Bayesian Forecasting and Portfolio Decisions Using Dynamic Dependent Sparse Factor Models

Xiaocong Zhou¹, Jouchi Nakajima^{2,*}, Mike West Department of Statistical Science, Duke University, Durham 27708-0251, U.S.A.

Abstract

We extend recently introduced *latent threshold dynamic models* to include dependencies among dynamic latent factors underlying multivariate volatility. With an ability to induce *time-varying sparsity* into factor loadings, these models now also allow time-varying correlations among factors; this may be exploited to improve volatility forecasts. We couple multi-period, out-of-sample forecasting with portfolio analysis using standard and novel *benchmark neutral* portfolios. Detailed studies of stock index and FX time series include: multi-period, out-of-sample forecasting, statistical model comparisons, portfolio performance using raw returns, risk-adjusted returns and portfolio volatility. We find uniform improvements on all measures relative to standard dynamic factor models. This is due to the parsimony of latent threshold models and their ability to exploit between-factor correlations to improve volatility characterization and prediction. These advances will interest financial analysts, investors and practitioners, as well as modeling researchers.

Keywords: Bayesian forecasting; Benchmark neutral portfolio; Dynamic factor models; Latent threshold dynamic models; Multivariate stochastic volatility; Portfolio optimization; Sparse time-varying loadings.

^{*}Corresponding author. Tel: +81-3-3279-1111; Address: Monetary Affairs Department, Bank of Japan, 2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

Email addresses: sailorzxc@gmail.com (Xiaocong Zhou), jouchi.nakajima@boj.or.jp (Jouchi Nakajima), mw@stat.duke.edu (Mike West)

¹Present affiliation: UBS Investment Bank, New York, NY, U.S.A.

²*Present affiliation:* Bank of Japan, Tokyo, Japan

1. Introduction

From early Bayesian approaches to factor volatility modeling (e.g. Aguilar et al., 1999; Pitt & Shephard, 1999; Aguilar & West, 2000), there has been increasing interest in refined models based on their practical benefits in financial studies, in particular (e.g. Quintana et al., 2003, 2010). Originally assuming constant factor loadings and no time dependence of latent factors for financial returns series, recent extensions introduced short-term time series models for factor loadings (Lopes & Carvalho, 2007; Carvalho et al., 2011). To date, little has been discussed about dependencies among factor processes, primarily due to the adoption of identifying constraints under which independent factor processes are mandated. With the increasing interest in sparse factor models– models in which multiple factor loadings are zero over some periods of time– this changes: such models now allow for dependencies among latent factor processes, and our main modeling goal here is to develop and exploit this in forecasting and portfolio decisions.

We make these developments in analyses of dynamic factor models using *latent thresholding*, an idea and methodology recently introduced and developed theoretically in Nakajima & West (2012a), with application to dynamic regression and time-varying VAR models. A follow-up application in Nakajima & West (2012b) adds dynamic sparsity to traditional factor models; the current paper extends this with the development of *dependent factor model* structures, novel portfolio constructions and their embedding in a complete analysis and forecasting system. In the applied studies of the paper, this is shown to be quite substantially beneficial in terms of improved forecasting performance and portfolio decision outcomes, as well as improved model fits on purely statistical criteria. We develop the applied examples using a range of stylized Bayesian portfolio decision constructs, and as part of this introduce a novel strategy that explicitly integrates a *benchmark neutral* strategy into more or less standard portfolio optimization. In addition to demonstrating the ability of sparse, dependent factor models to outperform standard models under this as well as other portfolio rules, this will be of interest to forecasting and financial decision makers in other contexts.

This work contributes modeling, forecasting and decision analytic advances to the growing literature on dynamic factor approaches in time series analysis. Beginning with earlier developments of dynamic factor models in econometric time series (e.g. Stock & Watson, 1989) and Geweke & Zhou (1996), these approaches have become popular in various macroeconomic applications (e.g. Forni et al., 2000; Stock & Watson, 2002; Koop & Potter, 2004; Bai & Ng, 2006; Del Negro & Otrok, 2008; Aruoba et al., 2009; Forni & Gambetti, 2010) as well as financial applications where multivariate volatility is represented in factor structures and other forms (e.g., Harvey et al., 1994; Pitt & Shephard, 1999; Aguilar & West, 2000; Han, 2005; Chib et al., 2006; Doz & Renault, 2006; Philipov & Glickman, 2006; Yu & Meyer, 2006; Asai et al., 2006; Fan et al., 2008). Recent developments of time-varying factor loadings models that provide part of the foundation for our work here have been particularly noted for the forecasting and statistical improvements they can generate (e.g. Lopes & Carvalho, 2007; Del Negro & Otrok, 2008). Our work builds on structural and dynamic model concepts from these areas, introducing dynamic, sparse factor models with dependencies among latent factor processes that are shown to be able to provide additional, substantial improvements in model fit, forecasting and portfolio decisions.

Section 2 summarizes the standard framework of dynamic factor models. Section 3 discusses model identification, sparse dynamic factors and the key rationale for dependent factor models. Section 4 discusses the latent thresholding concept and its application to factor models. Section 5 summarizes the new class of dynamic sparse factor models with time-varying volatility matrices allowing correlated factors. Sections 6 and 7 discuss analysis, model comparison, forecasting, and portfolio decisions in two case studies: of a 10-dimensional stock price index time series, and of a 20-dimensional FX time series. Some summary comments appear in Section 8. An appendix briefly outlines the Bayesian Markov chain Monte Carlo computational method for model fitting; this links to more extensive technical details in prior publications as well as software for interested readers.

Some notation: We use the distributional notation $\mathbf{y} \sim N(\mathbf{a}, \mathbf{A})$, $d \sim U(a, b)$, $p \sim B(a, b)$, $v \sim G(a, b)$, for the normal, uniform, beta, and gamma distributions, respectively. We use diag (a_1, \ldots, a_k) to refer a diagonal matrix whose diagonal elements are the arguments. We also use s : t to denote $s, s + 1, \ldots, t$ when s < t, for succinct subscripting; e.g., $\mathbf{y}_{1:T}$ denotes $\{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$.

2. Basic setting and background: Traditional dynamic factor models

2.1. Basic model context

We begin with traditional dynamic factor models with time-varying factor loadings and volatility components, as follows. The $m \times 1$ vector response time series y_t , (t = 1, 2, ...), follows

$$\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{B}_t \boldsymbol{f}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

$$\boldsymbol{f}_t = \boldsymbol{G} \boldsymbol{f}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Upsilon}_t), \tag{2}$$

where

- $c_t = (c_{1t}, \ldots, c_{mt})'$ is the $m \times 1$ time-varying local mean at time t;
- *f*_t = (f_{1t},..., f_{kt})' is a k × 1 vector of latent factors evolving according to a VAR(1) model with a diagonal (k × k) AR coefficient matrix *G* = diag(γ₁,..., γ_k);
- $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$ is a $k \times 1$ vector of factor innovations with time-varying variance matrix Υ_t with elements v_{ijt} ;
- B_t is the $m \times k$ time-varying factor loadings matrix;
- $\boldsymbol{\nu}_t = (\nu_{1t}, \dots, \nu_{mt})'$ is a $m \times 1$ residual vector with a diagonal time-varying volatility matrix $\boldsymbol{\Sigma}_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{mt}^2)$.

The ε_s and ν_t sequences are independent and mutually independent. Equations (1,2) define a broad class of dynamic latent factor models, and variants are in routine application in financial time series. The AR coefficient matrix G in the factor evolution model is assumed constant and diagonal; this could also be relaxed for other applications in which a persistent but time-varying matrix may be of interest, although the applications here do not suggest such an extension for the current analyses.

To ensure mathematical identification of factor models, and as a matter of modeling choice, we use the traditional lower triangular constraint on the dynamic factor loadings matrix process B_t (e.g. Geweke & Zhou, 1996; Aguilar & West, 2000; Lopes & West, 2004). Noting that B_t is "tall and skinny" – that is, the number of factors k will typically be far less than the number of series m– the upper triangular elements are $b_{ijt} = 0$ for $k \ge j > i \ge 1$, and the main diagonal elements are $b_{iit} = 1$ for i = 1 : k. Importantly, the traditional model has a *diagonal* factor innovations volatility matrix Υ_t , while admitting volatility models for the diagonal elements.

