Bayesian Latent Threshold Modeling:
Multivariate Time Series and Dynamic Networks

Duke University, Statistical Science Discussion Paper 2014-01
January 2014

Jouchi Nakajima
Bank of Japan &
Tokyo, Japan

Mike West
Duke University
Durham NC, USA

Abstract: We discuss dynamic network modeling for multivariate time series, exploiting dynamic variable selection and model structure uncertainty strategies based on the recently introduced concept of “latent thresholding.” This dynamic modeling concept addresses a critical and challenging problem in multivariate time series and dynamic modeling: that of inducing formal probabilistic structures that are able to dynamically adapt to temporal changes in the practical relevance/existence of relationships among variables, overlaying the more traditional needs for adapting to and estimating time variation in strengths of relationships when they exist. Bayesian methodology based on dynamic latent thresholding has shown its utility in initial studies in dynamic regression, econometric and financial factor models. From this basis, the current paper involves focused development of dynamic latent thresholding for time-varying, vector autoregressive (TV-VAR) models in increasingly high-dimensions, with the theme of inference on dynamics in network structure. We develop latent thresholding models for both time-varying VAR coefficient matrices and innovation precision matrices. This induces novel classes of dynamic processes over directed and undirected associations in multivariate time series, reflecting dynamics of lagged and contemporaneous “network” dependencies. Applied analyses involving foreign currency exchange rate (FX) time series and electroencephalography (EEG) time series in neurophysiology exemplify latent thresholding as a flexible approach to inferring patterns of temporal change in both structure and strength of existing relationships, opening up new methodology for dynamic network evaluation and prediction.


1 Introduction

In their seminal paper on modeling multiple time series (one of a series of influential time series papers of George Tiao and co-authors during the 1970s-90s), Tiao & Box articulate the need to: “understand the dynamic relationships ... (among multiple series that may be) contemporaneously related (and allowing for the reality that) one series may lead the others or (that) there may be feedback relationships” (Tiao and Box 1981). That paper on vector ARMA models can be regarded as foundational, and a key step in one of several paths that led to what are nowadays mainstream classes of “workhorse” models in multivariate time series and dynamic modeling, for forecasting, control and scientific structure evaluation: vector AR (VAR) and time-varying vector AR (TV-VAR) models.

As we engage in analysis of increasingly high-dimensional time series, the need for increasing attention to constraints on model parameterizations is heightened. While Bayesian methods for shrinkage of parameters– including full shrinkage to zero when data so suggest– are well developed, the transfer of basic Bayesian sparsity modeling ideas to dynamic contexts has continued to challenge methodological researchers. We address this challenge in the current paper. In the spirit of Tiao and Box (1981), we present developments of dynamic Bayesian modeling that address head-on– and develop a new and effective modeling approach to– the challenging questions of isolating and inferring contemporaneous and lagged relationships among multiple series. A critical focus for us is in admitting that these
relationships may change over time in their relevance— that is, whether they exist in a practical sense in any time period— as well as allowing for time-varyation in strengths of relationships when they are deemed to exist.

In the literature, VAR models have become standard tools (e.g. Valdés-Sosa 2004; Valdés-Sosa et al. 2005). A typical approach to network structure inference uses ideas such as statistical assessment of causality (Granger 1969). From another perspective, graphical modeling has been exploited for evaluating contemporaneous associations (Whittaker 1990; Lauritzen 1996; Dobra et al. 2004; Jones et al. 2005; Carvalho and West 2007). These approaches concern models with constant parameters and structure. However, our interest is in time-variation in patterns of dependencies. We refer to dynamic dependencies as defining dynamic network relationships, linking to growing literatures on dependencies in networks represented by empirical statistical models; here, TV-VAR models in which TV-VAR parameters, both contemporaneous and lagged, may shrink fully to zero for periods of time, while taking time-varying non-zero values during other epochs. TV-VAR models are increasingly popular (e.g. Fujita et al. 2007) in network studies. In this class of models— a key “workhorse” for applied multivariate time series analysis that for some years has helped define the success of Bayesian methods in dynamical systems— increasing dimension brings the increasing need for data-respecting constraints and coherent model structures. Latent thresholding is a generic and, as we demonstrate, effective approach to both global and local— in time— shrinkage to zero of TV-VAR parameter processes.

Our general framework builds on Bayesian sparsity modeling and incorporates a latent threshold model (LTM) concept. The mechanism that defines parsimony in dynamic time series model can effectively induce dynamic networks among responses. contacting related literatures, we note that Valdés-Sosa et al. (2005) employ sparsity modeling for VAR coefficients with standard penalized regression techniques in analysis of brain functional connectivity. In another line of research, Yoshida et al. (2005) propose a dynamic linear model with Markov switching for estimating time-dependent gene network structure. Our LTM approach differs from these approaches in that threshold mechanisms operate continuously in the parameter space and over time; as remarked above, this defines an holistic approach to modeling and representing arbitrary patterns of change over time in sparsity structure of network dependencies, as well as time-varying strengths of existing dependencies. A salient feature is our use of Cholesky-type models of volatility matrices that enables parallel computation via sets of univariate dynamic regressions (Lopes et al. 2012). This dramatically impacts and enables scalability to increasingly high-dimensional settings.

Some notation: We use the distributional notation $d \sim U(a,b), p \sim B(a,b), v \sim G(a,b)$, for the uniform, beta, and gamma distributions, respectively, and $y \sim N(m,V)$ for the multivariate normal with dimensions implicit or specifically stated. We also use $s : t$ to denote $s, s+1, \ldots, t$ when $s < t$, for succinct subscripting; e.g., $y_{1:T}$ denotes $\{y_1, \ldots, y_T\}$.

