Bayesian Analysis of Latent Threshold Models

by

Jochi Nakajima

Department of Statistical Science
Duke University

Date: _____

Approved:

Mike West, Supervisor

Surya Tapas Tokdar

Fan Li

Francesco Bianchi

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University

2012

ABSTRACT

Bayesian Analysis of Latent Threshold Models

by

Jochi Nakajima

Department of Statistical Science
Duke University

Date: _____

Approved:

Mike West, Supervisor

Surya Tapas Tokdar

Fan Li

Francesco Bianchi

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University

2012

Abstract

Time series modeling faces increasingly high-dimensional problems in many scientific areas. Lack of relevant, data-based constraints typically leads to increased uncertainty in estimation and degradation of predictive performance. This dissertation addresses these general questions with a new and broadly applicable idea based on latent threshold models. The latent threshold approach is a model-based framework for inducing data-driven shrinkage of elements of parameter processes, collapsing them fully to zero when redundant or irrelevant while allowing for time-varying non-zero values when supported by the data. This dynamic sparsity modeling technique is implemented in broad classes of multivariate time series models with application to various time series data. The analyses demonstrate the utility of the latent threshold idea in reducing estimation uncertainty and improving predictions as well as model interpretation.

Chapter 1 overviews the idea of the latent threshold approach and outlines the dissertation. Chapter 2 introduces the new approach to dynamic sparsity using latent threshold modeling and also discusses Bayesian analysis and computation for model fitting. Chapter 3 describes latent threshold multivariate models for a wide range of applications in the real data analysis that follows. Chapter 4 provides US and Japanese macroeconomic data analysis using latent threshold VAR models. Chapter 5 analyzes time series of foreign currency exchange rates (FX) using latent threshold dynamic factor models. Chapter 6 provides a study of electroencephalographic

(EEG) time series using latent threshold factor process models. Chapter 7 develops a new framework of dynamic network modeling for multivariate time series using the latent threshold approach. Finally, Chapter 8 concludes the dissertation with open questions and future works.

To Hero Nakajima

Contents

Abstract	iv
List of Tables	xii
List of Figures	xiii
List of Abbreviations and Symbols	xvii
Acknowledgements	xviii
1 Introduction	1
1.1 A new idea of dynamic sparsity modeling	1
1.2 Outline	3
2 Latent Threshold Modeling	6
2.1 Basic framework	6
2.2 Latent thresholds	8
2.2.1 Concept	8
2.2.2 Priors	10
2.2.3 Shrinkage probability	10
2.3 Bayesian computation	14
2.4 Simulation study	17
2.5 Generalization	20
2.5.1 Threshold and shrinkage indicators	21

2.5.2	Nonstationary latent process	22
2.5.3	General covariance matrix for latent process	22
3	Latent Threshold Multivariate Models	24
3.1	Latent threshold time-varying VAR (LT-VAR) models	24
3.1.1	Model structure	25
3.1.2	Time-varying covariance matrix	26
3.2	Latent threshold time-varying covariance matrix	29
3.3	Latent threshold factor (LTF) models	31
4	Application to Macroeconomic Time Series	36
4.1	Context and literature	36
4.2	US macroeconomic data analysis	37
4.2.1	Data and priors	37
4.2.2	Forecasting performance and comparisons	39
4.2.3	Summaries of posterior inferences	46
4.3	Japanese macroeconomic data analysis	50
4.3.1	Data	50
4.3.2	Forecasting performance and comparisons	52
4.3.3	Summaries of posterior inferences	53
4.4	Summary	57
5	Application to Financial Time Series	60
5.1	Context and literature	60
5.2	Identification	62
5.3	Exchange rate data analysis – Major currencies	63
5.3.1	Data and priors	63

5.3.2	Summaries of posterior inferences	66
5.3.3	Portfolio analysis	72
5.3.4	Diagnostics	80
5.4	Exchange rate and commodity price data analysis	82
5.4.1	Data and priors	82
5.4.2	Posterior inferences and portfolio analysis	83
5.5	Exchange rate and stock price data analysis	89
5.5.1	Data and priors	89
5.5.2	Posterior inferences and portfolio analysis	89
5.6	Exchange rate data analysis – World-wide currencies	94
5.6.1	Data and priors	94
5.6.2	Summaries of posterior inferences	95
5.6.3	Portfolio analysis	99
5.7	Summary	103
6	Application to EEG Time Series	104
6.1	Context and literature	104
6.2	Latent threshold factor process (LTFP) model	106
6.3	EEG data analysis	108
6.3.1	Data and priors	108
6.3.2	Summaries of posterior inferences	110
6.3.3	An extension and impulse response analysis	119
7	Dynamic Network Modeling	123
7.1	Context and literature	123
7.2	LT-VAR model revisited	125
7.3	Exchange rate data analysis	127

7.3.1	Data and priors	128
7.3.2	Summaries of posterior inferences	129
7.4	EEG data analysis	134
7.4.1	Data and priors	134
7.4.2	Summaries of posterior inferences	135
8	Concluding Remarks	141
A	Posterior Computation and MCMC Algorithm	144
A.1	MCMC algorithm for LTMs	144
A.2	MCMC algorithm for LT-VAR models	147
A.3	MCMC algorithm for LTF models	149
A.4	Software	151
B	Cholesky-type Time-varying Covariance Matrix	152
	Bibliography	156
	Biography	164