2.2. AR models for dynamic parameter processes

Complete model specification requires specific structures for the time-varying parameter processes c_t, B_t and the diagonal Σ_t and Υ_t . The simplest and most widely used are basic AR(1) models for univariate parameters, as follows.

Factor loadings B_t . Univariate AR(1) models for univariate factor loadings are increasingly popular in the literature (e.g. Lopes & Carvalho, 2007; Del Negro & Otrok, 2008). For each i = 2 : m and j < i, denote by β_{ijt} the loading relating series (row) i to factor (column) j in B_t , recalling that the upper right triangle elements are zero and the diagonals are unity. The AR(1) models are

$$\beta_{ijt} = \mu_{\beta ij} + \phi_{\beta ij}(\beta_{ij,t-1} - \mu_{\beta ij}) + \eta_{\beta ijt}, \quad \eta_{\beta ijt} \sim N(0, v_{\beta ij}), \tag{3}$$

with $|\phi_{\beta ij}| < 1$. Each β_{ijt} follows a (typically highly persistent) autoregressive process, allowing for time variation that may be substantial for some loadings, but close to constant for others.

Residual volatility matrix Σ_t . We use standard stochastic volatility models (e.g. Jacquier et al., 1994; Kim et al., 1998; Aguilar & West, 2000; Omori et al., 2007; Prado & West, 2010, chap. 7) for the set of univariate residual variances in the diagonal matrix Σ_t . With $\delta_{it} = \log \sigma_{it}^2$, this assumes a stationary AR(1) model for each *i*, given by

$$\delta_{it} = \mu_{\delta i} + \phi_{\delta i} (\delta_{i,t-1} - \mu_{\delta i}) + \eta_{\delta it}, \quad \eta_{\delta it} \sim N(0, v_{\delta i}), \tag{4}$$

with $|\phi_{\delta i}| < 1$ for each i = 1 : m.

Factor innovations volatility matrix Υ_t . In this standard model, $\Upsilon_t \equiv \Psi_t = \text{diag}(\psi_{1t}^2, \dots, \psi_{kt}^2)$, a diagonal matrix of latent factor volatilities. For these, we use the same, standard univariate stochastic volatility models as for the σ_{it}^2 . That is, with $\lambda_{it} = \log \psi_{it}^2$ we have

$$\lambda_{it} = \mu_{\lambda i} + \phi_{\lambda i} (\lambda_{i,t-1} - \mu_{\lambda i}) + \eta_{\lambda it}, \quad \eta_{\lambda it} \sim N(0, v_{\lambda i}), \tag{5}$$

with $|\phi_{\lambda i}| < 1$ for each i = 1 : k.

Level process c_t . This may be modeled in different ways, with opportunity for short-term prediction based on chosen independent variables and/or direct intervention. In model comparison and assessment here, we are interested in relative fit and forecasting performance of various models that differ in assumptions about the nature of the latent factor process, so we do not develop more customized components for c_t . We simply adopt a framework in which this term is also represented via a set of stationary, independent AR(1) processes,

$$c_{it} = \mu_{ci} + \phi_{ci}(c_{i,t-1} - \mu_{ci}) + \eta_{cit}, \quad \eta_{cit} \sim N(0, v_{ci}),$$
(6)

with $|\phi_{ci}| < 1$ for each i = 1 : m.

3. Sparse factor loadings and dependent factors

For larger numbers of time series m, it becomes more and more important to induce additional parsimony via some zeros in the (lower triangular entries of the) factor loadings matrices. This embodies the view that, while several latent factors underlie covariation in the response series, each factor will typically influence only a subset of the responses. It becomes increasingly untenable to link each response series to each factor for larger m, k; the more zeros appearing in B_t - that is, the *sparser* B_t is- the more stable and efficient should be resulting inferences, and the more easily interpretable will be the model.

Sparsity inducing priors for parameter matrices are increasingly adopted in dynamic modeling (e.g. Carvalho & West, 2007; George et al., 2008; Korobilis, 2012; Wang, 2010), and the application to factor loadings matrices natural translates from developments in sparse factor models in other areas (West, 2003; Carvalho et al., 2008; Yoshida & West, 2010). Below we discuss the specific approach to inducing time-varying patterns of sparsity based on latent threshold mechanisms. First, we discuss how this leads naturally to an interest in *dependent factor processes*– one of the key contributions of this paper.

The standard factor model of Section 2 is mathematically identified based on the assumed lower triangular form of B_t together with diagonal $\Upsilon_t = \Psi_t$. For example, rotating the factors via $B_t f_t = B_t^* f_t^*$ where $B_t^* = B_t L_t$ and $f_t^* = L_t^{-1} f_t$ for any non-singular $k \times k$ matrix L_t , destroys the lower triangular structure: B_t^* will be a full matrix, in general. However, taking L_t to be lower triangular with diagonal elements of unity means that B_t^* still has zeros above the diagonal and unit diagonal entries, so we have an identification problem unless we impose further constraints; assuming the factors to be uncorrelated, with $\Upsilon_t = \Psi_t$, diagonal, is the traditional, default constraint. Otherwise, there is a continuum of models defined by arbitrary patterns of factor dependencies– choices of L_t – that yield the same model.

A key point in this paper is to recognize that moving to *sparse* models releases this constraint; it allows for non-diagonal Υ_t , i.e., *dependent factors*. See this as follows. With B_t sparse below the diagonal, $B_t L_{t-}$ for any (non-diagonal) $k \times k$ lower triangular, unit diagonal matrix L_t - is *less sparse* than B_t and, indeed often is full. Any non-degenerate L_t at least changes the *pattern of zeros* in mapping from B_t to B_t^* , if not removing all zeros. Hence, a model with a specific pattern of zeros *cannot* be subjected to such a factor rotation without changing or losing the pattern of zeros; that is, zeros in the lower part of B_t define additional identifying constraints. As a result, with *any* zeros in the lower part of B_t , the factor process can have a non-diagonal variance matrix Υ_t that cannot be rotated without changing the model.

Reversing the above argument shows that correlated factor models can lead to higher degrees of sparsity- and hence increased parsimony in terms of fewer non-zero, time-varying parameters in the factor loadings matrices. A model with loadings matrix B_t^* that has few zeros can often be column-rotated to $B_t = B_t^* L_t^{-1}$ where L_t is a lower triangular matrix with diagonals of unity, and resulting in a B_t that is much sparser. The effects can be profound with larger m, when the reduction of the number of non-zero entries in the $m \times k$ matrix B_t relative to B_t^* can quickly become much larger than the additional k(k-1)/2 non-zeros in the lower triangular L_t .

This has been unrecognized and unexploited, to date; as we show below, moving to a dependent dynamic factor model this way can be exploited to yield substantial improvements in forecasting and decisions. The above discussion is quite general. Any application requires specific models for the parameter processes B_t incorporating sparsity, and the corresponding Υ_t . Latent thresholding to do this is now summarized.

4. Latent thresholding of parameter processes

4.1. Latent threshold model concept and background

We now discuss how latent thresholding ideas apply to induce (possibly many) data-respected zeros in B_t , and to allow the patterns of zeros to vary over time while estimating time-varying non-zero values. This extends prior work of Nakajima & West (2012a) who introduced the general idea, theory and methodology of latent thresholding for a range of dynamic models, with dynamic regression and timevarying VAR models as examples. Nakajima & West (2012b) applied the idea to the above traditional factor models with assumed *diagonal* forms for Υ_t . Below we extend this to apply to the non-diagonal Υ_t representing time-varying dependencies among latent factors.

The latent thresholding idea is simple: a parameter process, such as the AR(1) process β_{ijt} in equation (3), may be practically significant for some periods of times, but close to zero and, in terms of its practical impact on the model inferences and predictions, insignificant at others. A model with many small, practically insignificant parameters suffers from lack of robustness and increased noise in inferences due to estimation uncertainty for what are effectively zero terms. Parsimonious modeling will shrink such terms to zero when possible. The simple latent thresholding idea captures this through parameter process-specific thresholds; for example, the *effective* contribution to the model of β_{ijt} is set to zero so long as $|\beta_{ijt}|$ exceeds a threshold, while it maintains non-zero, time-varying values defined by equation (3) otherwise.