Section 2 summarizes the general class of TV-VAR$(p)$ models adopted and discusses aspects of model form, specification and sparsity. Section 3 describes the approach to dynamic sparsity using latent threshold modeling and its application in TV-VAR models with time-varying multivariate volatility. Section 4 outlines Bayesian analysis of the proposed model and discusses computation for model fitting, linking to recently published discussions of the idea in financial and econometrics applications with dynamic factor and regression models. Some of the scope as an approach to evaluating network dynamics is exhibited in two studies. The first, in Section 5, is an econometrics example concerning FX time series and exploring dynamic dependencies in international financial markets. The second, in Section 6, concerns dynamic network dependency analysis in multivariate EEG time series to elucidate brain functional connectivity. Section 7 provides some summary comments.
2 General Model Form

For the $m \times 1$-vector time series $y_t, (t = 1, 2, \ldots)$, we focus on TV-VAR($p$) models of the form

$$A_t y_t = \sum_{j=1}^{p} B_{jt} y_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Delta_t), \quad (2.1)$$

where

$$A_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -a_{21t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -a_{m1t} & \cdots & -a_{m,m-1,t} & 1 \end{pmatrix}, \quad \Delta_t = \begin{pmatrix} \sigma^2_{1t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2_{mt} \end{pmatrix} \quad (2.2)$$

and, for $j = 1 : p$, $B_{jt}$ is the $m \times m$ lag–$j$ matrix of TV-VAR coefficients. The model can obviously be extended to include time-varying intercepts and dynamic regressions on known predictors, although these are side-issues in the context of our main goals in this paper. We highlight the following.

**Implied reduced model form:** The model can be recast as

$$y_t = \sum_{j=1}^{p} \Gamma_{jt} y_{t-j} + u_t, \quad u_t \sim N(0, \Omega_t^{-1}), \quad \Omega_t = A_t^t \Delta_t^{-1} A_t \quad (2.3)$$

where $\Gamma_{jt} = A_t^{-1} B_{jt}$ for $j = 1 : p$. The innovations precision matrix $\Omega_t$ implies a Cholesky-style decomposition for the covariance matrix $\Omega_t^{-1} = A_t^{-t} \Delta_t A_t^{-1}$.

**Contemporaneous and lagged dependencies:** The structural model form of eqn. (2.1) provides key benefits in both interpretation and technical aspects of model fitting. With the diagonal residual covariance matrix $\Delta_t$, this form makes it explicit that $A_t$ alone defines the patterns of contemporaneous dependencies among the univariate elements of $y_t$, with the $B_{jt}$ then defining the lagged dependencies after conditioning on the contemporaneous relationships. This is the fundamental point for model interpretation.

**Identification:** A first key technical benefit of the structural model specification is that the model suffers no identification problems, and so requires no additional constraints on elements of the TV-VAR coefficient matrices; all relevant constraints (including dynamic sparsity) will come from the data: model fitting.

**Analysis decoupling and scalability:** A second, critical technical benefit is analysis decoupling. Note first that the model can be recast as a triangular set of univariate dynamic regressions:

$$y_{1t} = b_{1t}' z_{t-1} + \varepsilon_{1t},$$
$$y_{2t} = b_{2t}' z_{t-1} + a_{21t} y_{1t} + \varepsilon_{2t},$$
$$y_{3t} = b_{3t}' z_{t-1} + a_{31t} y_{1t} + a_{32t} y_{2t} + \varepsilon_{3t},$$
$$\vdots$$
$$y_{mt} = b_{mt}' z_{t-1} + a_{m1t} y_{1t} + \cdots + a_{m,m-1,t} y_{m-1,t} + \varepsilon_{mt}, \quad (2.4)$$

where

- $z_{t-1}$ is the $mp \times 1$ vector defined by $z_{t-1}' = (y_{1t-1}', \ldots, y_{t-p}');
• For each \( i = 1 : m \), the \( mp \times 1 \) vector \( b_{it} \) is formed by catenating the vectors defining row \( i \) in each of the \( B_{jt} \) in the order \( j = 1 : p \);
• \( \varepsilon_{it} \sim N(0, \sigma_{it}^2) \), for \( i = 1, \ldots, p \);
• \( \text{Cov}(\varepsilon_{it}, \varepsilon_{js}) = 0 \), for \( i \neq j \) as well as for all \( t, s \).

That is, for \( i = 1 : m \) the univariate series \( i \) is given by
\[
y_{it} = g'_{it} x_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_{it}^2),
g'_{it} = (b'_{it}, a_{i1t}, \ldots, a_{i,i-1}, t),
x'_t = (z'_{t-1}, y_{it}, \ldots, y_{i-1,t}).
\] (2.5)

Under conditionally independent model structures and priors over the coefficient processes and parameters across these \( m \) equations (as we specify below) it is clear that the \( m \) models can be analyzed separately, and in parallel. This eases computational burdens and enable scalability in \( m \), so offers a major advantage for analysis and prediction in higher-dimensional problems.

**Time-varying parameters:** All non-zero parameters in \( A_t, B_{1:p,t}, \Delta_t \) are in general time-varying; this allows for some to be more stable and even roughly constant over time, while others may vary more substantially.

**Sparsity and parsimony:** Fundamental to our main interests here, some or many of the lower triangular entries in \( A_t \) and elements of the \( B_{jt} \) may be exactly zero; that is, these matrices will generally have some sparsity patterns, and the degrees of sparsity will generally be higher for larger \( m \). The latter relates to the natural appearance– and practical need for– increasing sparsity of model parametrizations as dimension increases.

**Dynamic sparsity:** Sparsity patterns may be dynamic. An element of \( A_t \) or \( B_{jt} \) may be zero for some period of time, but then non-zero with time-varying values over other epochs. We refer to this as dynamic sparsity, and it is a main focus of the latent threshold mechanism in the following section.