List of Tables

2.1	Shrinkage probability $\Pr(s_t = 0)$ marginalized on $(\boldsymbol{\theta}_i, d_i)$	14
2.2	Simulation example: Posterior estimates.	17
4.1	Forecasting performance for US macroeconomic data over the 16 years (1996/Q4–2011/Q4).	40
4.2	Forecasting performance for US macroeconomic data over the 5 years (1996/Q4–2001/Q3).	42
4.3	Forecasting performance for Japanese macroeconomic data over the 5 years (2003/Q1–2007/Q4).	52
5.1	6 international major currencies for FX data.	64
5.2	Cumulative returns over 100 business days for major currency data. .	75
5.3	Portfolio risk ratio: Ex-ante ratio of portfolio standard deviation under the NT models relative to the LT models.	76
5.4	Additional international currencies for FX data.	83
5.5	20 international currencies for FX and 20 countries for stock price index.	90
5.6	Cumulative returns and average transaction over 50 business days for FX and stock data.	93
5.7	World-wide 20 international currencies.	94
5.8	Cumulative returns over 50 business days for world-wide FX data. . .	102
7.1	FX return data: 20 international currencies, ordered by the higher turnover in the global foreign exchange market.	128
7.2	19 EEG channels: (i) Baseline and (ii) factor-based ordering.	134

List of Figures

2.1	Example of LTM concept.	7
2.2	Likelihood $p(y_t d_1)$	9
2.3	Shrinkage probabilities given (w_i, d_i) for different ϕ_i 's.	12
2.4	Shrinkage probabilities given (ϕ_i, w_i) for different d_i 's.	13
2.5	Simulation example: Posterior estimates.	18
2.6	Simulation example: Priors and posteriors for thresholds.	18
4.1	US macroeconomic time series.	38
4.2	Posteriors of RMSFE from MCMC analysis of US macroeconomic data.	43
4.3	Log predictive density for LT-VAR and NT-VAR models.	45
4.4	Log predictive density ratios for LT-VAR over NT-VAR model.	45
4.5	Posterior probabilities of $s_{it} = 0$	47
4.6	Posterior trajectories of $a_{ij,t}$	47
4.7	Posterior trajectories of h_{it} and $\exp(h_{it}/2)$	48
4.8	Impulse response trajectories from NT-VAR and LT-VAR models.	49
4.9	Japanese macroeconomic time series.	51
4.10	Posterior means of β_t for Japanese macroeconomic data.	54
4.11	Posterior trajectories of $a_{ij,t}$	55
4.12	Posterior trajectories of h_{it} and $\exp(h_{it}/2)$	55
4.13	Impulse response trajectories from the NT-VAR and LT-VAR models.	56

4.14	Impulse response trajectories with credible intervals from the NT-VAR and LT-VAR models.	58
5.1	Daily exchange rates and returns for major currency data.	65
5.2	Estimated trajectories of univariate stochastic volatility fit to each series separately.	66
5.3	Posterior means and 95% credible intervals of time-varying factor loadings.	67
5.4	Time series trajectories of proportion of variances explained by factors.	69
5.5	Posterior estimates of factors and volatility.	70
5.6	Posterior means of series-specific stochastic volatility.	71
5.7	Results of FX analysis in model with $q = 4$ factors. Posterior estimates of factors.	72
5.8	Results of FX analysis in model with $q = 4$ factors. Posterior estimates of time-varying factor loadings.	73
5.9	Cumulative returns over 100 business days based on one-step ahead forecasting.	76
5.10	Portfolio risk: Ex-ante portfolio standard deviation based on one-step ahead forecasting.	77
5.11	Trajectories of portfolio weights based on one-step ahead forecasts.	78
5.12	Trajectories of portfolio weights based on five-step ahead forecasts.	79
5.13	QQ-plots for standardized errors against the standard normal.	81
5.14	Box-plots for posterior distributions of p -values in the normality tests using full posterior samples of standardised errors.	82
5.15	Posterior probability of $s_{it} = 0$ associated with the loadings for exchange rate and commodity data.	84
5.16	Posterior means of factors and log of commodity prices.	85
5.17	Posterior means and 95% credible intervals of time-varying regression coefficients.	86

5.18	Cumulative returns over 25 business days based on one-step ahead forecasting.	88
5.19	Trajectories of portfolio weights based on the portfolio.	88
5.20	Posterior estimates of factors and volatility for FX and stock price data.	91
5.21	Posterior probability of $s_{it} = 0$ associated with selected loadings. . . .	92
5.22	Heat map of posterior probabilities of $s_{it} = 0$	92
5.23	Posterior estimates of factors and volatility for world-wide FX data. .	96
5.24	Posterior estimates of correlations from $\mathbf{W}_t$	97
5.25	Posterior estimates of selected time-varying factor loadings.	98
5.26	Cumulative returns over 50 business days.	102
5.27	Analysis of world-wide FX returns with benchmark: Decoupling portfolio and standard portfolio.	103
6.1	Representation of the approximate 19-electrodes placement over the scalp.	109
6.2	Estimated latent factor process for EEG data.	111
6.3	Trajectories of posterior means and 95% credible intervals for the TV-AR coefficients.	112
6.4	Trajectories of posterior probability of $s_{it} = 1$ for the channel F7 . . .	113
6.5	Posterior probability of $s_{it} = 1$ with the lag-lead structure at selected time points.	116
6.6	Estimated factor loadings with the lag-lead structure at selected time points.	117
6.7	Posterior means of time-varying standard deviation for idiosyncratic shocks.	118
6.8	Impulse responses of 19 EEG channels to underlying factor process shock.	120
6.9	Impulse responses of 19 EEG channels to underlying factor process shock: Heat map.	122
7.1	Results of FX network analysis: Posterior probabilities of $s_{ijt} = 1$. . .	130