Latent thresholding is thus a strategy to *simplify* the traditional model to reduce parameter dimension, especially in larger models, via a natural approach to *time-varying sparsity*. Model fitting now defines

data-driven inferences on thresholds as well as sparsity patterns, and evaluation of the resulting impact of this implied dynamic parameter dimension reduction– when indicated by the data– on predictions. More detailed theoretical and methodological discussion, and examples, can be found in Nakajima & West (2012a).

4.2. Latent thresholding of factor loadings matrices

The basic AR(1) processes for the entries of B_t , above, is modified as follows. Change the notation so that b_{ijt} is now the (i, j) element of B_t . Then, the AR(1) forms of equation (3) (for i = 2 : m and j < i, j = 1 : k), define *underlying, latent processes* with the actual loadings given by

$$b_{ijt} = \beta_{ijt} s_{ijt} \quad \text{with} \quad s_{ijt} = I(|\beta_{ijt}| \ge d_{ij}), \tag{7}$$

for some *thresholds* d_{ij} . Hence the underlying β_{ijt} appear as non-zero, time-varying loadings in the observation equation (1) only when statistically relevant, with "practical significance" defined by the thresholds. Otherwise, a loading is shrunk fully to zero, embodying sparsity and parameter reduction in the factor loadings matrix. This leads to parsimonious structure in the factor model, strictly constraining some loadings to zero for periods of time, while allowing them to take time-varying, non-zero values elsewhere. The practical operation of the implicit switching mechanisms underlying this "dynamic sparsity" is of course guided by the data analysis and model fitting. Note also, importantly, that this allows for any one of the loadings to be non-zero for all time, as well as for full shrinkage of some entries to zero for all time.

5. Dependencies in sparse dynamic factor models

5.1. Dynamic models for factor innovations matrices

As discussed above, we now recognize that the diagonal limitation is overly restrictive when there is any sparsity in B_t and that allowing for factor dependencies can lead to sparser, more parsimonious models. Now, Υ_t is in general a non-diagonal time-varying variance matrix. We use the triangular reduction $A_t \Upsilon_t A'_t = \Psi_t$, where A_t is a lower triangular matrix with diagonal elements equal to one, and Ψ_t is diagonal. That is,

$$\boldsymbol{A}_{t} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{m1,t} & \cdots & a_{m,m-1,t} & 1 \end{pmatrix}$$
(8)

and $\Psi_t = \text{diag}(\psi_{1t}^2, \dots, \psi_{kt}^2)$. Note that, up to scalings by the ψ_{it} , $(A'_t)^{-1}$ is the Cholesky component of Υ_t . This decomposition of a volatility matrix is being increasingly used in related time series contexts (e.g. Pinheiro & Bates, 1996; Smith & Kohn, 2002; Cogly & Sargent, 2005; Primiceri, 2005; Lopes et al., 2010; Nakajima & West, 2012a). Our use of this parametrization here– for volatility matrices of innovations in dynamic latent factor processes– is novel, but of course builds on this prior literature.

Extending the traditional use of AR(1) parameter processes to the scalar elements of A_t gives a popular model for short-term dependent volatility matrices. Let $\alpha_t = (a_{21,t}, \ldots, a_{m,m-1,t})' = (\alpha_{1t}, \ldots, \alpha_{qt})'$ be the vector stack of the free elements in A_t , noting that q = k(k-1)/2. The underlying latent AR(1) dynamics are

$$\alpha_{ijt} = \mu_{\alpha ij} + \phi_{\alpha ij}(\alpha_{ij,t-1} - \mu_{\alpha ij}) + \eta_{\alpha ijt}, \quad \eta_{\alpha ijt} \sim N(0, v_{\alpha ij}), \tag{9}$$

with $|\phi_{\alpha ij}| < 1$.

For the univariate ψ_{it}^2 , we maintain the same, standard univariate stochastic volatility models of equation (5).

5.2. Aspects of prior specification and additional comments

Analysis is based on standard forms of priors for hyperparameters { μ_* , ϕ_* , v_* }, where * ranges over the model components {ci, βij , αij , δi , λi }. Specifically, we adopt normal or log-gamma priors for the μ_* , truncated normal or shifted beta priors for the ϕ_* , and inverse gamma priors for the v_* . We rarely have information for dependence of these parameters, so we assume prior independence. We also assume independent, truncated normal or shifted beta priors for each diagonal element of the factor AR coefficients γ_i .

A key component concerns structured priors for the latent threshold parameters d_{ij} . Evidently, we need to consider the scales of variation of the non-thresholded processes as part of the prior setup. Consider any one of the β_{ijt} AR(1) processes in equation (3). The stationary margin is $N(\mu_{\beta ij}, u_{\beta ij})$ where $u_{\beta ij} = v_{\beta ij}/(1 - \phi_{\beta ij}^2)$. If a threshold is large with respect to this distribution, then β_{ijt} will be zeroed out more often; if the threshold is small with respect to this distribution, β_{ijt} will more persistently beat the threshold and so contribute more significantly to the model. Following Nakajima & West (2012a) we use priors $d_{ij} \sim U(0, |\mu_{\beta ij}| + Ku_{\beta ij}^{1/2})$ with K = 3; this conditional prior– conditional on the hyperparameters defining the natural range of variation of the AR(1) process– spans the range of the process so allowing for smaller or larger thresholds and being otherwise diffuse to allow for adaptation to the data. Some interesting examples of posteriors on thresholds in other classes of models are given in Nakajima & West (2012a). Here we use this prior form for d_{ij} .

Some general comments on prior specification related to the AR model components are in order. As in other areas of multi-parameter modeling, while models are mathematically identifiable, some of the parameters, and parameter processes, may be only "weakly identified" in a practical sense. That is, there may be limited information on some parameter subsets and the posteriors, though proper, rather diffuse. In our models here, both the dynamic factors and factor loadings– in their non-zero phases– are time-varying latent variables; there is thus potential for fluctuations in loadings to explain observed data volatility when coupled with very low levels of variation in factors themselves. Informative priors over the autoregressive parameters for the factor loadings processes provide the necessary control to constrain their variability. We build these models with a strong prior view that the loadings parameters will vary over time, but really very slowly, consistent with low variation and high autocorrelation in the AR models, relative to the factors. This mandates the use of priors that concentrate close to one on the AR parameters while strongly favoring small innovations variances $v_{\beta ij}$. The same general point applies to prior specification for the hyperparameters of the AR processes α_{it} in the novel dependency model here.

5.3. Bayesian model fitting and forecasting via Markov chain Monte Carlo

Implementing Bayesian analysis uses Markov chain Monte Carlo (MCMC) methods, building on straightforward extensions of MCMC sampling scheme previously developed and in standard use for traditional latent factor models (Aguilar & West, 2000; Lopes & West, 2004), for latent threshold models (Nakajima & West, 2012a,b), and for univariate stochastic volatility models (Shephard & Pitt, 1997; Kim et al., 1998; Watanabe & Omori, 2004; Omori et al., 2007). The Appendix to this paper briefly outlines the overall MCMC strategy and its components, and indicates availability of efficient code used in the following applications. All aspects of posterior inference on model parameters and latent processes, and of forecast distributions to define predictions and feed into portfolio decisions, are based on large MCMC samples from posteriors based on historical data. In particular, out-of-sample/step-ahead forecasting over a time period from a baseline time point is trivially done in these models when the "current" posterior for model hyperparameters and historical trajectories of all latent processes components of the model are available as MCMC samples.

One obvious point is that, while the ensuing Bayesian analysis for model fitting provides inferences on– among other things– the full trajectories of the underlying processes β_{ijt} over time, their values are only relevant when they are non-zero.

6. A study of stock price index return series

6.1. Data and model setup

This section provides a first empirical study on daily stock index returns. The analysis uses m = 10 series of S&P Global 1200 sector index funds, listed in Table 1, over a time period of T = 1,251 business days beginning in November 2006 and ending in October 2011. The returns are computed as $y_{it} = 100(p_{it}/p_{i,t-1}-1)$, where p_{it} denotes the daily closing price.

