t}) \) 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 \( \Sigma_T \) from the implied approximate posterior margin

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

We then harness the reverse time process, depicted in Figure 2. As we iterate backwards in time, we condition on the previously sampled \( \Sigma_t \) to sample \( \Sigma_{t-1} \) as follows. Using the same ideas as in the proof of Theorem 3.2, we manipulate \( \Psi_t \) to sample the reverse time innovations as

\[
\begin{align*}
\tilde{\Psi}_t &\sim IW(r_t + 1 + q, G_t^{11} - G_t^{21} (G_t^{22})^{-1} G_t^{21}) \\
\tilde{\Psi}_t | \Psi_t &\sim N(G_t^{21'} (G_t^{22})^{-1}, \tilde{\Psi}_t, (G_t^{22})^{-1}).
\end{align*}
\]
and then set

\[ \Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Upsilon}_t \Sigma_t \tilde{\Upsilon}_t'. \]

Here, \( G_t = (r_t - 2) E_{g_{t-1|t-1}}[\Delta_t|y_{1:t-1}] + y_t y_t' \), with \( G_{t1} \), \( G_{t2} \), \( G_{t22} \) denoting the three unique \( q \times q \) sub-blocks \( (G_{t2} = G_{t22}) \). These terms, which can be regarded as sufficient statistics of the forward filtering procedure, can be written as

\[
\begin{align*}
G_{t1} &= (r_t - 2) S_{t-1} + z_t z_t' \\
G_{t2} &= (r_t - 2) F S_{t-1} + x_t x_t' \\
G_{t22} &= (r_t - 1) S_t,
\end{align*}
\]

with \( S_t \) as in \( (5.6) \). In practice, conditioned on \( \{n, S, F\} \), the sequence \( S_{1:T} \) is precomputed and simply accessed in backward sampling \( \Sigma_{0:T} \).

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

\[ \tilde{\Upsilon}_t = \Sigma_t \tilde{\Upsilon}_t' \Sigma_{t-1}^{-1} \quad \text{and} \quad \Psi_t = \Sigma_t - \tilde{\Upsilon}_t \Sigma_{t-1} \tilde{\Upsilon}_t'. \]

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

\[ r(\Delta_{1:T}^*) = \frac{p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T} \mid y_{1:T})}{q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T} \mid y_{1:T})} \]

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 \( \{\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T}\} \) and \( \{\Sigma_0, \tilde{\Upsilon}_{1:T}, \Psi_{1:T}\} \),

\[ \frac{p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T} \mid y_{1:T})}{q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T} \mid y_{1:T})} \propto \frac{p(y_{1:T} \mid \Sigma_0, \tilde{\Upsilon}_{1:T}, \Psi_{1:T}) p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T})}{q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \Psi_{1:T} \mid y_{1:T})}. \]

The augmented data likelihood is given by

\[ p(y_{1:T} \mid \Sigma_0, \tilde{\Upsilon}_{1:T}, \Psi_{1:T}) = \prod_{t=1}^{T} N(x_t \mid \tilde{\Upsilon}_t z_t, \Psi_t) N(z_t \mid 0, \Sigma_{t-1}). \]
As specified in eqn. (3.4), the prior of the reverse time process is given by

\[(5.16) \quad p(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T}) = \text{IW}(\Sigma_T \mid n + 2, nS)\]
\[
\times \prod_{t=1}^{T} \text{IW}(\tilde{\Psi}_t \mid n + q + 2, n\tilde{V})N(\tilde{\Upsilon}_t \mid \tilde{F}, \tilde{\Psi}_t, (nS)^{-1}).
\]

Similarly, the proposal density decomposes as

\[(5.17) \quad q(\Sigma_T, \tilde{\Upsilon}_{1:T}, \tilde{\Psi}_{1:T} \mid y_{1:T}) = \text{IW}(\Sigma_T \mid r_T + 1, G_{T}^{22}) \prod_{t=1}^{T} f_t(\tilde{\Upsilon}_t)h_t(\tilde{\Psi}_t \mid \tilde{\Upsilon}_t)\]

where

\[f_t(\tilde{\Upsilon}_t) = \text{IW}(\tilde{\Psi}_t \mid r_t + q + 1, G_{t}^{11} - G_{t}^{22}(G_{t}^{22})^{-1}G_{t}^{22})\]

and

\[h_t(\tilde{\Psi}_t \mid \tilde{\Upsilon}_t) = N(\tilde{\Upsilon}_t \mid G_{t}^{22}(G_{t}^{22})^{-1}, \tilde{\Psi}_t, (G_{t}^{22})^{-1}).\]

**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 \(\Sigma_T^*\) from \(g_{T \mid T}\) and a collection of \(T\) independent samples \(\{\tilde{\Upsilon}_{t}^*, \tilde{\Psi}_{t}^*\}\) from distributions based on \(g_{t \mid t}\). If the final approximate filtered distribution \(g_{T \mid T}\) 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 \(g_{T \mid T}\) decreases with \(T\). Furthermore, even if \(g_{T \mid T}\) 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 \(\Sigma_{T}^*\) and \(\{\tilde{\Upsilon}_{t}^*, \tilde{\Psi}_{t}^*\}\). 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 \(\{\Sigma_{T}, \tilde{\Upsilon}_{t-1:T}, \tilde{\Psi}_{t-1:T}\}\), 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 \(\{\Sigma_{T}, \tilde{\Upsilon}_{1:t-1:T}, \tilde{\Psi}_{1:t-1:T}\}\) and simply consider accepting or rejecting \(\{\tilde{\Upsilon}_{t}^*, \tilde{\Psi}_{t}^*\}\) at a single time step \(t\).
An Innovations-Based Sampler. Let $\theta_t = \{\tilde{Y}_t, \tilde{\Psi}_t\}$ and $S_t = \{\theta_1, \theta_2, \ldots, \theta_T, \Sigma_T\}$. For each $t$ we propose

$$S_t^* = \{\theta_1, \theta_2, \ldots, \theta_{t-1}, \theta_t^*, \theta_{t+1}, \ldots, \theta_T, \Sigma_T\},$$

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

$$\frac{r(\Delta_{1:T}^*)}{r(\Delta_{1:T})} = \frac{p(S_t^* \mid y_{1:T}) q(S_t \mid y_{1:T})}{q(S_t^* \mid y_{1:T}) p(S_t \mid y_{1:T})}$$

$$= \frac{p(y_{1:T} \mid \Sigma_0^*, Y_{1:T}^*, \Psi_1^*, \Psi_{t+1:T}, Y_{t+1:T})}{p(y_{1:T} \mid \Sigma_0, Y_{1:T}, \Psi_{1:T})} \frac{q(S_t) q(S_t^*)}{q(S_t^*) q(S_t)} \frac{p(\theta_t) p(\theta_t^*)}{q(\theta_t) q(\theta_t^*)} p_t(\theta_t^* \mid y_{1:T})$$

Here $p_t(\theta_t)$ denotes the matrix normal, inverse Wishart prior on the backwards innovations $\theta_t = \{\tilde{Y}_t, \tilde{\Psi}_t\}$ 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 $S_t^*$ and the previous $S_t$ are the backwards innovations at time $t$ (i.e., $\theta_t$). For the likelihood, the effects of the change in backwards innovations at time $t$ propagate to the forward parameters $\{\Sigma_0^*, Y_{1:T}^*, \Psi_1^*\}$ while leaving $\{Y_{t+1:T}, \Psi_{t+1:T}\}$ unchanged. See Figure 2.