7.2	Evolution of networks: Induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$ for select time points.	131
7.3	Posterior means and ± 2 standard deviation credible intervals of a_{ijt} . .	132
7.4	Posterior probabilities of $s_{ijt} = 1$ for induced dynamic network in $\mathbf{\Omega}_t$ and $\mathbf{B}_{1t}$ from baseline ordering and reverse ordering.	133
7.5	Results of EEG network analysis: Posterior probabilities of $s_{ijt} = 1$. .	136
7.6	Posterior probabilities of $s_{ijt} = 1$ for the induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$	137
7.7	Induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$ and selected network maps for select time points.	139
7.8	Posterior probabilities of $s_{ijt} = 1$ for induced dynamic network in $\mathbf{\Omega}_t$ from baseline ordering and factor-based ordering.	140
B.1	Trajectories of simulated $\mathbf{a}_t$ and $\mathbf{h}_t$, as well as resulting elements in the covariance matrix.	154
B.2	Autocorrelation function of resulting elements in the covariance matrix.	154

List of Abbreviations and Symbols

Symbols

$d \sim U(a, b)$	Uniform distribution.
$p \sim B(a, b)$	Beta distribution.
$v \sim G(a, b)$	Gamma distribution.
$\mathbf{V} \sim W(c, \mathbf{D})$	Wishart distribution.
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker product.
$\text{diag}(\cdot)$	Diagonal matrix.
$\mathbf{x} \circ \mathbf{y}$	Element-wise product.
$s : t$	Indices $s, s + 1, \dots, t$.

Abbreviations

DLM	Dynamic linear model.
LT	Latent threshold.
LTF	Latent threshold factor (model).
LTM	Latent threshold model.
NT	Non-threshold (model).
TV	Time-varying (parameter/model).
VAR	Vector autoregressive (model).

Acknowledgements

I would like to thank my advisor, Mike West, for his insightful guidance, invaluable advice, and warm encouragement throughout my years at DSS. I also wish to thank my committee members, Surya Tokdar, Fan Li, Susie Bayarri, and Francesco Bianchi for their suggestions. I also thank my friends and colleagues who have supported me during my time at Duke.

I would like to thank Yasuhiro Omori for providing me guidance and support early in my graduate career. I thank Toshiaki Watanabe and Genshiro Kitagawa for their support and great encouragement to my research. Further thanks go to Toshitaka Sekine, Shigenori Shiratsuka, Masazumi Hattori, Shigeru Iwata, and staff of the Bank of Japan for their constant support. Special thanks go to Mikio Fukase and Kiyoo Suzuki for teaching me mathematics and introducing such a wonderful world of statistics in my high school days.

Finally, I express my eternal gratitude to my parents, Hirofumi (Hero) and Junko Nakajima, and my sister, Tomoko Nakajima. I am grateful for their unconditional support and love. Since I came to Duke, Hero has been a great “Duke Watcher,” following my works and inspiring me with thoughtful discussions. I could not have done this work without him. To you, Hero, I dedicate this dissertation.

Introduction

1.1 A new idea of dynamic sparsity modeling

Time series modeling faces increasingly high-dimensional problems in many scientific areas. For analysis of such time series data, dynamic modeling strategies are pressed by the need to appropriately constrain parameters and time-varying parameter processes. Lack of relevant, data-based constraints typically leads to increased uncertainty in estimation and degradation of predictive performance. This dissertation addresses these general questions with a new and broadly applicable idea based on latent threshold models (LTMs). The LTM approach is a model-based framework for inducing data-driven shrinkage of elements of parameter processes, collapsing them fully to zero when redundant or irrelevant while allowing for time-varying non-zero values when supported by the data. This dynamic sparsity modeling concept is implemented in broad classes of multivariate time series models with great potential to reduce estimation uncertainty and improve predictions as well as model interpretation.

Much progress has been made in recent years in the general area of Bayesian

sparsity modeling: developing model structures via hierarchical priors that are able to induce shrinkage to zero of subsets of parameters. The standard use of sparsity priors for regression model uncertainty and variable selection (George and McCulloch 1993, 1997; Clyde and George 2004) has become widespread in areas including sparse factor analysis (West 2003; Carvalho et al. 2008; Yoshida and West 2010), graphical modeling (Jones et al. 2005), and traditional time series models (e.g., George et al. 2008; Chen et al. 2011). In the context of time series analysis, these general strategies have been usefully applied to induce “global” shrinkage to zero of parameter subsets, zeroing out regression coefficients in a time series model for all time (Carvalho and West 2007; George et al. 2008; Korobilis 2012; Wang 2010). The LTM mechanism builds on this idea but addresses the much more general question of time-varying inclusion of effects, i.e., dynamic sparsity modeling. The LTM strategy provides flexible, “local” and adaptive (in time) data-dependent variable selection for dynamic regressions and autoregressions, and for volatility matrices as components of more elaborate models.