RXI	Consumer Discretionary	6	EXI	Industrials
KXI	Consumer Staples	7	MXI	Materials
IXC	Energy	8	IXN	Technology
IXG	Financials	9	IXP	Telecommunications
IXJ	Healthcare	10	JXI	Utilities
	RXI KXI IXC IXG IXJ	RXI Consumer DiscretionaryKXI Consumer StaplesIXC EnergyIXG FinancialsIXJ Healthcare	RXIConsumer Discretionary6KXIConsumer Staples7IXCEnergy8IXGFinancials9IXJHealthcare10	RXIConsumer Discretionary6EXIKXIConsumer Staples7MXIIXCEnergy8IXNIXGFinancials9IXPIXJHealthcare10JXI

Table 1: Stock price indices: 10 sectors from S&P Global 1200 Index.

The following priors are used: $1/v_{ci} \sim G(40, 0.005)$, $1/v_{\beta ij} \sim G(20, 0.01)$, $1/v_{\alpha ij} \sim G(40, 0.005)$, $1/v_{\delta i} \sim G(2, 0.001)$, $1/v_{\lambda i} \sim G(2, 0.001)$; $(\phi_* + 1)/2 \sim B(20, 1.5)$ for all AR process components; $(\gamma_i + 1)/2 \sim B(1, 1)$; $\mu_{ci} \sim N(0, 1)$; and $\mu_{*ij} \sim N(0, 1)$ for each $* \in \{\beta, \alpha\}$; and, finally, $\exp(-\mu_{*i}) \sim G(3, 10)$ for each $* \in \{\delta, \lambda\}$. These priors follow the preceding literature to some extent, and also reflect the discussion of Section 5.2. The Bayesian MCMC analysis (see Appendix discussion) was run for a burn-in period of 20,000 samples prior to saving the following MCMC sample of size J = 100,000 for summary posterior inferences.

For the number of factors k, we examined analysis across models with up to k = 4. While model fitting using k = 3 suggests 3 clearly relevant factors, moving to k = 4 yields a fourth factor that is basically insignificant across the full time period, so indicating we may cut back; we therefore report summaries of posterior inference and portfolio study based on k = 3 factors. The selected ordering of the response time series in y_t , as listed in Table 1, leads to loosely "naming" the 3 factors as the RXI, KXI and IXC factors in order. The estimated patterns over time in the factors relate to major market events and structured changes in correlations among stock returns that can be plausibly understood in connection with these names. The structure is also such that we should expect strong dependencies among the factors: the three sectors represented by these first three indices are clearly strongly related so that, as they are used to name/define the three factors in that order, it is natural to expect strong factor dependencies as a result.

6.2. Summaries of posterior inferences

Figure 1 plots posterior estimates of the trajectories of underlying components of factor stochastic volatilities ψ_{it} and the posterior mean of the first factor f_{1t} . Recall that the ψ_{it} measure the instantaneous standard deviations of "structural" shocks underlying the dependent volatility and co-volatility patterns in the factors; their trajectories show marked differences. We see that ψ_{1t} traces global dynamics across the factors, exhibiting high volatility in the fourth quarter of 2008 reflecting the financial crisis as well as a hike in mid-2011 due to market turbulence triggered by Greece's debt crisis. In contrast, ψ_{2t} shows different and relatively lower levels of impact, but again keys out raised volatilities in the same time periods. Further, ψ_{3t} interestingly captures an early rise of volatility before the major onset of the financial crisis in 2008, presumably capturing industrial slowdown impacting the energy sector and presaging stock price collapse related heavily to the real estate industry. The estimated trajectory of the first factor f_{1t} itself shows well-identified common fluctuations among the sector stock indexes, exhibiting major events such as severely high volatility periods in the 2008 crisis onset and then again in the European sovereign debt crisis from 2010 to 2011.

To revisit the question of the number of factors, Figure 2 displays posterior estimates of the ψ_{t} trajectories from analysis using k = 4 factors. We see that the levels of the fourth component are really



Figure 1: Analysis of stock return data with k = 3: Posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of factor stochastic volatilities $\psi_{it} = \exp(\lambda_{it}/2)$ for i = 1:3, and posterior mean trajectory of the first factor f_{1t} .

practically negligible for most of the time period, indicating the irrelevance of this dimension, albeit with a minor uptick at the end. In addition, the posterior very strongly indicates that all of the time-varying loadings associated with a fourth factor are shrunk to zero across the time period via the latent threshold structure; this implies the analysis practically favours k = 3 factors.

From the k = 3 factor model, Figure 3 graphs estimated trajectories of correlations computed from the MCMC draws of $\Upsilon_t = A_t^{-1} \Psi_t (A'_t)^{-1}$. The factors are evidently correlated and exhibit interesting time variation through the sample period. Factor1 (RXI-leading) and Factor2 (KXI-leading) are very naturally highly correlated; the posteriors support values of correlations at or around 0.8 over all time, indicating the expected strong relationships as these two factors are primary drivers of consumer economics; however, they clearly also separately and individually represent other dynamics that are specific to the consumer discretionary (RXI) and staples (KXI) sectors. Correlations between Factor1 and Factor3 (IXC-leading), as well as between Factor2 and Factor3, show a marked drop during the second and third quarters of 2008, indicating diversification of stock price fluctuations across the industry-leading factors. We also note an additional temporal shift downwards in correlations around 2011 related to market turbulence in the European debt crisis. The right panels show the difference in posterior means between factors which confirm that, while the first two factors are highly dependent, they differ in practically measurable ways, especially during the recessionary period.

Figure 4 plots posterior probabilities of $s_{it} = 0$ for the factor loadings, showing considerable sparsity in the loadings matrix. The sparsity structure is clearly rather stable over the sample period. Note that the third series, IXC energy, relates strongly to the first two factors as well as defining, and indicating the need for, the third. A further notable point is the high degree of sparsity; the posterior indicates more than 50% sparsity and that this is quite stable over time. In the context of the high levels of factor dependencies inferred, this underscores the theoretical discussion of Section 3; repeat analysis constraining to independent factors– or simply directly transforming by rotation of the factors at each time– will yield a basically full factor loadings matrix. This is a nice illustration of the parsimony of the dependent sparse factor model: the identified model has about 12-13 effective non-zero parameters (3 a_{ijt} terms in A_t and around 9-10 effectively non-zero β_{ijt} in B_t); in contrast, the traditional full model with independent factors has 24 factor loadings to infer, so we have achieved around a 50% reduction.



Figure 2: Analysis of stock return data with k = 4: Posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of volatilities $\psi_{it} = \exp(\lambda_{it}/2)$, for i = 1:4.



Figure 3: Analysis of stock return data with k = 3: Posterior means and ± 1 standard deviation credible intervals (dotted) of factor innovation correlations from Υ_t (left), and difference of posterior means between factors (right).

The following section now explores the resulting implications for out-of-sample prediction and decisions based on those predictions.

6.3. Out-of-sample forecasting and portfolio analysis

A main goal of the work in this paper is improve forecasting performance of dynamic factor models by inducing sparse, parsimonious models through dependent factor structures and latent thresholding. To explore this in the stock index study, we consider dynamic portfolio allocations and make direct comparisons with the traditional models. The portfolio analysis uses sequential portfolio decisions based on short-term forecasts; we stress the exercise here involves wholly out-of-sample prediction for honest evaluation and comparison. We follow the prior literature (e.g., Quintana, 1992; Putnum & Quintana, 1994; Quintana & Putnum, 1996; Aguilar & West, 2000; Carvalho & West, 2007; Carvalho et al., 2011; Wang & West, 2009; Quintana et al., 2003, 2010) in utilizing Bayesian decision theory via Markowitz portfolio optimization.

A first study looks at one-day ahead forecasting and decisions. We use the same data set as in the previous subsection, then applying sequential portfolio reallocation analysis for an additional period of



Figure 4: Analysis of stock return data with k = 3: Posterior probabilities of $s_{it} = 0$ for factor loadings (implying full shrinkage to zero and hence a zero effective factor loading at those times).

100 business days. Specifically, we first fit the model based on the first $T_1 = 1, 151$ observations, $\boldsymbol{y}_{1:T_1}$, and compute the mean vector and covariance matrix of the implied one-step ahead forecast distribution $p(\boldsymbol{y}_{T_1+1}|\boldsymbol{y}_{1:T_1})$; this is easily done from the MCMC outputs. Then, we move ahead one business day to add the next observation \boldsymbol{y}_{T_2} ($T_2 = T_1 + 1$); we then rerun the entire MCMC based on the full updated data set $\boldsymbol{y}_{1:T_2}$, generating the next one-day-ahead forecast of \boldsymbol{y}_{T_2+1} for portfolio reallocation analysis. We repeat these sequential re-analyses for 100 business days.