Note that the proposed innovations sampler is quite computationally intensive since the accept-reject ratio calculation for each proposed $\{\tilde{Y}_t^*, \tilde{\Psi}_t^*\}$ requires recomputing $\Sigma_{0:t-1}^*$. In practice, we employ an approximate sampler that harnesses the fact that the effects of a given $\{\tilde{Y}_t^*, \tilde{\Psi}_t^*\}$ on $\Sigma_{t-\tau}$ decreases in expectation as $\tau$ increases. That is, $\{\tilde{Y}_t^*, \tilde{\Psi}_t^*\}$ 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 $\tau$ increasing until the Frobenius norm $||\Sigma_{t-\tau}^* - \Sigma_{t-\tau-1}|| < \epsilon$ for some pre-specified, small value $\epsilon > 0$. That is, we propagate the effects of $\{\tilde{Y}_t^*, \tilde{\Psi}_t^*\}$ 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 $\tau$ based on $F$ and $n$ since these hyperparameters determine (in expectation) the rate at which the effects of the proposed $\{\tilde{Y}_t^*, \tilde{\Psi}_t^*\}$ decay.

In contrast to sequential sampling of $\{\tilde{Y}_t, \tilde{\Psi}_t\}$ for $t = T, \ldots, 1$, one can imagine focusing on regions where the current $\Sigma_t$ 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. Recall that we sample the IW-AR hyperparameters $F$ and $S$ conditioned on the observations $x_{1:T}$ and a sampled sequence $z_{1:T}$, but marginalizing the latent augmented variance matrix sequence $\Delta_{1:T}$. We now discuss our prior and proposal specifications, and the resulting posterior computations for sampling $\{F, S\}$.

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) $F = \rho \cdot I_d$, 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 $V = (1 - 1' - \rho \cdot \rho') \circ S$ 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 $\rho_i$ 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 $W(v_0, V_0)$ denote a Wishart prior for $V$ and $\text{Beta}(c_{p_0,i}, c(1 - p_0,i))$ a beta prior for $p_i$. We use an independence chain sampler in which $V^*$ is proposed from a Wishart proposal $W(v_1, V_1)$ and $p_i^*$ from a beta proposal $\text{Beta}(d_{p_1,i}, d(1 - p_1,i))$. The accept-reject ratio is then calculated based on the ratio:

\[
\frac{r(V^*, F^*)}{r(V, F^*)} = \frac{p(x_{1:T} | z_{1:T}, F^*, V^*)p(z_{1:T} | F^*, V^*)p(V^*)p(F^*)q(V)q(F)}{p(x_{1:T} | z_{1:T}, F, V)p(z_{1:T} | F, V)p(V)p(F)q(V)q(F^*)}.
\]

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

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

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

(6.1)

where $A_i$ is the $q \times q$ autoregressive parameter matrix at lag $i$ and $\Sigma_t$ follows 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 Supplementary Material 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. Much of prior work with clinically relevant data sets has been on the evaluation of time:frequency structure in such series (Freyermuth, Ombao and von Sachs, 2010; Krystal et al., 2000; Ombao, von Sachs and Guo, 2005) and time-varying parameter vector autoregressions (Prado and West, 1997, 2010; West, Prado and Krystal, 1999). Existing models represent some aspects of cross-series structure in this inherently spatially distributed multiple time series context (Krystal, Prado and West, 1999; Prado, West and Krystal, 2001), but past studies have shown substantial residual dependencies among estimated innovations processes across EEG probe locations. 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.

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

\[ \nu_t = \sum_{i=1}^{r} A_{t,i} \nu_{t-i} + w_t, \quad w_t \sim N(0, I) \]  

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, Prado and Krystal (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 \( a_i \sim N(0, 10 I_q) \).

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 \( \rho \) and \( S \) based on an exploratory analysis of an earlier held-out section of the time series, \( \xi_{0:T}^0 \), also of length 1000. Specifically, this was based on estimating innovations \( x_{1:T}^0 \) from \( q \) separate, univariate TVAR models as in Krystal, Prado and West (1999). 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 \( U_{ii,0:T} \). 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 \( \rho_{0,i}^2 = (a_i(n + 1) - 1)/n \) and \( a_i \) is the approximate posterior mean of the IW-AR autoregressive parameter from the hold-out data analysis of \( U_{ii,1:T} \); second, we set \( \nu_0 = q + 2 \) and \( V_0 = (1 \cdot 1^t - \rho_0 \cdot \rho_0^t) \circ S_0 \), where \( S_0 \) is the sample mean of all of the the \( U_{0:T} \).

Although centered around a held-out-data-informed mean, the chosen Wishart prior for \( S \) is quite diffuse and the beta priors for the \( \rho_i \) 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 $\rho_1$ and $V_1$ that are used in specifying the Beta$(d\rho_{1,i}, d(1 - \rho_{1,i}))$ proposal for $\rho_i$ 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 $||\Sigma^*_t - \Sigma_{t-\tau}||_2 < \epsilon$ and uses $\epsilon = 1e - 4$.

Finally, using this estimated $x_{1:T}$ sequence for the test data, we analyzed the appropriateness of maintaining inverse Wishart margins for $\Sigma_t$ by examining the implied multivariate t distribution for $x_t$. The hyperparameter $S$ is taken to be the prior mean and a range of $n$ are considered. The results, which are provided in the Supplementary Material, suggest a good fit to the data. Note that the results below are based on an analysis of the actual observations $x_t; the estimated $x_t$ are simply used in the exploratory studies.

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 ini-
Figure 4. Estimated trajectories of some covariance terms $\Sigma_{ij,t}$ for $i \neq j$ for $j = 1$ (EEG channel 7) colored as in Figure 3.

Figure 5. Estimated trajectories of correlations between each of 3 channels and all other channels as a function of time. The correlations are computed based on posterior means of $\Sigma_t$ using MCMC samples [1000 : 10 : 5000] from 5 chains.

tialized 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 3 displays volatility trajectories for several of the EEG channels showing clear changes in volatility over the 8 seconds of data, while related temporal structure in cross-series covariances is evident across all pairs of channels as exemplified in Figure 4. These changes are also captured in Figure 5, which displays estimated trajectories of time-varying correlations between AR innovations for some pairs of EEG channels. For the model parameters, Figure 6 shows clear evidence of learning via changes from prior to posterior summaries for the $\rho_i$ and $S_{ii}$ elements; this figure also highlights the high temporal dependence in the IW-AR(1) model and heterogeneity across EEG channels. See the Supplementary Material for additional figures from this analysis.