In this dissertation, the LTM ideas are first introduced in dynamic regression models. This context also illustrates how the LTM idea applies to richer classes of multivariate dynamic models (West and Harrison 1997; Prado and West 2010; West 2013). The methodology is then incorporated into time-varying vector autoregressive (TV-VAR) models (Primiceri 2005; Prado and West 2010), time-varying covariance matrix processes (Quintana and West 1987; Bollerslev et al. 1992; Aguilar and West 2000; Shephard 2005; Carvalho and West 2007), and dynamic latent factor volatility models (Aguilar et al. 1999; Pitt and Shephard 1999; Aguilar and West 2000; Lopes and Carvalho 2007). Applications in time series analysis in macroeconomics, financial econometrics, and a psychophysiological study illustrate the approach, practical use, and impact of the LTM structure as well as highlighting connections with other

approaches, open questions and potential extensions. Dynamic shrinkage, which allows parameter processes to take non-zero, time-varying values over some periods of time but shrink to zero completely in others, is a natural extension of traditional Bayesian sparsity modeling to the time series context and can engender these benefits adaptively over time. Data induced dynamic sparsity feeds through to improvements in forecasting performance and contextually reasonable shrinkage of inferred state processes.

1.2 Outline

This dissertation consists of two parts. The first part describes a definition of the LTMs, discusses the approach, and develops the structured multivariate time series models, exploiting the LTM mechanism. Chapter 2 introduces the new approach to dynamic sparsity using latent threshold modeling and also discusses Bayesian analysis and computation for model fitting. Chapter 3 describes three latent threshold multivariate models useful for a wide range of applications in real data analysis. The first multivariate model is based on the TV-VAR models; each time-varying coefficient of lagged response is allowed to follow the LTM process, referred to as LT-VAR models. The second model is a time-varying covariance matrix where each off-diagonal covariance component is assumed to follow the LTM process. This LT-based covariance matrix is typically coupled with the LT-VAR model in real data analysis. The third multivariate model is a dynamic factor model where factor loadings follow the LTM process, labeled as LTF models.

The second part of the dissertation applies these multivariate LT models to multivariate time series data. Chapter 4 provides macroeconomic time series examples using the LT-VAR models. The analysis uses topical US and Japanese macroeconomic data to examine the forecasting performance of the LTM mechanism and

implications from estimated latent processes. Chapter 5 considers time series of foreign currency exchange rates (FX) using the LTF models with portfolio allocation studies. Chapter 6 provides a study of electroencephalographic (EEG) traces, time series of electrical potential fluctuations at various scalp locations of a human subject, reflecting dynamics of brain connectivities during a brain seizure. The analysis uses the LTF model with an additional factor process structure for unique characteristics of the EEG traces. Chapter 7 describes a new framework of dynamic network modeling for multivariate time series using the LTM approach. The analysis fits the LT-VAR models for time series of the FX returns and the EEG traces, and discusses dynamic networks induced by time-varying sparsity in the LTM structure. Chapter 8 concludes the dissertation with open questions and future works. Finally, the Appendices provide details of MCMC algorithms for the models proposed and discuss characteristics of the time-varying covariance matrices used in this dissertation.

To relate this dissertation to research articles, I reference the following papers:

- Nakajima, J. and West, M. (2012a), “Bayesian analysis of latent threshold dynamic models,” *Journal of Business and Economic Statistics*, forthcoming.
- — (2012b), “Dynamic factor volatility modeling: A Bayesian latent threshold approach,” *Journal of Financial Econometrics*, doi:10.1093/jjfinec/nbs013.
- — (2012c), “Evaluating EEG dynamics: An approach using latent threshold factor modelling,” Manuscript, Department of Statistical Science, Duke University.
- — (2012d), “Latent threshold models: Multivariate time series and dynamic networks,” Manuscript, Department of Statistical Science, Duke University.
- Zhou, X., Nakajima, J., and West, M. (2012), “Dynamic dependent factor

models: Improving forecasts and portfolio decisions in financial time series,”
Working Paper 12-09, Department of Statistical Science, Duke University.

Nakajima and West (2012a) describe the new idea of LTMs (Chapter 2 in this dissertation), develop the LT-VAR models and LT-based time-varying covariance matrix (Chapter 3), and the macroeconomic analyses (Chapter 4). Nakajima and West (2012b) develop the LTF models (Chapter 3) and provide their application to FX time series (Chapter 5). Zhou et al. (2012) consider correlated factors in the LTF models and provide analyses of stock price index and FX time series (Chapter 5). Nakajima and West (2012c) illustrate the LTF model with a TV-AR factor process with application to the EEG traces (Chapter 6). Nakajima and West (2012d) introduce the new framework of dynamic network modeling using the LTM mechanism and provide analyses of the FX and EEG time series (Chapter 7).