At any time t - 1, the one-step ahead forecast mean vector and variance matrix of y_t are denoted by g_t and Q_t , respectively. The total sum invested on each business day is restricted by $w'_t 1 = 1$, where w_t denotes the vector of portfolio weights allocated across the 10 stock price indices. Traditional Bayesian optimization (Markowitz, 1959) is utilized subject to specific constraints for forecast portfolio mean $w'_t y_t$ and variance $w'_t Q_t w_t$. We examine three variations of allocation rules, namely:

- 1. *Efficient Frontier*: optimize portfolio weights by minimizing a weighted average of expected return and variance; specifically, minimize $w'_t Q_t w_t \kappa_t w'_t y_t$ for some risk tolerance ratio $\kappa_t > 0$.
- 2. Target return, minimum risk: given a predetermined daily return target r_t , optimize the portfolio weights by minimizing the one-step ahead portfolio variance, $w'_t Q_t w_t$, among the restricted portfolios whose one-step ahead expectation is $w'_t g_t = r_t$;
- 3. Target risk, maximum return: given a predetermined tolerance level for portfolio variance ξ_t^2 , optimize the portfolio weights by maximizing the one-step ahead expected portfolio return among restricted portfolios with that risk, i.e., maximize $w_t'g_t$ subject to $w_t'Q_tw_t = \xi_t^2$;

These portfolio allocation rules are often used in practically relevant strategies. The efficient frontier rule addresses risk aversion preferences of investors; the target return, minimum risk rule allows for a range of increasingly aggressive strategies while quantifying implied risk; the target risk, maximum return strategy aims to elucidate investor risk preferences in advance, and then helps to understand how to advise on resulting return expectations. It is well-known that, while the strategies are mathematically reconcilable, they offer psychologically and typically practically different ways of exploring the formal decision problem.

We analyze and compare the latent threshold factor model with correlated factor innovations (*LTC*) and the standard model with independent factor innovations (*LT*), i.e., $A_t = I$. Table 2 reports cumulative returns of the portfolios resulting from the sequential investment over 100 business days, from several analyses that vary the degree of risk aversion tolerance κ_t , daily target returns r_t , and daily target risk tolerance ξ_t . We consider the chosen values as plausible levels for practical portfolio implementation selected based on statistics from the stock return data set. It is evident that the use of correlated factors remarkably dominates the uncorrelated factor model, across the range of portfolio allocation rules. The improvement of the forecasts is dramatic; based on the efficient frontier rule, the LTC model yields more than double the returns for $\kappa_t = 0.5$, and even close to quadruple for $\kappa_t = 5.0$; we see similarly significant results for the other two portfolio rules. In addition, we computed the sample variance of the realized returns; under the same level of variance, the LTC model tends to yield about double the realized returns compared to the LT model. We consider that these findings are natural because the correlations between factors reported above are considerably high throughout the sample period, therefore the structure of independent factor innovations loses much information that will otherwise impact on short-term

(1) Efficient Frontier		(2) Minimum Variance			(3) Maxi	(3) Maximum Return		
	LT	LTC		LT	LTC		LT	LTC
$\kappa_t = 0.5$	8.53	17.78	$r_t = 0.2$	15.25	20.14	$\xi_t = 1.58$	8.24	29.91
1.0	8.43	19.16	0.3	12.92	23.29	2.00	7.83	34.86
5.0	7.67	30.19	0.4	10.60	26.43	2.35	7.72	37.85

Table 2: Analysis of stock return data in a 3-factor model: Cumulative returns (%) for the latent threshold factor models with independent factor innovations (LT) and correlated factor innovations (LTC) based on portfolio implementation over 100 business days.



Figure 5: Analysis of stock return data in a 3-factor model: Cumulative returns (%) over 100 business days based on the efficient frontier portfolio for $\kappa_t = 1.0$.

forecasting. Figure 5 plots the trajectories of cumulative returns based on the efficient frontier portfolio algorithm of $\kappa_t = 1.0$ for each model, showing how the LTC model obviously dominates the LT model uniformly during the trial period of 100 business days. From parallel analyses, we have found that these results are robust across ranges of values of the key parameters $v_{\beta ij}$ and $v_{\alpha ij}$, while we stress the need for maintaining informative priors consistent with the discussion in Section 5.2. Readers interested in exploring these models may download the code provided, which allows experimentation and evaluation of the impacts of changes in prior specifications as well as other modeling choices.

7. A study of a 20-dimensional FX return series

This study concerns m = 20 series of daily foreign exchange (FX) rate returns. Analysis follows the same themes as in the previous section. Here, however, we focus more on dissecting the dynamic sparsity of the latent threshold mechanism, and evaluate and compare models under additional, practically very relevant portfolio strategies. As noted, part of this introduces a novel and practically important benchmark neutral portfolio construction, along with other more traditional constructions. We make comparisons with standard models across a range of portfolio studies, using both raw returns and risk-adjusted summaries, as well as in terms of predictive marginal likelihood measures.

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

Table 3: FX return data: 20 international currencies.

7.1. Data

We use 20 international currency exchange rates relative to US dollar, as listed in Table 3. The time period is T = 1,043 business days beginning in January 2008 and ending in December 2011. The returns are computed as the log difference of daily closing spot rates. For the number of factors and ordering of the variable in the y_t , our pre-analysis suggests k = 4 factors (using the same general ideas as in the previous section), and an appropriate choice of the first four series in y_t as (GBP, EUR, JPY, INR). Due to the triangular form of identification in the loadings matrix, the first four series associate with the factors; Factor1 measures overall FX movement across countries against US dollar led by the global proxy UK Pound; Factor2 essentially captures flows in European countries; Factor3 is led by JPY, which was regarded as a relatively less risky currency during the sample period and is also the natural primary proxy for measuring relative strength of the US dollar outside the EU; Factor4 is the INR-leading factor reflecting FX markets in emerging countries. For posterior computation, we use the same prior specifications as in the stock return analyses, with minor modifications on a few elements to adjust for scales, namely $1/v_{\delta i} \sim G(2, 0.01)$, and $\exp(-\mu_{*i}) \sim G(3, 0.03)$ for $* \in \{\delta, \lambda\}$.

7.2. Summaries of posterior inferences

Figure 6 shows estimates of posterior mean trajectories of the factors as well as the posterior estimates and credible intervals for trajectories of volatilities; the latter show both the underlying independent components of factor volatilities (the ψ_{it}) and the resulting factor innovation volatilities (the squareroots of diagonal elements of Υ_t). The factors show a high volatility period in late of 2008 tracing major turbulent fluctuations of the financial crisis triggered by the Lehman brothers shock. Focusing further around mid-2008, Factor1 (GBP-leading) and Factor2 (EUR-leading) imply seemingly correlated declines in the levels of the factors, which implies considerable appreciation of US dollar. During this period, Factor4 (INR-leading) exhibits almost negligible fluctuation. Another marked period is around the mid-2010 related to the European sovereign debt crisis; stochastic volatility component ψ_{1t} shows a sudden



Figure 6: Analysis of FX return data: Left column: posterior means of trajectories of factors f_{it} . Central column: posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of trajectories of factor stochastic volatilities, i.e, the square-roots of the diagonal elements of Υ_t over time t. Right column: Similar plots for factor volatilities $\psi_{it} = \exp(\lambda_{it}/2)$.



Figure 7: Analysis of FX return data: Posterior means and ± 1 standard deviation credible intervals (dotted) of trajectories of several correlations from Υ_t . Factors are labeled by leading currency.

hike, and all the factors including Factor4 exhibit higher fluctuations in their estimated trajectories. It is natural that the levels of marginal stochastic volatility are in decreasing order from Factor1 to Factor4 by construction and the identify of factors based on chosen order of returns series.

Figure 7 graphs estimated of trajectories of some of the factor correlations. That between GBP and EUR is substantially high in the beginning of the sample period and rises to levels very close to one. This is consistent with the role of Factor2 diminishing in the later stages of the sample period, where Factor1 essentially plays a more and more dominant role. For the other correlations, sudden change is observed around the financial crisis period, which implies a marked shift in factor relations across the currencies during these unprecedented initial events in world financial markets.