**Multivariate volatility:** The structuring of the time-varying precision matrix of the reduced form innovations shows that specified models of time-variation in \( (A_t, \Delta_t) \) implicitly define a multivariate volatility model for \( \Omega_t \) and its inverse. Cholesky-style multivariate volatility models of various forms are becoming increasingly popular in areas including finance and econometrics (Pinheiro and Bates 1996; Smith and Kohn 2002; Primiceri 2005; Nakajima and West 2013a), and they seem to be opening the path to more practically relevant, useful and elaborate models. Latent thresholding for dynamic sparsity of \( A_t \) enhances this potential, as our examples below demonstrate.

**Revisiting sparsity:** Patterns of zeros in the strict lower triangle of \( A_t \) at any specific time point can translate to some off-diagonal zeros in the precision matrix \( \Omega_t = A'_t \Delta_t^{-1} A_t \). Generally, the precision will be much less sparse than \( A_t \). Similarly, sparse \( B_{jt} \) matrices will in general induce much less sparse \( \Gamma_{jt} \) matrices in the reduced model form. This again speaks to the advantages of the initial structural form of the model and the strict partitioning into contemporaneous dependencies \( (A_t) \) and then conditional lagged dependencies (the \( B_{jt} \)).

## 3 Latent Thresholding in TV-VAR Models

The latent thresholding approach to dynamic sparsity modeling was introduced in Nakajima and West (2013a) and applied to forecasting applications in econometric and financial applications there as well as in Nakajima and West (2013b). Here we summarize the approach and key ideas in the context of the general class of TV-VAR\((p)\) models with multivariate volatility given in eqns. (2.1) and
(2.2), focusing on the use as a model to interrogate dynamics in relationships among the network of \( m \) series.

In the implied decoupled model forms of eqns. (2.5), relabel the elements of \( g_{it} \) as \( g_{ijt} \) so that \( g_{it} = (g_{i1t}, \ldots, g_{ik_{t}})' \) where \( k_{t} \) is the vector length; this is only for notational convenience. The latent thresholding approach defines

\[
g_{ijt} = \gamma_{ijt} s_{ijt} \quad \text{with} \quad s_{ijt} = I(|\gamma_{ijt}| \geq d_{ij}), \quad j = 1, \ldots, k_{t},
\]

where \( I(\cdot) \) is the indicator function and \( d_{i} = (d_{i1}, \ldots, d_{ik_{t}})' \) is a vector of latent thresholds \( d_{ij} \geq 0 \), for \( i = 1, \ldots, k_{t} \). The effective regression coefficient processes \( g_{it} \) are governed by underlying or latent processes \( \gamma_{it} \equiv (\gamma_{i1t}, \ldots, \gamma_{ik_{t}})' \). Various model assumptions can be made for these processes. Here, following Nakajima and West (2013a,b), we adopt arguably the simplest model form, taking each \( \gamma_{ijt} \) as a stationary AR(1) process,

\[
\gamma_{ijt} = \mu_{ij} + \phi_{ij} (\gamma_{ij,t-1} - \mu_{ij}) + \eta_{ijt}, \quad \eta_{ijt} \sim N(0, s_{ij}),
\]

with \( |\phi_{ij}| < 1 \), for \( j = 1, \ldots, k_{i} \), and assuming independence over \( i, j \).

A key structure here is that a regression coefficients collapses to zero when its absolute value falls below the threshold. The shrinkage region \((-d_{ij}, d_{ij})\) defines temporal variable selection; the covariate plays a role in predicting the response only when the associated latent parameter \( \gamma_{ijt} \) is “large enough”; the relevance of the covariate is hence dynamic, and the model can induce dynamic sparsity. Model fitting then allows the data to infer patterns of change over time in both the zero/non-zero patterns for coefficients as well as their time-varying values when deemed non-zero. Posterior inferences on the \( \gamma_{it} \) directly yield inferences on the elements of \( A_{it} \) and each of the \( B_{jt} \), and hence on dynamic sparsity and time-varying structure of both contemporaneous and lagged network relationships among the series. That is, we have a modeling strategy for dynamics in the structure and quantified form of networks in terms of both undirected associations between multivariate responses at a point in time, and of lagged/directional associations. Critically, the underlying AR(1) latent processes allow for ranges of variation in both the smoothness of temporal changes as well as the zero/non-zero process that thresholding defines.

There are various possible choices of models for the residual volatilities \( \sigma_{it} \), (\( i = 1, \ldots, m \)). In our examples, we exhibit two of the possibilities. The study of FX data uses the standard AR(1) univariate stochastic volatility process model that is commonly applied in financial time series and related areas (e.g. Jacquier et al. 1994; Kim et al. 1998; Aguilar and West 2000; Omori et al. 2007; Prado and West 2010, chap. 7). Here we take \( \delta_{it} = \log(\sigma_{it}^2) \) as

\[
\delta_{it} = \mu_{\delta_{i}} + \phi_{\delta_{i}} (\delta_{i,t-1} - \mu_{\delta_{i}}) + \eta_{\delta_{it}}, \quad \eta_{\delta_{it}} \sim N(0, v_{\delta_{i}}),
\]

and with \( |\phi_{\delta_{i}}| < 1 \), for \( i = 1, \ldots, m \). The resulting model for \( \Omega_{i} \) in a non-thresholded version of our model is then an extension of Cholesky-type multivariate stochastic volatility models (Lopes et al. 2012). Applying the LTM structure to the multivariate stochastic volatility process yields a (more flexible, parsimonious and scalable) dynamic sparsity variant of this basic model. In the second example on EEG data, we adopt a simpler approach to just track changes over time in the volatilities \( \sigma_{it} \), with no interest in imposing stationarity or in prediction. Here we use the venerable variance discounting model (West and Harrison 1997; Prado and West 2010, chap. 4), with immediate benefits in model fitting.
4 Bayesian Inference and Computation