2

Latent Threshold Modeling

2.1 Basic framework

The idea of the LTM is first introduced in simple dynamic regression models (e.g., West and Harrison 1997, Chapters 2 & 4), a subclass of dynamic linear models (DLMs). Suppose that y_t is a univariate response; $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})'$ is a $k \times 1$ vector of predictors; and $\mathbf{b}_t = (b_{1t}, \dots, b_{kt})'$ is a $k \times 1$ vector of time-varying coefficients. A basic form of the LTM is defined by

$$y_t = \mathbf{x}_t' \mathbf{b}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad (2.1)$$

$$b_{it} = \beta_{it} s_{it}, \quad (2.2)$$

$$s_{it} = I(|\beta_{it}| \geq d_i), \quad i = 1, \dots, k, \quad (2.3)$$

for $t = 1, \dots, n$, where $\mathbf{d} \equiv (d_1, \dots, d_k)$ is a vector of “latent thresholds” with $d_i \geq 0$, for $i = 1, \dots, k$, and $I(\cdot)$ denotes an indicator function. The time-varying coefficients, $\mathbf{b}_t$, are governed by latent time-varying parameters, $\boldsymbol{\beta}_t \equiv (\beta_{1t}, \dots, \beta_{kt})'$, and indicator variables, $\mathbf{s}_t \equiv (s_{1t}, \dots, s_{kt})'$. The fundamental idea is simple and clear; the i -th variable x_{it} has the time-varying coefficient whose value is shrunk to zero

FIGURE 2.1: Example of LTM concept. The dynamic regression coefficient process β_{1t} arises as a thresholded version of an underlying dynamic coefficient time series.

when its absolute value falls below the threshold, as illustrated in Figure 2.1. The x_{it} plays a role in predicting y_t in the observation equation (2.1) only when the absolute value of β_{it} is large enough, which leads to temporal variable selection. A key is that the relevance of variables can be dynamic; x_{it} may have a non-zero (time-varying) coefficient in some time periods but exactly zero in others, depending on the data and context. The model structure is thus flexible in addressing dynamic regression model uncertainty, i.e., selecting potential predictor variables at each time point.

Any form of model can be defined for β_t ; throughout this dissertation, the following vector autoregressive (VAR) process is assumed:

$$\beta_{t+1} = \mu + \Phi(\beta_t - \mu) + \eta_t, \quad \eta_t \sim N(\mathbf{0}, \mathbf{V}), \quad (2.4)$$

with individual AR parameters ϕ_i in the $k \times k$ matrix $\Phi = \text{diag}(\phi_1, \dots, \phi_k)$ and innovation variance matrix $\mathbf{V} = \text{diag}(v_1^2, \dots, v_k^2)$. Assuming $|\phi_i| < 1$ implies an individual

stationary AR(1) process with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$ and marginal distribution:

$$\beta_{it} \sim N(\mu_i, w_i^2), \quad w_i^2 = v_i^2 / (1 - \phi_i^2), \quad (2.5)$$

for $i = 1, \dots, k$.

The model of eqns.(2.1)-(2.4) is referred to as the “Latent Threshold Model” (LTM). Note that if $d_i \equiv 0$, the LTM reduces to a special case of the DLM. This LTM structure applies in a wide range of classes of dynamic time series models as developed in the following chapters. Connections with previous work include threshold mechanisms in regime switching and tobit/probit models (e.g., West and Mortera 1987; Polasek and Krause 1994; Wei 1999; Galvao and Marcellino 2010), as well as mixtures and Markov-switching models (Chen et al. 1997; West and Harrison 1997; Kim and Nelson 1999; Kim and Cheon 2010; Prado and West 2010) that describe discontinuous shifts in dynamic parameter processes. The LTM approach differs fundamentally from these approaches in that threshold mechanisms operate continuously in the parameter space and over time. This leads to a new paradigm and practical strategies for dynamic threshold modeling in time series analysis, with broad potential application.

2.2 Latent thresholds

2.2.1 Concept

To highlight the role of the threshold, suppose that $k = 1$ and $\mu_1 = 0$ as a simple case, and set $\sigma = 1$ with no loss of generality. At each time t , marginalizing over β_{1t} under eqn.(2.5) yields

$$p(y_t | d_1) = \pi_1 g(y_t) + (1 - \pi_1) f_N(y_t | 0, 1), \quad (2.6)$$

FIGURE 2.2: Likelihood $p(y_t|d_1)$ with $\mu_1 = 0$, $\phi_1 = 0.99$, $v_1 = 0.8$, and $\sigma = 1$. The mixing weight π_1 is reported for each d_1 in parentheses.

with

$$\begin{aligned}\pi_1 &= \Pr(|\beta_{1t}| \geq d_1) = 2\Phi(-d_1/w_1), \\ g(y_t) &= \frac{h(y_t)}{\pi_1} f_N(y_t|0, 1 + v_1^2), \\ h(y_t) &= 1 - \Phi\left(\frac{d_1 - a^2 y_t}{a}\right) + \Phi\left(\frac{-d_1 - a^2 y_t}{a}\right),\end{aligned}$$

where $a^2 = v_1^2/(1 + v_1^2)$, and $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal. The continuous mixture density (2.6) smoothly transitions from the zero-mean normal component when d_1 is large, to that from the regression with the normal prior on β_{1t} when d_1 is small or zero. This is illustrated in Figure 2.2. The mixing weight π_1 involves the latent threshold d_1 and the marginal variance w_1^2 of the underlying process. This also indicates the importance of the scale w_1 in considering relevant values of the threshold d_1 .