7. IW-AR(2) and Higher Order Models. The constructive approach for IW-AR(1) models extends to higher orders 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

$$
\begin{pmatrix}
\Delta_{t-1} & \Delta_{t-1}\Gamma'_t \\
\Gamma_t & \Sigma_t
\end{pmatrix}
\sim \text{IW}_3(n + 2, nS_3),
$$

where

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

Then, for all $t$,

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

$$
\Sigma_t \sim \text{IW}_q(n + 2, nS).
$$

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:

$$
\gamma_t = \Gamma_{1t}\Sigma_{t-2} + \Gamma_{2t}\phi_{t-1}
$$

$$
\phi_t = \Gamma_{2t}\Sigma_{t-1} + \Gamma_{1t}\phi_{t-1}.
$$
The “memory” induced by these off-diagonal elements is evident as the full conditional distribution for $\Sigma_t$ 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

$$(7.7) \quad \Sigma_t = \Omega_t + \Gamma_t \Delta_{t-1} \Gamma_t'$$

with time $t$ innovation matrices $\Gamma_t$ ($q \times 2q$) and $\Omega_t$ ($q \times q$) independent of $\Delta_{t-1}$ and distributed as

$$(7.8) \quad \Omega_t \sim IW_q \left( n + 2 + 2q, n \left( S - [H \quad G] S_2^{-1} [H \quad G]' \right) \right)$$

$$(7.9) \quad \Gamma_t \mid \Omega_t \sim N \left( [H \quad G] S_2^{-1}, \Omega_t, (nS_2)^{-1} \right).$$

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

$$(7.10) \quad S_3 = \begin{pmatrix} S & G' & G'S^{-1}G' \\ G & S & G' \\ G^{-1}G & G & S \end{pmatrix} = \begin{pmatrix} 0 & S_2 & S_2 [0 \quad GS^{-1}] \\ S_2 & S_2 & [0 \quad GS^{-1}] \end{pmatrix}^{'},$$

then

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

$$(7.12) \quad \Gamma_t \mid \Omega_t \sim N \left( [0 \quad GS^{-1}], \Omega_t, (nS_2)^{-1} \right).$$

Furthermore, taking $G = FS$ leads to

$$(7.13) \quad S_3 = \begin{pmatrix} S & SF' & SF^{2'} \\ FS & S & SF' \\ F^2S & FS & S \end{pmatrix} = \begin{pmatrix} 0 & S_2 & S_2 [0 \quad F] \\ S_2 & S_2 & [0 \quad S] \end{pmatrix}^{'},$$

and

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

$$(7.15) \quad \Gamma_t \mid \Omega_t \sim N \left( [0 \quad F], \Omega_t, (nS_2)^{-1} \right).$$

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

$$(7.16) \quad \left| S_3 \right| = \left| S_2 \right| \left| S - FSF' \right| = \left| S \right| \left| S - FSF' \right|^2.$$
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

$$
\Sigma_t = \Psi_t + \Upsilon_t \Sigma_{t-1} \Psi_t' - \Phi_t \Psi_{t-1} \Sigma_{t-2} \Psi_{t-1}' \Phi_t' + \Xi_t,
$$

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

$$
E[\Sigma_t | \Delta_{t-1}] = F \Sigma_{t-1} F' + H \Psi_{t-1} H' + c_{n,2q} W + c_{n,2q} W \text{tr}(\Psi_{t-1} (nV)^{-1}) \left(1 + \text{tr}(\Sigma_{t-1} (nS)^{-1})\right),
$$

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 $\Delta_{t-1}$. In the limit as $n \to \infty$,

$$
E[\Sigma_t | \Delta_{t-1}] = W + F \Sigma_{t-1} F' + H \Psi_{t-1} H'
= S + F(\Sigma_{t-1} - S) F' + H((\Sigma_{t-1} - S) - (\Upsilon_{t-1} \Sigma_{t-2} \Upsilon_{t-1}' - FSF')) H'.
$$

Derivations are provided in the Supplementary Material.

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.

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, Chatfield and Walker (2002); Pitt and Walker (2005). 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 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 \( \Sigma_t \). 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 (2006b) specifies the transition distribution for \( \Sigma_t \) given \( \Sigma_{t-1} \) 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 (asymptotically) 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, 2006a), 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, Jasiak and Sufana (2009) define Wishart processes via noncentral Wishart transitions, with interpretations as functions of sample variance matrices of a collection of latent vector autoregressive processes. Specifically, when the degree of freedom \( n \) is integer, \( \Sigma_t = \sum_{k=1}^n x_k,t x'_k,t \), with each \( x_k \) independently defined via \( x_{k,t} = M x_{k,t-1} + e_{k,t} \) and \( e_{k,t} \sim N(0, \Sigma_0) \). As such, the Wishart process provides an analytic expression for the predictive
distribution at finite horizons, whereas the IW-AR only provides predictions in the
mean or predictions via simulation. For stationary autoregressions, one can analyze
the asymptotic marginal distribution of $\Sigma_t$ and show that the process is asymptot-
ically strictly stationary. 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. One scenario where this arises is in modeling realized covariance
matrices, as recently explored in ?, though generally the volatility process is la-
tent. Extensions to embedding the proposed Wishart autoregressive process within
a standard stochastic volatility framework rely on estimation of a model that is
Wishart-autoregressive in mean based on nonlinear filtering approximations of la-
tent 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 straight-
forward.

Our proposed IW-AR is a fundamentally different approach to AR(1) modeling
of variance matrices; it is a pure state-space formulation and defines foundation for
new lines of research on multivariate variance matrix processes. Instead of directly
specifying a transition distribution, we harness inverse Wishart theory to define an
autoregression based on sequences of pairwise joint distributions with consistent
marginals. The construction yields direct analysis of a number of theoretical prop-
erties without relying on asymptotic arguments: the IW-AR is a stationary process
with inverse Wishart margins, fundamental to Bayesian model fitting, and can be
specified as time-reversible or irreversible. Furthermore, the IW-AR shares some
of the benefits of the framework of Gouriéroux, Jasiak and Sufana (2009), such as
invariance to quadratic transformations.

AR models for Cholesky elements. Several recent works use linear, normal AR(1)
models for off-diagonal elements of the Cholesky of $\Sigma_t$ and for the log-diagonal
elements (Cogley and Sargent, 2005; Lopes, McCulloch and Tsay, 2010; Primiceri,
2005; ?, ?), building on the Cholesky-based heteroscedastic model of Pourahmadi
(1999), and a natural parallel of Bayesian factor models for multivariate volatility.
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 inter-
pretable. We can cast our IW-AR within a similar framework. The inverse Wishart
margins for $\Sigma_t$ and $\Sigma_{t-1}$ translate to Wishart margins for the precision matrices
$\Sigma_t^{-1}$ and $\Sigma_{t-1}^{-1}$. 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 vari-
ables, 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 suggests opportunity to develop more effective computational methods using some variant of particle filtering (cf. Doucet, de Freitas and Gordon, 2005) and particle learning (Carvalho et al., 2010; Lopes et al., 2010). 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, Doucet and Holenstein, 2010) to provide samples from the smoothing density. The algorithm only relies on an ability to simulate from \( p(\Sigma_t \mid \Sigma_{t-1}) \). 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, Massam and West, 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.