Key and critical practical/technical benefits of casting the multivariate LTM TV-VAR\(p\) model in the structural form are that: (i) the decoupled univariate models can be analyzed independently and in parallel; and (ii) they each have the form of a LTM dynamic regression (eqns. (2.5) and (3.2)) for which the general MCMC methods of Nakajima and West (2013a) (and accompanying software) can be directly applied. The latter develops Bayesian analysis via Markov chain Monte Carlo (MCMC) methods, extending traditional sampling methods for dynamic univariate regression models (e.g. West and Harrison 1997; Prado and West 2010) to incorporate the latent threshold structure. These extensions add conditional posterior samplers with standard Metropolis Hastings algorithms to sample the sequence of the latent process vectors \(\gamma_{it}\), as well as the latent threshold parameters \(d_{i}\). We summarize the key components here; full details and links to software appear in Nakajima and West (2013a).

With data \(y_{1:T}\), we simulate the full joint posterior \(p(\gamma_{1:T}, \theta_{i}, d_{i}, \delta_{1:T}, \theta_{i}|y_{1:T})\), including the set of latent process state parameters and model hyper-parameters, for each of the models \(i = 1 : m\) in parallel. For equation \(i\), these are:

- The dynamic regression parameter process states, \(\gamma_{i,1:T} = \{\gamma_{it} | t = 1 : T\}\);
- Hyper-parameters defining each of the component univariate AR(1) models, \(\theta_{i} = \{\mu_{ij}, \phi_{ij}, v_{ij}; j = 1 : k_{i}\}\);
- The vector of latent thresholds, \(d_{i} = (d_{i1}, \ldots, d_{ik_{i}})^{T}\);
- The log volatility process \(\delta_{i,1:T}\) and its hyper-parameters \(\vartheta_{i} = \{\mu_{bi}, \phi_{bi}, v_{bi}\}\).

We outline key analysis components for each of the sets of parameters and states; see also Appendix A of Nakajima and West (2013a) for full technical details.

**Dynamic regression parameter process states** \(\gamma_{i,1:T}\): Conditional on model hyper-parameters, volatility process states and the data, we use a Metropolis-within-Gibbs sampling strategy for simulation of \(\gamma_{i,1:T}\); this sequences through each \(t\), using a Metropolis-Hastings sampler for each \(\gamma_{it}\) given \(\gamma_{i-t} = \gamma_{i,1:T}\backslash\gamma_{it}\). We draw a Metropolis proposal of \(\gamma_{it}\) trivially from the underlying non-threshold model obtained by ignoring the latent threshold, i.e., setting \(s_{ijt} = 1\); the resulting proposal distribution is simply a normal conditional posterior distribution from the dynamic linear regression model. The step is completed by accepting the proposal with the relevant Metropolis-Hastings probability. Analysis iterates through this process over \(t = 1 : T\) to resample the entire sequence of \(\gamma_{it}\) vectors.

**AR hyper-parameters** \(\theta_{i}\): With prior independence across \(i = 1 : m\) and \(j = 1 : k_{i}\), we use traditional forms of priors for the AR model parameters \((\mu_{ij}, \phi_{ij}, v_{ij})\). Analysis uses normal prior for \(\mu_{ij}\), shifted beta priors for \(\phi_{ij}\), and inverse gamma priors for \(v_{ij}\). Resulting conditional posteriors are be sampled directly or via Metropolis-Hastings accept/reject steps.

**Latent threshold parameters** \(d_{i}\): Nakajima and West (2013a) discuss in detail the specification of priors on the threshold parameters, and we use the same forms here. Specifically,

\[
d_{ij}(\mu_{ij}, \phi_{ij}, v_{ij}) \sim U(0, |\mu_{ij}| + Kw_{ij}^{1/2})
\]

where \(w_{ij} \equiv v_{ij}/(1 - \phi_{ij}^{2})\). This is a uniform prior distribution ranging over a range based on the stationary scale of the corresponding latent process \(\gamma_{ijt}\). For \(K\) fixed at a value in the \(3 - 5\) range, this prior conditionally covers the range of the process \(\gamma_{ijt}\) as reflected in its stationary distribution. Within the MCMC, thresholds are then sampled independently across \(i\) via Metropolis-Hastings steps with candidates drawn from the conditional uniform prior based on “current” values of the conditioning parameters \((\mu_{ij}, \phi_{ij}, v_{ij})\).
Table 1: FX return data: 20 international currencies, ordered by the higher turnover in the global foreign exchange market in April 2007 reported by BIS (2010).