Figure 8 displays posterior estimates of the trajectories of the latent time-varying factor loadings, as well as the posterior *shrinkage probabilities* of $s_{it} = 0$ for selected loadings. Recall that $s_{it} = 0$ implies that the corresponding factor loading is fully shrunk to zero at that time, and so there is no effective relationship between the corresponding response series and factor for periods when $s_{it} = 0$. The panels evidently show both local and global shrinkage patterns in the factor loadings. The NZD-Factor1 loading takes relevant positive values in 2008 and 2009, while turns out to be irrelevant after the beginning of 2010; the posterior distribution of β_{it} declines toward zero and the corresponding shrinkage probability significantly rises up. Meanwhile, the other loadings for NZD are estimated to be relevant during the sample period. The NZD-Factor3 loading is estimated to be negative, which implies the NZD fluctuates inversely to the JPY-leading factor. This negative loading associated with the JPY-leading factor is typically observed across other several countries; the JPY exchange rate was regarded as a relatively less risky asset, along with CHF, during the recent crisis times.

For the RUB series, the loadings of Factor1 and Factor3 are basically shrunk to zero for entire range of the sample period, while the RUB-Factor2 loading has a significant role in describing fluctuations of RUB returns. The RUB-Factor4 loading shows an interesting behavior; the shrinkage probability is high around the financial crisis of 2008, while the loading turns out to be primarily positive with almost no shrinkage after 2009. In contrast, the BRL-Factor4 shows higher sparsity probabilities in the second half



Figure 8: Analysis of FX return data: Posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of trajectories of selected time-varying factor loadings. The dashed lines show posterior means of the corresponding latent thresholds. Posterior probabilities of $s_{it} = 0$ (implying full shrinkage to zero and hence a zero effective factor loading at those times) are plotted below each trajectory.

of the sample period, which indicates a shift in some relationships between the INR-related market and those series.

7.3. Multi-period, out-of-sample forecasting performance

We evaluate the contribution of the latent threshold mechanism and correlated factor components based on forecasting performance. Out-of-sample forecasts are obtained over 50 business days using the 20-dimensional FX times series data as in the previous section. First, we fit the model based on the first $T_1 = 1,201$ observations, $y_{1:T_1}$, and produce resulting out-of-sample predictive distribution over the following 5 business days $t = T_1 + 1, \ldots, T_1 + 5$. Then, the analysis moves ahead one business day to observe the next observation y_{T_2} ($T_2 = T_1 + 1$), and reruns the MCMC based on the updated data $y_{1:T_2}$, generating forecasts of the next 5 business days $t = T_2 + 1, \ldots, T_2 + 5$. This is repeated until we obtain 50 sets of daily 5-step ahead forecasts. We compare the LTC model to the NT model– the standard dynamic factor model with no thresholding and independent factor innovations, i.e., $A_t = I$.

Out-of-sample predictive fit can be explored with log predictive density ratios (LPDRs). For forecasting h days ahead from day t and comparing models M_1 and M_2 , this is

$$LPDR_t(h) = \log\{p_{M_1}(y_{t+h}|y_{1:t})/p_{M_0}(y_{t+h}|y_{1:t})\}$$

where $p_M(\mathbf{y}_{t+h}|\mathbf{y}_{1:t})$ is the predictive density under model M. Relative forecasting accuracy is represented by this evaluated at the observed data. In the context of Bayesian model comparison (e.g., West & Harrison, 1997, Chapters 10 & 12), cumulative sums of $\text{LPDR}_t(1)$, defining log model marginal likelihoods, formally evaluate evidence of the proposed model versus standard factor models. The $\text{LPDR}_t(h)$ for h > 1 provide further insights into relative forecasting ability at longer horizons.

Figure 9 shows the log predictive density ratio for LTC over NT model from out-of-sample forecasts over 50 business days, plotted for each horizon. The LPDR values are all positive, indicating higher predictive density for LTC model at each horizon, h = 1, ..., 5. The remarkably high values per quarter

imply strong support for LTC over NT. Cumulating values over the 50 days with h = 1, the log Bayes factor is 1969.0 for LTC/NT; this indicates most substantial formal evidence in favor of the LTC model.

7.4. Portfolio analysis

We discuss portfolio studies following similar themes to those of the stock index studies of the previous section, while now using more realistic and structured portfolio strategies extending the target return, minimum variance portfolio. We again fix the total sum of investment by restricting $w'_t 1 = 1$, where w_t here denotes a vector of portfolio weights allocated for 20 currencies.

Suppose that an investor is faced with a certain benchmark investment along with allocating resources to the 20 currencies. The benchmark can be any asset– a stock price index, oil price, or another currency excluding those analyzed. The investor deals with the portfolio of the currencies under two condition; (i) ex-ante forecast return attains some additional margin over the benchmark; and (ii) the implemented portfolio has zero correlation with the benchmark. We assume that the investor pursues a profit dominating the benchmark, and importantly, is unconcerned with fluctuations of the benchmark asset. This *decoupling* strategy more heavily emphasizes forecast accuracy of this higher-dimensional series than does the standard portfolio allocation; see works related to the idea of this decoupling strategy in financial markets (e.g., Gulko, 2002; Frauendorfer et al., 2007), and related studies of decoupling of liquid Treasury returns from equity returns during periods of crisis (Harper, 2003), for example. From the viewpoint of hedge-fund managers and investment advisors, decoupling strategies that emphasize (and can achieve) returns above a benchmark and that are uncorrelated with that benchmark are key and critical in defining a competitive market position.

We derive an explicit form for the optimized weights for the decoupling strategy. Define z_t as the return of the benchmark asset at time t, and write the mean and covariance structure of y_t and z_t jointly as follows:



 $E\begin{pmatrix} \mathbf{y}_t\\ z_t \end{pmatrix} = \begin{pmatrix} \mathbf{g}_t\\ x_t \end{pmatrix}, \quad V\begin{pmatrix} \mathbf{y}_t\\ z_t \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_t & \mathbf{q}_t\\ \mathbf{q}'_t & \gamma_t \end{pmatrix},$

Figure 9: Log predictive density ratios LPDR_t(h) for LTC model over the NT model at horizons h = 1 : 5 in analysis of FX returns.

with a $m \times 1$ vector of covariances $q_t = Cov(y_t, z_t)$ between asset returns and the benchmark under our "current" forecast distribution. At time t, we minimize the ex-ante portfolio variance $w'_t Q_t w_t$, subject to (i) $w'_t g_t = x_t + r_t$, (ii) $Cov(w'_t y_t, z_t) = 0$, and (iii) $w'_t 1 = 1$, where r_t is the expected margin gained over the forecast mean of the baseline asset. This quadratic optimization yields an explicit solution $w_t = K_t C'_t (C_t K_t C'_t)^{-1} a_t$, where $K_t = Q_t^{-1}$, $C_t = [g_t; q_t; 1]'$, and $a_t = (r_t, 0, 1)'$. We implicitly assume that there are no transaction costs to reallocate the resources to arbitrary long or short positions across the currencies, or that they are costs that will be similar, and negligible in their impact on resulting cumulative returns, across different models using the same decision strategy.

For the baseline asset, we use the S&P500 index from the US stock market. To obtain the covariance structure between the 20 currencies and the baseline asset, we reanalyze the factor models fit to the m = 21 vector $(z_t, y'_t)'$. We find that analysis indicates k = 5 factors and the ordering of the first five variables in y_t as (US-stock, GBP, EUR, JPY, INR), a natural extension of the earlier FX 4-factor model. Prior specifications and MCMC simulation details are as in the previous analysis and detailed in Section 7.1. In addition to the restricted condition, $w'_t 1 = 1$, we examine an unrestricted allocation that ignores this sum-to-one condition; that is, a strategy simulating an investor able to borrow or short to unrestricted levels with no cost (i.e., a major investment bank). The study uses a range of daily target margins of $r_t = 0.02\%$, 0.04%, and 0.06%, corresponding to monthly (about 20 business days) returns of approximately 0.4%, 0.8% and 1.2%, respectively. We regard these settings as well in the realm of realistic investor behavior. Beyond this, we have rerun the analyses using far more aggressive, and riskier, settings; they do indeed yield results that confirm– in fact more strongly– our general conclusions below on the uniform dominance of the LTC model over the NT model.