2.2.2 Priors

Conditional on the hyper-parameters $\boldsymbol{\theta}_i \equiv (\mu_i, \phi_i, v_i)$ underlying the stationary distribution of eqn.(2.5), a general form of prior for the latent threshold is specified by

$$d_i | \boldsymbol{\theta}_i \sim U(0, |\mu_i| + K_i w_i), \quad K_i \sim p(K_i). \quad (2.7)$$

The threshold d_i follows the uniform prior that respects the range of the underlying latent process β_{it} as reflected in its marginal distribution. The hyper-parameter K_i defines how large thresholds may be and is assumed exchangeably distributed with a specified common pdf $p(K_i)$. As shown later, simulation experiments conclude that taking K_i fixed is typically adequate, and setting $K_i = 3$ is broad enough to span the underlying range of variation of the state process in its marginal distribution. Note that the full model specification will include priors on the LTM parameters $\boldsymbol{\theta}_i$, so inducing marginal priors for thresholds as

$$p(d_i | K_i) = \int f_U(d_i | 0, |\mu_i| + K_i w_i) p(\mu_i, w_i) d\mu_i dw_i. \quad (2.8)$$

These prior forms are revisited in a simulation study below (Section 2.4).

2.2.3 Shrinkage probability

Latent thresholding is a direct extension of standard Bayesian variable selection (e.g., George and McCulloch 1993, 1997; Clyde and George 2004; West 2003; Carvalho et al. 2008) to the time series context, providing a general scheme of dynamic regression model uncertainty. Standard Bayesian model selection methods assign a non-zero prior probability to zero values of regression parameters, and continuous priors centered at zero otherwise. The extension to time-varying parameters requires a non-zero marginal prior probability $\Pr(b_{it} = 0)$ coupled with a continu-

ous prior on non-zero values, but that also respects the time series context and the need to induce dependencies in the b_{it} being zero/non-zero over time. A highly positively dependent AR model for β_{it} respects relatively smooth variation over time in the parameters, while the thresholding mechanism induces persistence over periods of time in the occurrences of zero/non-zero values in the effective coefficients b_{it} 's. The threshold parameter defines the marginal probability of a zero coefficient at any time and implicitly the persistence in terms of joint probabilities over sequences of consecutive zeros/non-zeros. In particular, it is useful to understand the role of the threshold d_i in defining the marginal probability $\Pr(b_{it} = 0)$, which is the key, sparsity prior parameter analogous to a prior variable exclusion probability in standard Bayesian variable selection for regression models.

A simulation investigates how the underlying process β_{it} and latent thresholds d_i relate to shrinkage patterns arising on the coefficient b_{it} . First, consider key probabilities $p_1 = \Pr(s_{it} = 1)$, $p_{00} = \Pr(s_{it} = 0 | s_{i,t-1} = 0)$, and $p_{11} = \Pr(s_{it} = 1 | s_{i,t-1} = 1)$ for several values of (ϕ_i, w_i, d_i) ; $\mu_i = 0.5$, fixed, with no loss of generality. Figure 2.3 plots the probabilities for $\phi_i = 0.7$ and 0.9 , setting different d_i 's in each panel and varying w_i on the x -axis. The panels show clearly different curves between the cases of $d_i < \mu_i$ and $d_i > \mu_i$. In the former case ($d_i = 0.3, 0.4$), as w_i gets larger, p_1 goes down for small values of w_i and then increases. For the small values of w_i , the latent process β_{it} does not tend to drop below the threshold due to a relatively small variance. As the variance gets larger, the probability that β_{it} stays below the threshold increases. At a certain value of w_i , the probability p_1 starts to increase because the probability that β_{it} goes below the lower threshold ($-d_i$) increases as w_i gets larger. For the latter case ($d_i = 0.6 : 0.8$), because the marginal mean of the β_{it} process is below the threshold, β_{it} more often floats above d_i or below $-d_i$ as the variance gets larger. Figure 2.4 shows the probabilities for $d_i = 0.4$

(i) $\phi_i = 0.7$

(ii) $\phi_i = 0.9$

FIGURE 2.3: Shrinkage probabilities given (ϕ_i, w_i, d_i) . $p_1 = \Pr(s_{it} = 1)$, $p_{00} = \Pr(s_{it} = 0 | s_{i,t-1} = 0)$, $p_{11} = \Pr(s_{it} = 1 | s_{i,t-1} = 1)$, and $\mu_i = 0.5$.

(i) $d_i = 0.4$

(ii) $d_i = 0.7$

FIGURE 2.4: Shrinkage probabilities given (ϕ_i, w_i, d_i) . $p_1 = \Pr(s_{it} = 1)$, $p_{00} = \Pr(s_{it} = 0 | s_{i,t-1} = 0)$, $p_{11} = \Pr(s_{it} = 1 | s_{i,t-1} = 1)$, and $\mu_i = 0.5$.

Table 2.1: Shrinkage probability $\Pr(s_t = 0)$ marginalized on $(\boldsymbol{\theta}_i, d_i)$ with $v_i^{-2} \sim G(20, S_0)$. Standard deviations are in parentheses.

	$S_0 = 0.01$	0.1	1.0	10.0
$K = 1$	0.658 (0.137)	0.626 (0.162)	0.563 (0.187)	0.478 (0.185)
2	0.799 (0.143)	0.779 (0.164)	0.742 (0.184)	0.695 (0.197)
3	0.885 (0.132)	0.872 (0.149)	0.850 (0.168)	0.825 (0.178)
4	0.932 (0.114)	0.924 (0.127)	0.910 (0.141)	0.896 (0.150)

and 0.7, setting different ϕ_i 's in each panel and varying w_i on the x -axis. The conditional probabilities p_{00} and p_{11} exhibit different curves with respect to values of ϕ_i 's, although the marginal probability p_1 is stable.