An alternative approach to scaling the IW-AR to high-dimensional time series is to harness a factor model. In particular, one could straightforwardly extend to a model in which a low-dimensional, normally-distributed latent factor has volatility following an IW-AR process. Although the computations presented in this paper focused on a simple independent-normal or normal-autoregressive observation model, one could embed the IW-AR process in a wealth of more complicated frameworks. The theoretical properties of the volatility process studied herein remain unchanged.
References.


APPENDIX A: PROOFS OF THEORETICAL PROPERTIES

A.1. Proof of Theorem 3.1. For all $h > 0$ and all $\tau > 0$, the conditional independencies defined by the construction of eqn. (2.3) and (2.4), and further outlined in Section 2.2 (e.g., $\phi_t$ is conditionally independent of $\{\phi_{t-1}, \Sigma_{t-1}, \Sigma_{t-2}\}$ given $\Sigma_t$), imply that (cf. ?)

\[(A.1)\]

\[p(\Sigma_{t:t+h}, \phi_{t:t+h}) = \prod_{\tau=1}^{h} p(\Sigma_{t+\tau-1}, \Sigma_{t+\tau}, \phi_{t+\tau}) \prod_{\tau=2}^{h} p(\Sigma_{t+\tau-1})\]

\[= \prod_{\tau=1}^{h} \text{IW}_{2q}(n + 2, n \Delta_0) \prod_{\tau=2}^{h} \text{IW}_{q}(n + 2, nS) = p(\Sigma_{t+\tau:t+h+\tau}, \phi_{t+\tau:t+h+\tau}),\]

where $\Delta_0$ is the $2q \times 2q$ scale matrix parameter of the joint density. 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. Assuming a valid model, eqn. (A.2) implies the stationarity of $\{\Sigma_t\}$.

A.2. Proof of Theorem 3.3. For the case of $A$ invertible, and assuming $\Sigma_t$ follows an IW-AR(1) specified by the joint distribution of eqn. (2.4), standard theory implies that

\[(A.3)\]

\[
\begin{pmatrix}
    A' & 0 \\
    0 & A'
\end{pmatrix}
\begin{pmatrix}
    \Sigma_{t-1} & \Sigma_{t-1} Y_t' \\
    Y_t \Sigma_{t-1} & \Sigma_t
\end{pmatrix}
\begin{pmatrix}
    A & 0 \\
    0 & A
\end{pmatrix}
\sim \text{IW}_{2q}(n + 2, n \hat{S} \hat{F})
\begin{pmatrix}
    A' \Sigma_{t-1} A \\
    \hat{Y}_t A' \Sigma_{t-1} A 
\end{pmatrix}
\begin{pmatrix}
    A' \Sigma_{t-1} A \\
    \hat{Y}_t A' \Sigma_{t-1} A 
\end{pmatrix}
\sim \text{IW}_{2q}(n + 2, n \hat{S} \hat{F})
\]

implying

\[(A.4)\]

\[
\begin{pmatrix}
    A' \Sigma_{t-1} A \\
    \hat{Y}_t A' \Sigma_{t-1} A 
\end{pmatrix}
\sim \text{IW}_{2q}(n + 2, n \hat{S} \hat{F})
\]

for $\hat{F} = A' F (A')^{-1}$ and $\hat{S} = A' S A$. That is, $A' \Sigma_t A$ follows an IW-AR(1) defined by $\hat{F}$, $\hat{S}$, and $n$. 
A.3. Proof of Corollary 3.1. The result follows immediately from Theorem 3.3 taking \( A = (S^{-1/2})' \).

A.4. Proof of Theorem 3.4. Let \( v_{t_k} \) denote the \( k \)th row of \( \Upsilon_t \) and \( f_k \) the \( k \)th column of \( F \). Then, for the IW-AR(1) we can write the \((i, j)\) element of \( \Sigma_t \) as

\[
\Sigma_{t,ij} = \Psi_{t,ij} + v_{t_i} \Sigma_{t-1} v_{t_j}'.
\]

(A.5)

Taking the expectation conditioned on \( \Sigma_{t-1} \),

\[
E[\Sigma_{t,ij} | \Sigma_{t-1}] = \frac{nV_{ij}}{n + q} + \sum_{k=1}^{q} \sum_{\ell=1}^{q} \Sigma_{t-1,k\ell} E[v_{tk}v_{t\ell}]
\]

\[
= \frac{nV_{ij}}{n + q} + \sum_{k=1}^{q} \sum_{\ell=1}^{q} \Sigma_{t-1,k\ell} \left\{ \left[ \frac{nV_{ij}}{n + q} (nS)^{-1} \right]_{k\ell} + F_{ik}F_{j\ell} \right\}
\]

(A.6)

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

\[
E[\Sigma_t | \Sigma_{t-1}] = \frac{1 + \text{tr}(\Sigma_{t-1}(nS)^{-1})}{n + q} nV + F\Sigma_{t-1}F'.
\]

(A.7)

A.5. 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 \( t - 1 \)

\[
E[\Sigma_{t-1} | \Sigma_0] = F^{t-1}\Sigma_0F^{t-1'} + \frac{n}{n + q} \left( S - F^{t-1}SF^{t-1'} \right) + O \left( \frac{1 \cdot 1'}{n} \right).
\]

(A.8)

To prove that eqn. (A.8) holds for general \( t \), we apply iterated expectations to the conditional expectation of eqn. (3.5):

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

(A.9)

\[
= F^t\Sigma_0F'^t + \frac{n}{n + q} \left( S - F^{t}SF^{t'} \right) + \frac{V}{n + q} \text{tr} \left( E[\Sigma_{t-1} | \Sigma_0]S^{-1} \right),
\]

(A.10)

where we have used the definition \( V = S - FSF' \). Since

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

(A.11)
we conclude that, indeed,

(A.12) \[ E[\Sigma_t \mid \Sigma_0] = F^t\Sigma_0F'^t + \frac{n}{n + q} \left( S - F^tSF'^t \right) + O \left( \frac{1}{n} \right). \]

Then also

(A.13) \[ \lim_{n \to \infty} E[\Sigma_t \mid \Sigma_0] = S + F^t(S_0 - S)F'^t. \]

A.6. Proof of Theorem 3.6. For the case of \( F = ERE' \) and \( S = EQE' \), we can write eqn. (2.4) as

(A.14) \[ \left( \begin{array}{cc} \Sigma_{t-1} & \Sigma_{t-1}Y_t \\ Y_t\Sigma_{t-1} & \Sigma_t \end{array} \right) \sim \text{IW}_2 \left( n + 2, n \left( \begin{array}{cc} E & 0 \\ 0 & E \end{array} \right) \left( \begin{array}{cc} Q & QR \\ QR & Q \end{array} \right) \left( \begin{array}{cc} E' & 0 \\ 0 & E' \end{array} \right) \right). \]

The result follows immediately taking \( A = E \) in Theorem 3.3 such that \( \hat{F} = E'(ERE')^{-1} = R \) and \( \hat{S} = E'(EQE')E = Q \).