(1) Cumulative returns					
	NT	LTC			
(i) Restricted allocation					
Target margin: $r_t = 0.02$	-1.676	1.551			
0.04	-1.851	1.441			
0.06	-2.027	1.330			
(ii) Unrestricted allocation					
$r_t = 0.02$	-2.828	-0.110			
0.04	-2.974	-0.106			
0.06	-3.121	-0.101			
(2) Sharpe ratio					
	NT	LTC			
(i) Restricted allocation					
m 0.09	1 661				

(1) Cumulative noturn

-	NT	LTC			
(i) Restricted allocation					
$r_t = 0.02$	1.661	5.058			
0.04	1.437	5.066			
0.06	1.203	5.061			
(ii) Unrestricted allocation					
$r_t = 0.02$	0.118	4.104			
0.04	-0.125	4.276			
0.06	-0.371	4.442			

Table 4: Analysis of FX returns with benchmark: (1) Cumulative returns (%) and (2) Sharpe ratio for (i) restricted ($w'_t \mathbf{1} = 1$) and (ii) unrestricted allocations over 50 business days.

Table 4 reports realized cumulative returns of the portfolio allocation from the sequential investment over 50 business days based on daily out-of-sample one-day-ahead forecasts. The table shows Sharpe ratios (Sharpe, 1994), here defined as a cumulative sum of risk-adjusted portfolio returns divided by their standard deviation; the risk-adjusted portfolio return is the portfolio return minus an equally-weighted



Figure 10: Analysis of FX returns with benchmark: Cumulative returns (%) over 50 business days for target margin $r_t = 0.06\%$ with restricted (top) and unrestricted (bottom) allocations.



Figure 11: Analysis of FX returns with benchmark: (i) One-day returns from decoupling portfolio in LTC model (solid) and S&P500 baseline asset (dotted). (ii) Cumulative returns from LTC model with standard portfolio (dashed) and decoupling portfolio (solid). The target margin is $r_t = 0.06\%$.

portfolio return as a baseline portfolio. The use of correlated factor innovations and the latent threshold structure evidently and uniformly dominates the other models regardless of the target margins. Figure 10 shows the cumulative returns across time periods using a target margin of $r_t = 0.06\%$. The LTC model performs better than the NT model, which implies the correlated factor innovation and the latent threshold mechanism contribute to forecasting accuracy in the portfolio allocation experiment.

To assess the role of decoupling portfolio rules, Figure 11(i) plots return trajectories from portfolios

using the LTC model and the S&P500 index as the benchmark. The decoupling portfolio primarily realizes less volatile returns than the benchmark asset due to the offset strategy for correlation between them; the standard deviation of the returns is 1.60% for the decoupling portfolio and 0.54% for the baseline asset. This evidently provides a beneficially, more stable portfolio. We also compute portfolios based on the standard portfolio allocation, which excludes the constraint related to the benchmark asset (i.e., setting $x_t = 0$ and $q_t = 0$). Figure 11(ii) shows cumulative returns of the decoupling and standard portfolios from the LTC model. It is obvious that the decoupling portfolio yields better performance across the experimental period. The final resulting portfolio performance is returns of 1.33% for the decoupling portfolio and -0.09% for the standard. The benchmark asset is considerably volatile, while it leads to a higher target return imposing the decoupling portfolio for some times; this mechanism boosts the decoupling portfolio to the higher returns.

8. Concluding remarks

Our substantive examples in short-term forecasting and portfolio decisions for financial return time series illustrates a number of important aspects of dynamic latent factor modeling: in particular, we can identify naturally emerging patterns of remarkably correlated dynamics in factors, coupled with dynamic shrinkage in factor loadings. Dynamic sparsity patterns induced by the latent threshold structure underlie plausible parameter estimates for complicated flow and dynamics in the stock price index and FX returns, and improved short-term prediction accuracy in the realistic portfolio allocation analysis. From careful examination using various portfolio allocation rules reflecting practically relevant strategies, improvements in forecasting performance and resulting portfolio investments are evident, they are robust regardless of the portfolio allocation strategies, and they can be practically very substantial.

There are several methodological and computational areas for further investigation. Developing analysis techniques for sequential particle filtering and particle learning (e.g. Carvalho et al., 2010) is a possible future step towards a real-time, forward/sequential implementation of these models, while there are openings and opportunities for correlated dynamic factor models in areas of macroeconomic time series as well as high-dimensional microeconomic panel data, as well as multiple other areas of finance. We are also interested in computational implementations, and the need for fast, efficient and effective modular code to utilize for these models and potential future extensions. We note the current status on our code used for the current paper at the end of the following appendix.

Appendix: Bayesian computation

Bayesian computations via Markov chain Monte Carlo (MCMC) extend prior strategies in technically direct ways. The initial MCMC sampling scheme for latent threshold models (Nakajima & West, 2012a,b) is extended to this new class of correlated dynamic factor volatility models with additional MCMC components that, in terms of modern applications of Bayesian methods, are relatively routine to implement. We here summarize the several components of the MCMC computation to define posterior samples for all model hyperparameters and historical values of the trajectories of the several latent processes, based on observed data $y_{1:T}$.

First, each of the latent variable processes $c_{0:T}$, $f_{0:T}$, and $\alpha_{0:T}$ has a conditional posterior distributional form emerging from the theory of linear, Gaussian state space models. This enables us to directly utilize the standard forward filtering, backward sampling (FFBS) strategy for state space models (e.g. Prado & West, 2010). As often discussed in the literature, the FFBS method is an efficient algorithm that regenerates full trajectories of the latent variables over whole a data period t = 0 : T at each iterate of the overall MCMC. This is applied to each of the above component series, conditional on all other variables and latent processes, sequentially within each of the MCMC iterates.

The second component concerns the latent threshold factor loadings processes $b_{1:T}$ and the latent threshold parameters $d \equiv \{d_{ij}\}$. Conditional on thresholds d and all other quantities as well as the

data, we apply the extension of a Metropolis Hastings algorithm of Nakajima & West (2012a) to generate sample of β_{ijt} according to the observational eqn. (1) and the state eqn. (3). Note that, due to the diagonal structure of Σ_t , we can jointly sample β_{ijt} 's in each row of the loadings matrix. Write β_{jt} for the vector of β_{ijt} 's in the j^{th} row of the loadings matrix. The conditional posterior distribution of β_{jt} given $\beta_{j,-t} = \beta_{j,0:T} \setminus \beta_{j,t}$ and other parameters can then be assessed by the Metropolis-within-Gibbs sampling strategy with a proposal drawn from the underlying *non-threshold* model. The latter is a trivially computed normal distribution (see Nakajima & West, 2012a, Section 2.3). The procedure sequences through this sampling scheme over t = 0: T to obtain the entire sequence $\beta_{j,0:T}$, applied separately for j = 1: m. Sampling the latent threshold parameters in d from their conditional posteriors uses another simple Metropolis Hastings algorithm; the candidate is drawn from its uniform prior distribution developed above, as described in Nakajima & West (2012a).

Fourth, for the volatility sequences $\{\delta_{i,1:T}\}_{i=1:m}$ and $\{\lambda_{j,1:T}\}_{j=1:k}$, we apply the standard MCMC technique for univariate stochastic volatility models (Shephard & Pitt, 1997; Kim et al., 1998; Watanabe & Omori, 2004; Omori et al., 2007). We separately generate these process across i = 1:m and j = 1:k in parallel, developing efficient resampling of full trajectories of each volatility process over t = 1:T.

Fifth, for all AR parameters $\{\mu_*, \phi_*, v_*\}$, we generate samples from their conditional posterior distributions using Metropolis Hasting algorithms separately across *; see Appendix of Nakajima & West (2012a) for details.

The final component is the set of AR(1) coefficients $\gamma_{1:k}$ defining the diagonal factor VAR coefficient matrix G. Under the independent shifted-beta prior for each γ_i , we generate a sample from its conditional posterior distribution using a Metropolis Hastings algorithm. A truncated normal prior reduces the generation to a direct draw from the normal posterior distribution.

Software implementing these computations is freely available from the authors to interested readers. All the results reported here were produced in analysis using custom code in Ox (Doornik, 2006).

Acknowledgments

The authors thank the editor, associate editor and two anonymous referees for their constructive comments on an earlier version of the paper.