Next, consider the shrinkage probability marginalized over $(\boldsymbol{\theta}_i, d_i)$ with respect to the following priors: $\mu_i \sim N(0, 1)$, $(\phi_i + 1)/2 \sim B(20, 1.5)$, $v_i^{-2} \sim G(20, S_0)$, and d_i follows the prior of eqn.(2.7). Table 2.1 reports the shrinkage probability $\Pr(s_t = 0)$ for selected S_0 and K_i . Note that the prior mean of v_i is $S_0/19$. For each K_i , the marginal shrinkage probability decreases as S_0 gets larger, although the change seems modest when $K = 3$ or 4. For each S_0 , the shrinkage probability increases as K_i gets larger. The following analyses assume $K_i = 3$ throughout this dissertation; those standard priors imply the marginal shrinkage probability is about 80-90%.

2.3 Bayesian computation

For estimation of the LTM, a Markov chain Monte Carlo (MCMC) algorithm is developed by extending standard analytic and MCMC methods for the dynamic regression model (West and Harrison 1997; Prado and West 2010) to incorporate the latent threshold structure. Based on observations $\mathbf{y}_{1:T} = (y_1, \dots, y_T)$, the MCMC algorithm is constructed for simulation of the full joint posterior $p(\boldsymbol{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \mathbf{d} | \mathbf{y}_{1:T})$, where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$. This section outlines components of the MCMC computa-

tions, and details are provided in Appendix A.1.

First, sampling of the VAR model parameters Θ of the latent process conditional on $(\beta_{1:T}, \mathbf{d})$ is standard, reducing to the set of conditionally independent posteriors $p(\theta_i | \beta_{i,1:T}, d_i)$ due to the diagonal assumption for Φ and $\mathbf{V}$. Under traditional priors for θ_i , the procedure utilizes standard Metropolis-Hastings (MH) methods of sampling parameters in the univariate AR(1) model.

The second MCMC component is new, addressing the key issue of sampling the latent state process $\beta_{1:T}$. Conditional on $(\Theta, \sigma, \mathbf{d}, \mathbf{y}_{1:T})$, a direct Metropolis-within-Gibbs sampling strategy is used for simulation of $\beta_{1:T}$; it sequentially generates β_t given $\beta_{-t} = \beta_{1:T} \setminus \beta_t$ through each t . The proposal distribution for this MH sampler is based on the conditional posterior from usual dynamic models without thresholds, formally arising by fixing $\mathbf{s}_t = \mathbf{1}$, which results in $N(\beta_t | \mathbf{m}_t, \mathbf{M}_t)$, where

$$\begin{aligned} \mathbf{M}_t &= \{ \sigma^{-2} \mathbf{x}_t \mathbf{x}_t' + \mathbf{V}^{-1} (\mathbf{I} + \Phi' \Phi) \}^{-1}, \\ \mathbf{m}_t &= \mathbf{M}_t [\sigma^{-2} \mathbf{x}_t y_t + \mathbf{V}^{-1} \{ \Phi (\beta_{t-1} + \beta_{t+1}) + (\mathbf{I} - 2\Phi + \Phi' \Phi) \boldsymbol{\mu} \}], \end{aligned}$$

for $t = 2, \dots, T-1$. For $t = 1$ and $t = T$ a slight modification is required, with details in Appendix A.1. The MH algorithm uses this as the proposal distribution to generate a candidate β_t^* . Note that this is a natural and reasonable proposal strategy; the proposal density will be close to the exact conditional posterior in dimensions such that the elements of β_t are large, and smaller elements in candidate draws will tend to agree with the likelihood component of the exact posterior as they imply limited or no impact on the observation equation by construction. The MH algorithm is completed by accepting the candidate with probability

$$\alpha(\beta_t, \beta_t^*) = \min \left\{ 1, \frac{N(y_t | \mathbf{x}_t' \mathbf{b}_t^*, \sigma^2) N(\beta_t | \mathbf{m}_t, \mathbf{M}_t)}{N(y_t | \mathbf{x}_t' \mathbf{b}_t, \sigma^2) N(\beta_t^* | \mathbf{m}_t, \mathbf{M}_t)} \right\},$$

where $\mathbf{b}_t = \beta_t \circ \mathbf{s}_t$ is the current LTM state and $\mathbf{b}_t^* = \beta_t^* \circ \mathbf{s}_t^*$ the candidate.

It is possible to develop a block sampling extension of this MH strategy by using a forward-filtering, backward sampling (FFBS) algorithm (de Jong and Shephard 1995; Durbin and Koopman 2002) on the non-thresholded model to generate proposals for the full sequence $\beta_{1:T}$. This follows related approaches using this idea of a global FFBS-based proposal (Prado and West 2010; Niemi and West 2010). However, the resulting acceptance rates decrease exponentially with T and some simulation experiments indicate unacceptably low acceptance rates in several examples, especially with higher levels of sparsity when the proposal distribution from the non-threshold model agrees less and less with the posterior under the LTM. A modified multi-block approach has also been examined; the FFBS is applied separately with a block conditional on state vectors in the adjacent blocks, while it yielded limited success in improving acceptance rates. Although the simpler single-move strategy has the potential to mix slowly, it is computationally efficient and so can be run far longer than blocking approaches to balance mixing issues. Experiments have shown practically acceptable acceptance rates and mixing in multiple examples illustrated here, and hence it is adopted as standard. Further, the single-move sampler is technically preferable for extending the model by adding some constraint on β_t or s_t ; we only need to add rejection sampling in the MH proposal step to check whether the proposal point satisfies the constraint.