The derivation of the conditional mean of eqn. (3.11) is exactly as in the general IW-AR case, noting that \( \text{tr}(\hat{R}_{t-1}(nQ)^{-1}) = \sum_i \hat{R}_{t-1,ii}/(n\xi_i) \).

The off-diagonal terms of \( E[\Sigma_t \mid \hat{\Sigma}_{t-1}] \) follow immediately by examining eqn. (3.11). Specifically, for \( i \neq j \),

(A.15) \[ E[\Sigma_{t,ij} \mid \hat{\Sigma}_{t-1}] = \rho_i \rho_j \hat{\Sigma}_{t-1,ij} \quad \text{and} \quad E[\Sigma_{t,ij} \mid \Sigma_0] = \rho_i \rho_j \hat{\Sigma}_{0,ij}. \]

For the diagonal terms, we derive

(A.16) \[ E[\hat{\sigma}^2_{t,ij} \mid \hat{\Sigma}_{t-1}] = \rho_j^2 \hat{\sigma}^2_{t-1,ij} + \frac{n}{n + q} \xi_j(1 - \rho_j^2) + \sum_i \frac{\xi_j (1 - \rho_j^2)}{\xi_i (n + q)} \hat{\sigma}_{t-1,ii} \]

(A.17) \[ = \frac{n}{n + q} \xi_j(1 - \rho_j^2) + \sum_i \left( \frac{\xi_j (1 - \rho_j^2)}{\xi_i (n + q)} + \rho_j^2 \delta_{ij} \right) \hat{\sigma}_{t-1,ii}. \]

Equivalently, letting \( \hat{\sigma}^2_t = [\hat{\Sigma}_{t,11}, \hat{\Sigma}_{t,22}, \ldots, \hat{\Sigma}_{t,qq}]' \), \( \xi = \text{diag}(Q) \) and \( \xi_{-1} = \text{diag}(Q^{-1}) \),

(A.18) \[ E[\hat{\sigma}^2_t \mid \hat{\Sigma}_{t-1}] = E[\hat{\sigma}^2_t \mid \hat{\sigma}^2_{t-1}] = \frac{n}{n + q} (I - R^2)\xi + \left[ \frac{1}{n + q} (I - R^2)\xi\xi_{-1}' + R^2 \right] \hat{\sigma}^2_{t-1}. \]

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

(A.19) \[ E[\hat{\sigma}^2_t \mid \hat{\sigma}^2_0] = B'\hat{\sigma}^2_0 + \sum_{r=0}^{t-1} B^r \alpha. \]
Since $B$ represents a matrix (strictly) convex combination of $\frac{\xi_{\ell-1}^t}{n+q}$ and $I$, the maximum eigenvalue of $B$ is bounded by

$$\left\| (I - R^2) \max \left\{ \text{eig} \left( \frac{\xi_{\ell-1}^t}{n+q} \right) \right\} \cdot I + R^2 \cdot 1 \right\|_0.$$  \hspace{1cm} (A.20)

Here, $\max \{ \text{eig}(A) \}$ denotes the maximum eigenvalue of $A$. The term $\xi_{\ell-1}^t$ is a rank 1 matrix implying that the only non-zero eigenvalue is equal to $\text{tr}(\xi_{\ell-1}^t) = q$. Thus, regardless of $n$, $B$ has eigenvalues with modulus strictly less than 1 since $\frac{\xi_{\ell-1}^t}{n+q}$ 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 $\hat{\Sigma}_0$ exponentially fast regardless of $n$. Furthermore, since the eigenvalues of $B$ have modulus less than 1,

$$E[\hat{\sigma}^2_t | \hat{\sigma}^2_0] = B^t \hat{\sigma}^2_0 + (I - B)^{-1} (I - B^t) \alpha,$$  \hspace{1cm} (A.21)

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

$$\lim_{t \rightarrow \infty} E[\hat{\sigma}^2_t | \hat{\sigma}^2_0] = (I - B)^{-1} \alpha$$  \hspace{1cm} (A.22)

$$= \left[ \left( I - \frac{\xi_{\ell-1}^t}{n+q} \right)^{-1} (I - R^2)^{-1} \right] \frac{n}{n+q} (I - R^2) \xi$$  \hspace{1cm} (A.23)

$$= \frac{n}{n+q} \left( I - \frac{\xi_{\ell-1}^t}{n+q} \right)^{-1} \xi$$  \hspace{1cm} (A.24)

$$= \xi.$$  \hspace{1cm} (A.25)

The last equality follows from matrix inversion and the fact that $\xi_{\ell-1}^t \xi = q$. The proof is completed by combining the fact that $\lim_{t \rightarrow \infty} E[\hat{\Sigma}_{ij,t} | \hat{\Sigma}_0] = 0$ for $i \neq j$ and $\lim_{t \rightarrow \infty} E[\hat{\sigma}^2_t | \hat{\sigma}^2_0] = S$. That is, $\lim_{t \rightarrow \infty} E[\hat{\Sigma}_t | \hat{\Sigma}_0] = Q$.

A.7. Proof of Corollary 3.2. As in Theorem 3.6, $\hat{\Sigma}_t = E' \Sigma_t E$ implying that $\Sigma_t = E \hat{\Sigma}_t E'$. Therefore, using the result of Theorem 3.6, $\lim_{t \rightarrow \infty} E[E \hat{\Sigma}_t E' | E \hat{\Sigma}_0 E' ] = EQE'$. That is, $\lim_{t \rightarrow \infty} E[\Sigma_t | \Sigma_0] = S$. Again by Theorem 3.6, the rate is exponential in terms of the eigenvalues of $B = \frac{1}{n+q} (I - R^2) \xi_{\ell-1}^t + R^2$.