This work was completed while Jouchi Nakajima was a PhD student, and Xiaocong Zhou an undergraduate student, in Statistical Science at Duke University. Any opinions, findings and conclusions or recommendations expressed in this paper do not necessarily reflect the views of the UBS Investment Bank nor of the Bank of Japan.

This work was partially supported by 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 National Science Foundation.

References

- Aguilar, O., Prado, R., Huerta, G. & West, M., 1999. Bayesian inference on latent structure in time series (with discussion), in: Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, Oxford, pp. 3–26.
- Aguilar, O. & West, M., 2000. Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economic Statistics 18, 338–357.
- Aruoba, S.B., Diebold, F.X. & Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business & Economic Statistics 27, 417–427.
- Asai, M., McAleer, M. & Yu, J., 2006. Multivariate stochastic volatility: A review. Econometric Reviews 25, 145–175.

- Bai, J. & Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference for factoraugmented regressions. Econometrica 74, 1133–1150.
- Carvalho, C.M., Chang, J., Lucas, J.E., Nevins, J.R., Wang, Q. & West, M., 2008. High-dimensional sparse factor modeling: Applications in gene expression genomics. Journal of the American Statistical Association 103, 1438–1456.
- Carvalho, C.M., Johannes, M.S., Lopes, H.F. & Polson, N.G., 2010. Particle learning and smoothing. Statistical Science 25, 88–106.
- Carvalho, C.M., Lopes, H.F. & Aguilar, O., 2011. Dynamic stock selection strategies: A structured factor model framework (with discussion), in: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M. & West, M. (Eds.), Bayesian Statistics 9. Oxford University Press, pp. 69–90.
- Carvalho, C.M. & West, M., 2007. Dynamic matrix-variate graphical models. Bayesian Analysis 2, 69–98.
- Chib, S., Nardari, F. & Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. Journal of Econometrics 134, 341–371.
- Cogly, T. & Sargent, T.J., 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. Review of Economic Dynamics 8, 262–302.
- Del Negro, M. & Otrok, C., 2008. Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Federal Reserve Bank of New York Staff Reports, No.326.
- Doornik, J., 2006. Ox: Object Oriented Matrix Programming. Timberlake Consultants Press, London.
- Doz, C. & Renault, E., 2006. Factor stochastic volatility in mean models: A GMM approach. Econometric Reviews 25, 275–309.
- Fan, J., Fan, Y. & Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. Journal of Econometrics 147, 186–197.
- Forni, M. & Gambetti, L., 2010. The dynamic effects of monetary policy: A structural factor model approach. Journal of Monetary Economics 57, 203–216.
- Forni, M., Hallin, M., Lippi, M. & Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. Review of Economics and Statistics 82, 540–554.
- Frauendorfer, K., Jacoby, U. & Schwendener, A., 2007. Regime switching based portfolio selection for pension funds. Journal of Banking and Finance 31, 2265–2280.
- George, E.I., Sun, D. & Ni, S., 2008. Bayesian stochastic search for VAR model restrictions. Journal of Econometrics 142, 553–580.
- Geweke, J.F. & Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies 9, 557–587.
- Gulko, L., 2002. Decoupling. Journal of Portfolio Management 28, 59-66.
- Han, Y., 2005. Asset allocation with a high dimensional latent factor stochastic volatility model. Review of Financial Studies 19, 237–271.
- Harper, R.B., 2003. Asset allocation, decoupling, and the opportunity cost of cash. Journal of Portfolio Management 29, 25–35.

- Harvey, A.C., Ruiz, E. & Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.
- Jacquier, E., Polson, N.G. & Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.
- Kim, S., Shephard, N. & Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.
- Koop, G. & Potter, S.M., 2004. Forecasting in dynamic factor models using Bayesian model averaging. Econometrics Journal 7, 550–565.
- Korobilis, D., 2012. VAR forecasting using Bayesian variable selection. Journal of Applied Econometrics, forthcoming.
- Lopes, H.F. & Carvalho, C.M., 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference 137, 3082–3091.
- Lopes, H.F., McCulloch, R.E. & Tsay, R., 2010. Cholesky stochastic volatility. Technical Report. University of Chicago, Booth Business School.
- Lopes, H.F. & West, M., 2004. Bayesian model assessment in factor analysis. Statistica Sinica 14, 41-67.
- Markowitz, H., 1959. Portfolio Selection: Efficient Diversification of Investments. John Wiley and Sons, New York.
- Nakajima, J. & West, M., 2012a. Bayesian analysis of latent threshold dynamic models. Journal of Business and Economic Statistics doi: 10.1080/07350015.2012.747847.
- Nakajima, J. & West, M., 2012b. Dynamic factor volatility modeling: A Bayesian latent threshold approach. Journal of Financial Econometrics doi: 10.1093/jjfinec/nbs013.
- Omori, Y., Chib, S., Shephard, N. & Nakajima, J., 2007. Stochastic volatility with leverage: Fast likelihood inference. Journal of Econometrics 140, 425–449.
- Philipov, A. & Glickman, M.E., 2006. Factor multivariate stochastic volatility via Wishart processes,. Econometric Reviews 25, 311–334.
- Pinheiro, J.C. & Bates, D.M., 1996. Unconstrained parametrizations for variance-covariance matrices. Statistics and Computing 6, 289–296.
- Pitt, M. & Shephard, N., 1999. Time varying covariances: A factor stochastic volatility approach (with discussion), in: Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M. (Eds.), Bayesian Statistics VI. Oxford University Press, pp. 547–570.
- Prado, R. & West, M., 2010. Time Series Modeling, Computation, and Inference. Chapman & Hall/CRC, New York.
- Primiceri, G.E., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72, 821–852.
- Putnum, B.H. & Quintana, J.M., 1994. New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination, in: Proceedings of the ASA Section on Bayesian Statistical Science, American Statistical Association. pp. 232–237.
- Quintana, J.M., 1992. Optimal portfolios and forward currency contracts, in: Berger, J.O., Bernardo, J.M., Dawid, A.P. & Smith, A.F.M. (Eds.), Bayesian Statistics IV. Oxford University Press, pp. 753–762.

- Quintana, J.M., Carvalho, C.M., Scott, J. & Costigliola, T., 2010. Futures markets, Bayesian forecasting and risk modeling, in: O'Hagan, A. & West, M. (Eds.), The Handbook of Applied Bayesian Analysis. Oxford University Press, Oxford, pp. 343–365.
- Quintana, J.M., Lourdes, V., Aguilar, O. & Liu, J., 2003. Global gambling, in: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M. & West, M. (Eds.), Bayesian Statistics VII. Oxford University Press, pp. 349–368.
- Quintana, J.M. & Putnum, B.H., 1996. Debating currency markets efficiency using dynamic multiplefactor models, in: Proceedings of the ASA Section on Bayesian Statistical Science, American Statistical Association. pp. 55–60.
- Sharpe, W.F., 1994. The Sharpe ratio. Journal of Portfolio Management 21, 49–58.
- Shephard, N. & Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.
- Smith, M. & Kohn, R., 2002. Parsimonious covariance matrix estimation for longitudinal data. Journal of the American Statistical Association 97, 1141–1153.
- Stock, J.H. & Watson, M.W., 1989. New indexes of coincident and leading economic indicators, in: NBER Macroeconomics Annual. MIT Press, Cambridge, MA. volume 4, pp. 351–409.
- Stock, J.H. & Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, 20, 147–162.
- Wang, H., 2010. Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. Computational Statistics and Data Analysis 54, 2866–2877.
- Wang, H. & West, M., 2009. Bayesian analysis of matrix normal graphical models. Biometrika 96, 821– 834.
- Watanabe, T. & Omori, Y., 2004. A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). Biometrika 91, 246–248.
- West, M., 2003. Bayesian factor regression models in the "large p, small n" paradigm, in: Bernardo, J., Bayarri, M., Berger, J., David, A., Heckerman, D., Smith, A. & West, M. (Eds.), Bayesian Statistics 7, Oxford. pp. 723–732.
- West, M. & Harrison, P.J., 1997. Bayesian Forecasting and Dynamic Models. Springer-Verlag, New York. 2nd edition.
- Yoshida, R. & West, M., 2010. Bayesian learning in sparse graphical factor models via annealed entropy. Journal of Machine Learning Research 11, 1771–1798.
- Yu, J. & Meyer, R., 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison. Econometric Reviews 25, 361–384.