<table>
<thead>
<tr>
<th></th>
<th>Currency</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EUR Euro</td>
<td>1 EUR</td>
</tr>
<tr>
<td>2</td>
<td>JPY Japanese Yen</td>
<td>2 JPY</td>
</tr>
<tr>
<td>3</td>
<td>GBP British Pound Sterling</td>
<td>3 GBP</td>
</tr>
<tr>
<td>4</td>
<td>CHF Swiss Franc</td>
<td>4 CHF</td>
</tr>
<tr>
<td>5</td>
<td>AUD Australian Dollar</td>
<td>5 AUD</td>
</tr>
<tr>
<td>6</td>
<td>CAD Canadian Dollar</td>
<td>6 CAD</td>
</tr>
<tr>
<td>7</td>
<td>SEK Swedish Krona</td>
<td>7 SEK</td>
</tr>
<tr>
<td>8</td>
<td>NOK Norwegian Krone</td>
<td>8 NOK</td>
</tr>
<tr>
<td>9</td>
<td>NZD New Zealand Dollar</td>
<td>9 NZD</td>
</tr>
<tr>
<td>10</td>
<td>SGD Singapore Dollar</td>
<td>10 SGD</td>
</tr>
<tr>
<td>11</td>
<td>KRW South Korean Won</td>
<td>11 KRW</td>
</tr>
<tr>
<td>12</td>
<td>ZAR South African Rand</td>
<td>12 ZAR</td>
</tr>
<tr>
<td>13</td>
<td>RUB Russian Ruble</td>
<td>13 RUB</td>
</tr>
<tr>
<td>14</td>
<td>INR Indian Rupee</td>
<td>14 INR</td>
</tr>
<tr>
<td>15</td>
<td>BRL Brazilian Real</td>
<td>15 BRL</td>
</tr>
<tr>
<td>16</td>
<td>TWD Taiwanese Dollar</td>
<td>16 TWD</td>
</tr>
<tr>
<td>17</td>
<td>THB Thai Baht</td>
<td>17 THB</td>
</tr>
<tr>
<td>18</td>
<td>IDR Indonesian Rupiah</td>
<td>18 IDR</td>
</tr>
<tr>
<td>19</td>
<td>CLP Chilean Peso</td>
<td>19 CLP</td>
</tr>
<tr>
<td>20</td>
<td>PHP Philippine Peso</td>
<td>20 PHP</td>
</tr>
</tbody>
</table>

Stochastic volatility processes \((\delta_{i,1:T}, \vartheta_i)\): In the case of models using the standard log-volatility process models, sampling the log volatility process \(\delta_{i,1:T}\) employs the standard MCMC technique (Shephard and Pitt 1997; Kim et al. 1998; Watanabe and Omori 2004; Omori et al. 2007). The associated AR(1) parameters \(\vartheta_i\) are then directly drawn using Gibbs/Metropolis-Hastings steps, based on standard priors as for \(\theta_i\). In the case of models using the simpler variance discounting approach, volatilities are exactly sampled from conditional posteriors using the forward filter/backward sampling algorithm arising from the discount model theory (Prado and West 2010, section 4.6, p143).

5 Analysis of FX Time Series Interrelationships

The first study applies the LTM-based network model to a series of daily foreign exchange (FX) rate returns from world-wide international financial markets. Our focus here is sparsity patterns on time-varying relationships underlying the FX series and induced dynamic networks among international financial markets. In econometrics, correlation-based network analysis has been discussed for studying relations among stock prices (e.g. Vandewalle et al. 2001; Onnela et al. 2004). Tse et al. (2010) describe financial networks in which nodes are individual US stocks and edges are determined by cross correlation of the stock price/return and trading volumes with some cutoff levels of sample correlation. Granger causality and cointegration models have also been developed, using rolling cointegration methods to investigate evolving pattern of interdependence among international stock markets Awokuse et al. (2009). From a rather different perspective, Carvalho and West (2007) explore dynamic Gaussian graphical models and provide analysis of FX and stock price index time series (see also Prado and West 2010, chap.10). For our FX time series analysis, we note connections with previous works using dynamic factor volatility models (Aguilar and West 2000; Lopes and West 2004; Lopes and Carvalho 2007; Nakajima and West 2013b; Zhou et al. 2012). None of these prior approaches is able to address time-varying parameter issues, nor the critical questions of time-variation in the existence/non-existence of practically relevant associations, whether contemporaneous or lagged, however; this is what we address directly.

5.1 Data and priors

The data are \(m = 20\) daily international currency exchange rates relative to the US dollar over a time period of \(T = 1,304\) business days beginning in January 2007 and ending in December 2011; the currencies are listed in Table 1. The returns are computed as \(y_{it} = 100(p_{it}/p_{i,t-1} - 1)\), where \(p_{it}\) denotes the daily closing spot rate. Figure 1 displays a representative return series, the EUR currency against the US dollar, and a trajectory of estimated univariate stochastic volatility process fitted to the single series as an exploratory reference. A striking feature is volatile and turbulent movements triggered by the global financial crisis around 2008. We observe other high volatility periods around the mid-2010 related to the European sovereign debt crisis.
Figure 1: EUR return series (upper) and the trajectory of its volatility as estimated in an initial univariate SV model (lower).

An important aspect of the Cholesky-type time-varying covariance/precision matrix in real data analysis is the specified ordering of series. As well discussed by Primiceri (2005), the order matters because we impose the lower triangular structure on the $A_t$. Here, the currencies are ordered by the turnover in the global foreign exchange market in April 2007 reported by BIS (2010), as in Table 1. We adopt an ordering based on the notion of dominant strength of transmission to currencies of the effects of an idiosyncratic structural shock to one currency; we regard a currency dominant in this sense based on larger trade volume, and base our chosen order on the latter for this reason. Other orderings can be argued on other bases, of course. The key point here is that of model specification based on known or expected contextual information, similar to that relied upon in specifying variable order in factor models (e.g. Aguilar and West 2000; Lopes and West 2004; Lopes and Carvalho 2007; Nakajima and West 2013b; Zhou et al. 2012). We also revisit this question from an empirical viewpoint later, via reanalysis of the data under a different ordering, in fact precisely the reverse ordering adopted here on the above basis.

Following prior work with related data, we set $p = 1$ for the VAR lag. The following priors are used: $1/\nu_{ij} \sim G(100, 0.001)$; $1/\nu_{\delta i} \sim G(3, 0.03)$; $(\phi_{ij} + 1)/2 \sim B(20, 1.5)$; $(\phi_{\delta i} + 1)/2 \sim B(20, 1.5)$; $\mu_{ij} \sim N(0, 1)$, and $\exp(-\mu_{\delta i}) \sim G(3, 0.03)$. The MCMC analysis was run for a burn-in period of 10,000 samples prior to saving the following MCMC sample of size $J = 100,000$ for summary posterior inferences. Computations were performed using custom code in Ox (Doornik 2006); the code is available to interested readers.