APPENDIX B: DERIVATION OF FORWARD FILTERING BACKWARD SAMPLING ALGORITHM

B.1. Approximate Forward Filtering. The inverse Wishart prior on $\Delta_1$ can be analytically updated to an inverse Wishart posterior conditioned on $y_1$:

$$p(\Delta_1 | y_1) = \text{IW}_{2q} \left( n + 3, n \begin{pmatrix} S & SF' \\ FS & \Sigma \end{pmatrix} + y_1 y_1' \right).$$  \hspace{1cm} (B.1)
To propagate to $t = 2$, we use the Chapman-Kolmogorov equation, integrating over $\Delta_1$:

$$p(\Delta_2 \mid y_1) \propto \int p(\Delta_2 \mid \Delta_1) p(\Delta_1 \mid y_1) d\Delta_1$$

$$\propto \int \delta_{\Sigma_1 = \Psi_1 + \Upsilon_1 \Sigma_0 \Upsilon_1} p(\Psi_2) p(\Upsilon_2 \mid \Psi_2) p(\Delta_1 \mid y_1) d\Upsilon_1 d\Psi_1 d\Sigma_0$$

$$\propto \text{IW}_q(\Psi_2 \mid n + q + 2, nV) N(\Upsilon_2 \mid F, \Psi_2, (nS)^{-1})$$

(\text{B.2})

$$\text{IW}_q(\Sigma_1 \mid n + 3, nS + x_1 x_1')$$

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 $\Sigma_1$. Integrating the elements used to compute $\Sigma_1$ (a component of $\Delta_2$), 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 $\Delta_2$. Namely, the $\Psi_2$ prior and $\Sigma_1$ posterior degrees of freedom do not match, nor do the $\Sigma_1$ posterior scale matrix and the $\Upsilon_2$ 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,

$$p(\Delta_2 \mid y_1) \approx \text{IW}(n + 2, nE[\Delta_2 \mid y_1])$$

(\text{B.3})

$$= g_{2|1}(\Delta_2 \mid y_1).$$

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

(\text{B.4})

$$g_{1|0}(\Delta_1) = p(\Delta_1) = \text{IW}(r_1, (r_1 - 2)E[\Delta_1]), \quad E[\Delta_1] = \begin{pmatrix} S & SF' \\ FS & S \end{pmatrix},$$

$$g_{1|1}(\Delta_1 \mid y_1) = p(\Delta_1 \mid y_1) = \text{IW}(r_1 + 1, (r_1 - 2)E[\Delta_1] + y_1 y_1').$$
Thus, sampling
\( g_{2|1}(\Delta_2 \mid y_1) = \text{IW}(r_2, (r_2 - 2)E_{g_{1|1}}[\Delta_2 \mid y_1]), \)
(B.5)
\( g_{2|2}(\Delta_2 \mid y_{1:2}) = \text{IW}(r_2 + 1, (r_2 - 2)E_{g_{1|1}}[\Delta_2 \mid y_1] + y_2y'_2). \)

Here, the predictive mean is derived as

\[
E_{g_{1|1}}[\Delta_2 \mid y_1] = \left( \begin{array}{c}
\Sigma_1 \\ FS_1 \\ FS_1 + \frac{nV}{n+q}(1 + \text{tr}(S_1(nS)^{-1}))
\end{array} \right)
\]
(B.6)

with \( S_1 = E_{g_{1|1}}[\Sigma_1 \mid y_1] = \frac{(r_1-2)S + x_1x_1'}{r_1-1} \). 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 \( \Sigma_2 \) is conditionally independent of \( y_1 \) given \( \Sigma_1 \).

The forward filter then recursively defines

\[
g_{t|t-1}(\Delta_t \mid y_{1:t-1}) = \text{IW}\left(r_t, (r_t - 2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] \right),
\]
(B.7)
\[
g_{t|t}(\Delta_t \mid y_{1:t}) = \text{IW}\left(r_t + 1, (r_t - 2)E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] + y_ty'_t \right),
\]
with

\[
E_{g_{t-1|t-1}}[\Delta_t \mid y_{1:t-1}] = \left( \begin{array}{c}
S_{t-1} \\ FS_{t-1} \\ FS_{t-1} + \frac{nV}{n+q}(1 + \text{tr}(S_{t-1}(nS)^{-1}))
\end{array} \right)
\]
and

\[
S_t = \frac{(r_t - 2)\left(FS_{t-1} + \frac{nV}{n+q}(1 + \text{tr}(S_{t-1}(nS)^{-1}))\right) + x_tx'_t}{r_t - 1}.
\]

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

\[
p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid \Sigma_t, y_{1:T}) = p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid \Sigma_t, y_{1:t})
\]
(B.10)
\[
\propto p(\Sigma_{t-1}, \Upsilon_t, \Psi_t \mid y_{1:t})\delta_{\Sigma_t = \Psi_t + \Upsilon_t\Sigma_{t-1}\Upsilon'_t}
\]
(B.11)
\[
\propto g_{t|t}(\Delta_t \mid y_{1:t})\delta_{\Sigma_t = \Psi_t + \Upsilon_t\Sigma_{t-1}\Upsilon'_t}.
\]
(B.12)

Thus, sampling \( \{\Sigma_{t-1}, \Upsilon_t, \Psi_t\} \) from this conditional posterior is equivalent to fixing \( \Sigma_t \) in the \( \Delta_t \) matrix and sampling a valid \( \{\Sigma_{t-1}, \Upsilon_t, \Psi_t\} \) conditioned on \( \Sigma_t \) 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. (5.7),

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

implying that

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

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 \( \{ \Sigma_{t-1}, \Psi_t \} \) conditioned on \( \Sigma_t \) and \( y_{1:t} \). Specifically, as discussed in Section 2, there exists a \( \{ \tilde{\Psi}_t, \tilde{\Psi}_t \} \) such that \( \tilde{\Psi}_t \Sigma_t = \tilde{\Sigma}_{t-1} \tilde{\Psi}_t' \) and

\[(B.15) \quad \tilde{\Psi}_t \bigg| y_{1:t} \sim \text{IW}(r_t + 1 + q, G_t^{11} - G_t^{21} (G_t^{22})^{-1} G_t^{21}),\]

\[\tilde{\Psi}_t \bigg| \tilde{\Psi}_t, y_{1:t} \sim N( (G_t^{21}) (G_t^{22})^{-1} \tilde{\Psi}_t, (G_t^{22})^{-1} )\]

with

\[(B.16) \quad \Sigma_{t-1} = \tilde{\Psi}_t + \tilde{\Sigma}_t \tilde{\Psi}_t'.\]

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

**B.3. Hyperparameter Sampling.** We examine the conditional and marginal likelihood terms of the accept-reject ratio of eqn. (5.20).

**Conditional Likelihood** \( p(x_{1:T} \big| z_{1:T}, F, S) \). Since \( x_t \sim N( \Psi_t z_t, \Psi_t ) \) with \( \Psi_t = N(F, \Psi_t, (nS)^{-1}) \) and \( \Psi_t \sim \text{IW}(n + q + 2, nV) \), marginalizing \( \Psi_t \) yields a multivariate t distribution (see eqn. (D.4)). This results in

\[(B.17) \quad p(x_{1:T} \big| z_{1:T}, F, S) = \prod_{t=1}^{T} t_{n+q+2} \left( x_t \big| F z_t, c_{n,q+2} \{ 1 + z_t'(nS)^{-1} z_t \} V \right).\]

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

\[(B.18) \quad p(z_{1:T} \big| F, S) = \int \prod_t p(z_t \big| \Sigma_{t-1}) p(\Sigma_{0:T} \big| F, S) d\Sigma_{0:T}.\]
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

\[
p(z_{1:t}) = \frac{p(z_t \mid \Sigma_{t-1}) p(\Sigma_{t-1} \mid z_{1:t-1})}{p(\Sigma_{t-1} \mid z_{1:t})} p(z_{1:t-1})
\]

(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,

\[
g_{t|t}(\Sigma_t \mid z_{1:t}) = \text{IW} (r_t, (r_t - 2)\Sigma_{t|t})
\]

\[
g_{t|t+1}(\Sigma_t \mid z_{1:t+1}) = \text{IW} (r_t + 1, (r_t - 2)\Sigma_{t|t} + z_{t+1}z_{t+1}')
\]

with \( \Sigma_{t|t} = E_{y_{t-1}|t}[\Sigma_t \mid z_{1:t}] \). The matched-means are recursively computed using

\[
(r_{t-1} - 1)\Sigma_{t-1|t} = (r_{t-1} - 2)\Sigma_{t-1|t-1} + z_tz_t'
\]

\[
\Sigma_{t|t} = F\Sigma_{t-1|t}F' + c_{n,q} \left[ 1 + \text{tr}\{\Sigma_{t-1|t}(nS)^{-1}\} \right]
\]

with initial condition \( \Sigma_{0|0} = S \).

Using this approximate filter in eqn. (B.19) for the marginal likelihood and canceling terms yields

\[
p(z_{1:t} \mid F, S) = \frac{1}{\pi T^{q/2}} \prod_{t=1}^{T} \frac{|(r_{t-1} - 2)\Sigma_{t-1|t-1}|^{(r_{t-1}+q-1)/2} \Gamma\left(\frac{r_{t-1}+q}{2}\right)}{|(r_{t-1} - 2)\Sigma_{t-1|t-1} + z_tz_t'|^{(r_{t-1}+q)/2} \Gamma\left(\frac{r_{t-1}}{2}\right)}.
\]

Note that we could have employed our filter for \( \Delta_t \) 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 the filter 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.
APPENDIX C: INVERSE WISHART AND MATRIX NORMAL DISTRIBUTIONS

A \( q \times q \) positive definite and symmetric matrix \( \Sigma \) has an inverse Wishart distribution \( \Sigma \sim IW(n, D) \) when its density function is

\[
p(\Sigma) = c|\Sigma|^{-(q+n)/2} \text{etr}\{-\Sigma D/2\}
\]

where \( c \) is the constant given by

\[
c^{-1} = |D|^{-(n+q-1)/2} 2^{(n+q-1)q/2} \pi^{q(q-1)/4} \prod_{i=1}^{q} \Gamma((n + q - i)/2).
\]

Here \( n > 0 \) is a degree of freedom parameter and \( D \) is a \( q \times q \) positive definite and symmetric matrix, referred to as the scale parameter matrix. If \( n > 2 \) then the mean is defined and is \( E(\Sigma) = D/(n-2) \). The notation is sometimes modified to explicitly reflect the dimension, via \( \Sigma \sim IW_q(n, D) \).

The inverse Wishart and Wishart are related by \( \Omega = \Sigma^{-1} \) when \( \Sigma \sim IW(n, D) \) and \( \Omega \sim W(h, A) \) with \( h = n + q - 1 \) and \( A = D^{-1} \). The map from Wishart to inverse Wishart, and back, is derived by direct transformation using the Jacobians

\[
\frac{\partial \Omega}{\partial \Sigma} = |\Sigma|^{-(q+1)} \quad \text{and} \quad \frac{\partial \Sigma}{\partial \Omega} = |\Omega|^{-(q+1)}.
\]

The constant \( c \) in the densities of each is the same, expressed in terms of either parametrisation \( (h, A) \) or \( (n, D) \). Note that both \( h \) and \( n = h - q + 1 \) are often referred to as degrees of freedom, and it is important to avoid notational confusion.

The \( r \times q \) random matrix \( A \) has a matrix normal distribution, denoted by \( A \sim N(M, U, V) \) when its density function is given by

\[
p(A) = (2\pi)^{-rq/2} |U|^{-q/2} |V|^{-r/2} \text{etr}\{-A - M\}'U^{-1}(A - M)V^{-1}/2
\]

with mean matrix \( M \) \( (r \times q) \), column (or left) variance matrix \( U \), \( (r \times r) \), and row (or right) variance matrix \( V \), \( (q \times q) \). The distribution is defined when either or both of the variance matrices are non-negative definite, and it is non-singular if and only if each variance matrix is positive definite.

Matrices \( A = (a_{i,j}) \) and \( M = (m_{i,j}) \) have rows \( a_i \) and \( m_i \), and columns \( a_{i,j} \) and \( m_{i,j} \) while the variance matrices have elements \( U = (u_{i,j}) \) and \( V = (v_{i,j}) \) for \( i = 1, \ldots, r, j = 1, \ldots, q \). All marginal and conditional distributions of elements of \( A \) are normal: \( p(A) \) has multivariate normal margins for rows, \( a'_{i,:} \sim N(m'_{i,:}, u_{i,:}) \) and for columns, \( a_{:,j} \sim N(m_{:,j}, v_{:,j}) \), for \( i = 1, \ldots, r \) and \( j = 1, \ldots, q \). For any two rows \( (i, s) \), \( \text{Cov}(a'_{i,:}, a'_{s,:}) = u_{i,s} \) and for any two columns \( (j, t) \), \( \text{Cov}(a_{:,j}, a_{:,t}) = v_{j,t} \). The marginal distribution of any pair of elements \( a_{i,j}, a_{s,t} \) is bivariate normal with \( \text{Cov}(a_{i,j}, a_{s,t}) = u_{i,s} v_{j,t} \). Stacking columns of each of \( A \) and \( M \) into \( rq \times 1 \) vectors \( \text{vec}(A) \) and \( \text{vec}(M) \) yields a multivariate normal \( \text{vec}(A) \sim N(\text{vec}(M), V \otimes U) \) where \( \otimes \) denotes Kronecker product.
APPENDIX D: MULTIVARIATE T DISTRIBUTIONS FOR \( P(X_T) \) AND \( P(X_T | Z_T) \)

The \( q \)-dimensional multivariate t distribution with \( \nu \) degrees of freedom and parameters \( \mu \) and \( \Sigma \) has density

\[
\tau_{\nu}(x | \mu, \Sigma/\nu) = a_{\nu,q} |\Sigma|^{-1/2} \left( 1 + \frac{1}{\nu} (x - \mu)' \Sigma^{-1} (x - \mu) \right)^{-(\nu+q)/2}
\]

where

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

For the proposed IW-AR, since \( x_t \sim N(0, \Sigma_t) \) and \( \Sigma_t \sim IW(n+2, nS) \), marginalizing \( \Sigma_t \) yields the multivariate t distribution with density

\[
p(x_t) = t_{n+2} \left( 0, \frac{n}{n+2} S \right).
\]

Additionally, since \( \Upsilon_t | \Psi_t \sim N(F, \Psi_t, (nS)^{-1}) \) standard theory gives \( \Upsilon_t z_t | \Psi_t, z_t \sim N(F z_t, \Psi_t (1 + z_t' (nS)^{-1} z_t)) \). Marginalizing \( \Upsilon_t \) from the distribution of eqn. (4.1), we have

\[
p(x_t | z_t) = t_{n+q+2} \left( F z_t, (1 + z_t' (nS)^{-1} z_t) \right) \frac{nV}{n + q + 2}.
\]

APPENDIX E: TRANSITION DISTRIBUTION AS A MIXTURE OF INVERSE WISHARTS

Based on eqn. (2.6) and recalling \( \phi_t = \Upsilon_t \Sigma_{t-1} \), we have

\[
\Sigma_t = \Psi_t + \phi_t \Sigma_{t-1} \phi_t'.
\]

The distribution of \( \Psi_t \) in eqn. (2.7) implies

\[
\Sigma_t | \Sigma_{t-1}, \phi_t \sim IW_q(n + q + 2, nV + (n + q)\phi_t \Sigma_{t-1} \phi_t')
\]

such that \( E[\Sigma_t | \Sigma_{t-1}, \phi_t] = E[\Psi_t] + \phi_t \Sigma_{t-1} \phi_t' \).