5.2 Summaries of posterior inferences

Figure 2 displays posterior probabilities of $s_{ijt} = 1$ for the elements of $A_t$ and the induced posterior probabilities of non-zero elements in the precision matrix $\Omega_t = A_t'\Delta_t^{-1}A_t$, at $t = 651$ (2009/Jun/30). This is the mid-time point of the sample period, when the volatility of the world market was temporarily lowered three quarters after the Lehman Brothers shock. The panel of $\Omega_t$ exhibits interesting patterns of sub-networks among international currencies. The posterior probabilities of non-zero “links” among the first 10 currencies are evidently high, which implies considerable contemporaneous associations among the currencies of major industrialized countries. The second sub-group consists of the last 10 currencies from emerging countries, exhibiting relatively more sparse structure than the
Figure 2: FX analysis: A snapshot at $t = 2009/6/30$. The heat-maps show posterior probabilities of $s_{ijt} = 1$ for elements of $A_t$, the induced posterior probabilities of non-zeros in $\Omega_t$, and posterior probabilities $s_{ijt} = 1$ for the VAR coefficients in $B_{1t}$ at this single time point.

first sub-group, although relevant edges are observed among several currencies such as KRW, RUB, INR, TWD, and IDR. The first five currencies in the second sub-group (i.e., 11:KRW–15:BRL) show spotted edges to some currencies in the first sub-group with high posterior probability. A sparse block of the left-bottom sub-matrix (colored in blue) indicates conditionally independent structure between the first sub-group and the last five currencies in the second sub-group in terms of functional, contemporaneous network connectivities.

Figure 3: FX analysis: Binary heat-map showing thresholded posterior probabilities of non-zero elements in $\Omega_t$ for select time points. Black indicates posterior probability of a non-zero entry exceeding 0.9.

The third frame in Figure 2 shows posterior probabilities of $s_{ijt} = 1$ for the VAR coefficients in $B_{1t}$ at $t = 651$ (2009/Jun/30). The first lag coefficients are considerably sparse as expected in modeling financial returns; some relevant coefficients in the KRW currency model are exceptions to this. The structure of the inferred sparsity patterns are in fact quite stable throughout the sample period. To visualize this, we provide animated summaries of posteriors– heat-maps of the full sequence of posterior sparsity probabilities– for each of $B_{1t}$, $A_t$ and $\Omega_t$ over all $t$. The FX sparsity video in the Supplementary Material confirms that changes in the sparsity probabilities are evident, but the overall network structure– in terms of high/low probabilities– is remarkably stable.

Figure 3 shows snapshots of network associations in the innovations in the reduced form of the model, in terms of posterior probabilities of non-zero entries in $\Omega_t$ plotted for selected time points. The heat-maps are binary: black indicates posterior probability of a non-zero entry exceeding 0.9. The first panel shows the estimated network on 2007/Mar/27, more than one year before the financial crisis, exhibiting a similar pattern to that of Figure 2. The second panel reveals a stable network,
but with evolving connections among some currencies during the financial crisis. Interestingly, most of the industrialized countries are in one sub-network as can be seen in the left-top sub-matrix, and a new relevant sub-network arises in the middle of the matrix around currencies (5:AUD–10:SGD). Turbulent fluctuations triggered by the US financial crisis spread over the world market, and most of the currencies are seemingly linked in those periods, although our network model provides the parsimonious structure which identifies some sub-groups of currencies. The third panel shows a further evolving network during the European sovereign debt crisis. One interesting change is that CHF has fewer incident edges than before. CHF was regarded as a safer currency than other major currencies, and therefore it moved differently from the others, experiencing a historically huge appreciation.

These posterior probabilities indicate relevance of edges in the dynamic network of associations among innovations in the reduced, causal TV-VAR($p$) model. Analysis also provides posterior estimates of time-varying parameters. Figure 4 plots the posterior means and ±2 standard deviation credible intervals of selected $a_{ijt}$ coefficients, as well as posterior probabilities of $s_{ijt} = 0$ plotted below each trajectory. The first panel shows the example of EUR (as series $j$) predicting JPY (as response series $i$). After the outbreak of the US financial crisis in September 2008, the coefficient rapidly declines. It keeps moving up and down afterwards with temporal changes in posterior shrinkage probabilities. The behavior reflects market sentiments that JPY was regarded as a safer currency, which weakened the relation of JPY with EUR when the financial markets were in a downward trend with high volatilities. The second panel shows the parallel summary for GBP rather than JPY, implying a clear drop during the European crisis from 2010 to 2011. The third panel shows the corresponding inferences on influences of GBP on CHF, exhibiting sharp declines in 2008 and 2010 related to the crisis periods. Note that, from these example results, the ability to separately infer existence of relationships (dynamic sparsity) from time-varying values of state parameters that are inferred as likely non-zero over various time periods.

We reanalyzed the data: model with the series in reverse order. Figure 5 shows posterior probabilities of non-zeros in $\Omega_t$ and $B_{1t}$ at 2010/Oct/19. The panel from the latter ordering is plotted in the same order as the former. The reverse ordering exhibits a less sparse pattern in the industrialized
countries and a sparser pattern among the emerging countries, although the lack of association between industrialized and emerging countries is robust, compared to the baseline ordering. Ordering is essentially required in the Cholesky-type time-varying models, and we recognize that addressing it formally is an open research challenge. An analysis using reversible-jump MCMC methods might address different orderings in future work, while we do note that choice of ordering is largely irrelevant if analysis is largely focused on forecasting questions.

6 Analysis of EEG Time Series Network Dynamics

This section applies the LTM-based dynamic network models to multivariate time series of electroencephalographic (EEG) data. EEG recordings are measurements of electrical potential fluctuations at various scalp locations of a human subject, reflecting physiological behavior of the brain-cell networks. Analysis of multichannel EEG traces has become a standard tool to understand mechanisms of electroconvulsive therapy (ECT), known as effective treatments for major depression with electrically induced seizures in patients (Weiner and Krystal 1994). A main statistical interest here is to explore relations across EEG channels via dynamic network modeling to reveal underlying characteristics of brain-regional patterns of communication. In previous works, various classes of dynamic time series models have been studied to describe features of EEG time series (e.g., Kitagawa and Gersch 1996; West et al. 1999; Prado et al. 2001; Prado 2010a,b; Prado and West 2010), and have begun to open up the statistical-neuroscience research area.

6.1 Data and priors

This analysis uses EEG traces recorded in one seizure of one patient, previously analyzed in West et al. (1999); Prado et al. (2001) and Prado and West (2010). A set of 19 EEG series are recorded in parallel from electrodes placed on a patient’s scalp using the International 10-20 EEG System for
Figure 6: EEG data: Scalp placement of 19 electrodes defining channels/time series.

electrode location; see Figure 6. The original data set is stored with sampling rate at 256Hz to total about 26,000 observations. Following West et al. (1999) and Prado et al. (2001), the original series are subsampled every sixth observation after removing about 2,000 observations from the beginning (up to a higher amplitude portion of the seizure). The subsampling is a standard practice adopted for computational efficiency and no significant information of the characteristics in the original data is lost (West et al. 1999). For a more detailed description of the data, see Prado (1998) and the reference listed above. We use 3,000 observations from the sub-sampled series.

We use a baseline ordering of the series as listed in Table 2(i). Starting from the vertex channel Cz, known as a central probe for brain activity that can be expected to lead some of the EEG trace development, to a degree, at other locations, we list probes in order going around inside channels clockwise and then outside channels. We arrange the baseline ordering so that the responses have this spatial configuration, as relevant in previous works. Results based on another ordering, noted in Table 2(ii), are also assessed and discussed later in this section.

<table>
<thead>
<tr>
<th>(i) Baseline</th>
<th>(ii) Factor-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Cz</td>
<td>1 Cz</td>
</tr>
<tr>
<td>11 O1</td>
<td>11 F3</td>
</tr>
<tr>
<td>2 Pz</td>
<td>2 Pz</td>
</tr>
<tr>
<td>12 T5</td>
<td>12 T3</td>
</tr>
<tr>
<td>3 P3</td>
<td>3 P3</td>
</tr>
<tr>
<td>13 T3</td>
<td>13 F7</td>
</tr>
<tr>
<td>4 C3</td>
<td>4 T5</td>
</tr>
<tr>
<td>14 F7</td>
<td>14 Fp2</td>
</tr>
<tr>
<td>5 F3</td>
<td>5 P4</td>
</tr>
<tr>
<td>15 Fp1</td>
<td>15 O2</td>
</tr>
<tr>
<td>6 Fz</td>
<td>6 C3</td>
</tr>
<tr>
<td>16 Fp2</td>
<td>16 Fp1</td>
</tr>
<tr>
<td>7 F4</td>
<td>7 C4</td>
</tr>
<tr>
<td>17 F8</td>
<td>17 F4</td>
</tr>
<tr>
<td>8 C4</td>
<td>8 T6</td>
</tr>
<tr>
<td>18 T4</td>
<td>18 F8</td>
</tr>
<tr>
<td>9 P4</td>
<td>9 O1</td>
</tr>
<tr>
<td>19 T6</td>
<td>19 T4</td>
</tr>
<tr>
<td>10 O2</td>
<td>10 Fz</td>
</tr>
</tbody>
</table>

Table 2: EEG analysis: EEG channel orderings in dynamic model specification: (i) Baseline ordering; (ii) Factor-based ordering.

We examined models with lags up to $p = 6$, seeing significant over-fitting tendency in posterior estimates and considerably sparse patterns in posterior estimates of coefficient matrices of higher lagged predictors when $p > 2$. As a result, we cut-back to look in more detail at the case of $p = 2$. That is, for $j > 2$ the inferred coefficient matrices $B_{jt}$ are basically empty, indicating the need to reduce the model order. This is one key side-benefit of the LTM approach; it can act as an automatic Bayesian Occam’s Razor, wiping out parameter processes of over-elaborate models and automatically reducing to what, per the data analysis, are appropriately parsimonious models. In the analysis, the priors used are as follows: $1/v_{ij} \sim G(200, 10^{-4})$, $(\phi_{ij} + 1)/2 \sim B(20, 1.5)$, and $\mu_{ij} \sim N(0, 1)$. For the time-varying
volatility process we have, as earlier noted, the flexible, adaptive variance discount model (West and Harrison 1997; Prado and West 2010, chap. 4) as used in previous works in related EEG studies (West et al. 1999; Prado et al. 2001; Nakajima and West 2013c), and with volatility discount factor value \( \lambda = 0.99 \). MCMC sample sizes and strategy are as in the FX study above.