The mixing distribution is derived as

\[
p(\phi_t | \Sigma_{t-1}) = \int p(\phi_t | \Sigma_{t-1}, \Psi_t) p(\Psi_t) d\Psi_t,
\]
where based on properties of the matrix normal distribution,

\[(E.4) \quad \phi_t | \Sigma_{t-1}, \Psi_t \sim N(F\Sigma_{t-1}, \Psi_t, \Sigma'_{t-1}(nS)^{-1}\Sigma_{t-1}).\]

Equivalently,

\[(E.5) \quad \phi'_t | \Sigma_{t-1}, \Psi_t \sim N(\Sigma_{t-1}F', \Sigma'_{t-1}(nS)^{-1}\Sigma_{t-1}, \Psi_t).\]

Since \(\Psi_t\) is inverse Wishart distributed, standard results imply that \(\phi'_t | \Sigma_{t-1}\) is matrix T distributed (\(T\)).

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

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

\[(F.1) \quad E[\Psi_t | \Psi_{t-1}] = H\Psi_{t-1}H' + \frac{nW}{n+2q}(1 + tr(\Psi_{t-1}(nV)^{-1})).\]

Since \(E[\Sigma_t | \Delta_{t-1}] = E[\Psi_t | \Delta_{t-1}] + E[\Upsilon_t \Sigma_{t-1} \Upsilon'_t | \Delta_{t-1}],\) in deriving the conditional mean of \(\Sigma_t\) given \(\Delta_{t-1}\) we first need

\[(F.2) \quad E[\Upsilon_t \Sigma_{t-1} \Upsilon'_t | \Delta_{t-1}] = E[E[\Upsilon_t \Sigma_{t-1} \Upsilon'_t | \Psi_t | \Delta_{t-1}]\]

\[= E\left[\sum_{k,\ell} \Sigma_{t-1,kt}E[\Upsilon_{t,ik} \Upsilon_{t,j\ell} | \Psi_t] | \Delta_{t-1}\right]

\[= E\left[\sum_{k,\ell} \Sigma_{t-1,kt} \left(\Psi_{t,ij}(nS)^{-1}_{k\ell} + F_{ik}F_{j\ell}\right) | \Delta_{t-1}\right]

\[= E[tr(\Sigma_{t-1}(nS)^{-1})\Psi_{t,ij} + F_i \Sigma_{t-1} F'_j | \Delta_{t-1}]

\[= tr(\Sigma_{t-1}(nS)^{-1})E[\Psi_{t,ij} | \Delta_{t-1}] + F_i \Sigma_{t-1} F'_j,

implying

\[(F.3) \quad E[\Upsilon_t \Sigma_{t-1} \Upsilon'_t | \Delta_{t-1}] = \Psi_{t-1} + F\Sigma_{t-1} F'.\]

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

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

with \(E[\Psi_t | \Psi_{t-1}]\) as in eqn. (F.1).
In the limit as $n \to \infty$, we have
\begin{align*}
E[\Sigma_t | \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} - HY_{t-1}\Sigma_{t-2}Y_{t-1}')H' \\
&= S - FSF' + F\Sigma_{t-1}F' \\
&\quad + H((\Sigma_{t-1} - Y_{t-1}\Sigma_{t-2}Y_{t-1}') - (S - FSF'))H' \\
&= S + F(\Sigma_{t-1} - S)F' \\
&\quad + H((\Sigma_{t-1} - S) - (Y_{t-1}\Sigma_{t-2}Y_{t-1}' - FSF'))H'.
\end{align*}

APPENDIX G: 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 $a_i = \text{diag}(A_i)$. The autoregressive process of eqn. (6.1) can be equivalently represented as
\begin{equation}
\xi_t = \begin{bmatrix} \text{diag}(\xi_{t-1}) & \cdots & \text{diag}(\xi_{t-r}) \end{bmatrix} \begin{bmatrix} a_1' & \cdots & a_r' \end{bmatrix} + x_t.
\end{equation}

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

APPENDIX H: EXTENDED FIGURES FROM THE EEG ANALYSIS

We begin with a brief exploratory analysis of the appropriateness of inverse Wishart margins for $\Sigma_t$ in the EEG data by examining the implied multivariate t distribution for $x_t$ in eqn. (D.2), where $x_{1:T}$ is estimated from the $q$-variate observations $\xi_{1:T}$ as described in the main paper. The hyperparameter $S$ was set to the prior mean and a range of values of $n$ were examined from $n = 1$ to $n = 10$. Results were very similar for $n$ in the range of 3 to 6. In Figure 7, we examine QQ plots for $n = 6$ and the margins $x_{t,i}$: the empirical versus theoretical quantiles for components $i = 1, \ldots, 9$. Theoretical quantiles are estimated from a large set of multivariate t samples.

The subsequent three figures provide additional outputs from the EEG data analysis exemplifying the application of the IW-AR(1) model for multivariate volatility matrices in dynamic state space models.
FIG 7. QQ plots of the implied $x_t$ multivariate t margin of eqn. (D.2) using $n = 6$. The plots examine the empirical quantiles (x-axis) versus theoretical quantiles (y-axis) for components $x_{t,i}$, $i = 1, \ldots, 9$. 
FIG 8. An extension of Figure 3 to include estimated trajectories of volatilities $\Sigma_{i,i,t}^{1/2}$ 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 9. An extension of Figure 4 to include estimated trajectories of covariance terms $\Sigma_{ij,t}$ for $i \neq j$ for $j = 1$ (EEG channel 7) colored as in Figure 3.
FIG 10. An extension of Figure 5 to include estimated trajectories of correlations between each of 6 channels and all other channels as a function of time.

EMILY FOX
DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON BOX 354322
SEATTLE, WA 98195-4322
E-MAIL: ebfox@stat.washington.edu

MIKE WEST
DEPARTMENT OF STATISTICAL SCIENCE
DUKE BOX #90251
DURHAM NC 27708-0251
E-MAIL: Mike.West@duke.edu