6.2 Summaries of posterior inferences

Figure 7 displays posterior probabilities of \( s_{ijt} = 1 \) for the elements in \( A_t \) and induced posterior probabilities of non-zeros in \( \Omega_t \), plotted at the chosen time point \( t = 1,501 \). The panels exhibit a remarkable band pattern, indicating a spatial dependence among EEG channels. Figure 8 shows heat-maps for the induced dynamic networks of innovation processes, based on the posterior probabilities of non-zeros in \( \Omega_t \). Each sub-map is plotted at the location of its channel and each sub-map shows the posterior probabilities that the channel of its plotted location (marked by \( \otimes \)) has the edge in the induced network. The displayed heat-map image is created by linearly interpolating the probabilities over the grid defined by the approximate electrode locations. The results show high probabilities of connections in neighbors of the channel. The band pattern in the \( \Omega_t \) indicates relevant spatial relation between channels in the inside circle and in the outside circle from the baseline ordering. The posterior estimates exhibit notable asymmetry between frontal and occipital regions, as well as left and right hand sides (Prado et al. 2001; Nakajima and West 2013c). In the prefrontal region, Fp1 is simultaneously related only to neighboring frontal channels, while FP2 is related to channels over the central and even to the occipital channel in the right-hand side. Interestingly, the frontal channels, such as F7, F3, FZ, and F4, are typically conditionally independent with the parietal and occipital right sites. For temporal channels, T3 is connected to neighboring parietal and occipital regions up to T5. T5 is linked with a wide range of regions except Cz, although T6 is connected only with a few neighboring channels.

These characteristics of networks among the EEG channels are dynamic in part; Figure 9 shows an evolution of the posterior probabilities of non-zero values in \( \Omega_t \) and induced connections for selected channels. High posterior probabilities for edges connected to P3 (the second column) are observed in the left-hand side and center regions in the first period. Then, the area of high probabilities expands to the right-hand side of the scalp in the later periods. The region of the relevant edges linked with Fp1 also expands to the central sites, and the area linked with T4 around the parietal sites changes through the time periods.

As with our FX study, we provide additional summaries of posterior sparsity probabilities via temporal animations available in the Supplementary Material. The "EEG sparsity video" there displays the time evolution over the full time period of sparsity probabilities for both \( A_t \) and the implied \( \Omega_t \). These exhibit variation over time, while clearly showing very strong stability in the structure--i.e., in terms of pairs of nodes that have either high or low probability of contemporaneous linkages that are sustained over time. The video includes animation over the full time period of the “head-shot” spatial probability maps for the three selected nodes that are excerpted at just four time points in Figure 9.

Finally, we examine the ordering question via reanalysis using an alternative ordering labeled as “factor-based;” see Table 2. Prado et al. (2001) estimate a dynamic factor model for the same EEG time series and provides estimates of factor weights relating the 19 channels to the factor process. Here, the channels are arranged by their estimated posterior means of the factor weight, from high to low. These weights can be regarded as proxies of relations between channels and the underlying factor process, so implicitly link to associations among channels. Figure 10 exhibits the posterior sparsity probabilities for \( \Omega_t \) over \( t = 2,001 \) using baseline versus factor-based orderings. The network patterns are substantially similar; viewing similar displays across multiple time points confirms this general agreement, indicating that the above estimation results are reasonably robust with respect to the ordering in this analysis.
Figure 7: EEG analysis: Posterior probabilities of \( s_{ijt} = 1 \) for elements of \( A_t \), induced posterior probabilities of non-zeros in \( \Omega_t \), and posterior probabilities \( s_{ijt} = 1 \) for coefficients in \( B_{jt} \), \( j = 1, 2 \). The panels are plotted at \( t = 1,501 \). See Figure 2 for the heat-map scale.
Figure 8: EEG analysis: Posterior probabilities of $s_{ijt} = 1$ for non-zeros in $\Omega_t$, for $t = 1,501$. Each sub-map shows posterior probabilities that the channel of its plotted location (marked by $\otimes$) has an edge—i.e., a non-zero value—in the induced network. See Figure 2 for the heat-map scale.
Figure 9: Evolution of networks for EEG analysis: Induced posterior probabilities of dynamic network in $\Omega_t$ and selected network maps for select time points. $\otimes$ refers to the channel location. See Figure 2 for heat-map scale.
7 Concluding remarks

We have illustrated the use of latent threshold modeling in multivariate time series analysis using flexible TV-VAR models incorporating Cholesky-type decomposition volatility. These models allow for temporal changes in patterns of relationships among series that are either contemporaneous or lagged, while allowing for dynamics in the processes by which relationships are practically meaningful or not as defined by the novel latent threshold mechanism. This defines a general framework for assessing and inferring dynamics in network connectivities as well as a basis for parsimonious inference and prediction enabled by the dynamic Bayesian sparsity modeling approach. Substantive examples, using time series from quite different contexts, address dynamic relations among responses based on relevance of edges in induced networks, and reveal considerable utility of latent threshold mechanism in analysis of complicated high-dimensional real data. Dynamic shrinkage is the key approach to reasonably induce dynamic association with responses based on precision matrices; the Cholesky-type modeling of the covariance matrix is suitable for the latent threshold modeling as a natural extension of traditional dynamic linear models from viewpoints of high-dimensional time series modeling and efficient computational strategy.

There are a number of methodological and computational areas for further investigation. Among them, we remark possible parallelization that can exploit CUDA/GPU (Suchard, Wang, Chan, Frelinger, Cron and West 2010; Suchard, Holmes and West 2010) implementations for massively parallel computation approach, exploiting conditionally independent structure of the triangular dynamic regression equations. We use time series of $m = 19$ and 20 in this paper, but the approach is in principle scalable to much higher dimensions as a result of the model decoupling concept. Among other technical and theoretical areas, the question of variable ordering might be addressed using priors over orderings and enabled by reversible-jump MCMC methods, while models incorporating additional model components including independent variables will extend the use of such models in predictive applications.

Supplementary Material

We have noted in the text that two animations of summary posterior inferences are available in the Supplementary Material. The FX and EEG sparsity videos provide additional insights into temporal patterns of evolution of dynamic network structures—contemporaneous and lagged— in each of the two applied studies reported here.

We also provide freely available software in the Supplementary Material. All MCMC computations were performed using custom code in Ox (Doornik 2006); the code is available to interested readers.
Acknowledgements

The research reported here was partly supported by a grant from the National Science Foundation [DMS-1106516 to M.W.]. 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 Bank of Japan nor of the NSF.

References