The final MCMC component is the generation of thresholds $\mathbf{d}$. The algorithm adopts a direct MH approach with candidate drawn from the prior. Some experiments with the direct MCMC summarized here show that mixing is good and convergence clean as in standard DLM analyses; adding the threshold processes and parameters does not introduce any significant conceptual complications as shown by a simulation study in the next section.

2.4 Simulation study

For illustration of the model mechanism and estimation methodology, a sample of size $T = 500$ comes from the LTM with $k = 3$ predictors and where only the first two predictors are relevant. The x_{it} 's are generated from i.i.d. $U(-0.5, 0.5)$ and $\sigma = 0.15$. For $i = 1, 2$, the simulation uses parameters $(\mu_i, \phi_i, v_i, d_i) = (0.5, 0.99, 0.1, 0.4)$; while for $i = 3$, $\beta_{3t} = 0$ for all t . Figure 2.5 graphs the true values of the time-varying coefficients, indicating the within-threshold sparsity periods by shading. The following prior distributions are used: $\mu_i \sim N(0, 1)$, $(\phi_i + 1)/2 \sim B(20, 1.5)$, $v_i^{-2} \sim G(3, 0.03)$, and $\sigma^{-2} \sim G(3, 0.03)$. The prior mean and standard deviation of ϕ_i are $(0.86, 0.11)$; those for each of v_i^2 and σ^2 are $(0.015, 0.015)$. The conditional prior for thresholds is $U(d_i | 0, |\mu_i| + K_i v_i)$ with $K_i = 3$. This implies the marginal prior probability of shrinkage, $p(b_{it} = 0) \approx 0.85$. MCMC used $J = 50,000$ iterates after discarding a burn-in period of 5,000 samples.

Figure 2.5 shows summary results. The posterior means of the time-varying coefficients trace their true values, and the posterior probabilities of $s_{it} = 0$ appropriately identify the periods of non-zero effective coefficients for β_{1t} and β_{2t} as well as those of zero coefficients for β_{3t} . Table 2.2 reports the posterior estimates for selected parameters; posterior means are close enough to the true values that the corresponding

Table 2.2: Simulation example: Posterior estimates for selected parameters with credible intervals based on 2.5%, 97.5% quantiles of posterior MCMC draws.

	True	Mean	Stdev.	95% C.I.
μ_1	0.5	0.575	0.455	-0.502, 1.343
ϕ_1	0.99	0.991	0.005	0.978, 0.998
v_1	0.1	0.089	0.012	0.067, 0.115
σ	0.15	0.154	0.006	0.148, 0.167
d_1	0.4	0.220	0.128	0.008, 0.464

FIGURE 2.5: Simulation example: Trajectories of dynamic regression parameters. True values (top), posterior means, 95% credible intervals (middle), and posterior probabilities of $s_{it} = 0$ (bottom). The shadows in the top panels refer to the periods when $s_{it} = 0$. The thresholds in the middle panels refer to their posterior means.

FIGURE 2.6: Simulation example: Priors (solid lines) and posteriors (histograms) for thresholds. In the lower row, the graphs are simply truncated on the d_i axes for visual clarity.

95% credible intervals include them.

Figure 2.6 displays marginal prior density of the latent thresholds from eqn.(2.8), and resulting posteriors. The first two columns exhibit clear and strong evidence for low threshold values with information updated from prior to posterior given data. The posterior on the threshold d_3 is appropriately very close to the prior, favoring slightly larger values, indicating the lack of information in the data about the actual value of d_3 .

There is no inherent interest in the thresholds d_i as parameters to be inferred; the interest is wholly in their roles in inducing data relevant sparsity in the posterior for primary time-varying parameters and the resulting predictive distributions. Correspondingly, analysis does not focus on the underlying values of the latent processes β_{it} when they are below threshold; all that matters is that we then know $b_{it} = 0$. Depending on the model and data context, any one of the threshold parameters may be well-estimated, or identified, in the sense of the posterior being rather precise relative to the prior; in contrast, the data may be basically uninformative about other thresholds, so that their posteriors are minor modifications of their priors. In the dynamic regression context, a data set in which there is little evidence of a regression relationship between y_t and x_{it} , like $i = 3$ in the simulation example, will lead to very high levels of thresholding on β_{it} . Then, posteriors will suggest larger values of d_i while providing no information about the β_{it} process; the latent thresholds are simply vehicles to inducing relevant time-varying sparsity mechanisms, and whether or not they are well-estimated from the data is of no intrinsic importance or interest. At the other extreme, a strong relationship sustained over time is consistent with a low threshold and the posterior will indicate such.

Analyses with larger and moderately smaller values of K_i yield substantially similar inferences on the dynamic state vectors and their sparsity patterns over time.

Some experiments indicate that $K_i < 1$ has substantial impact on the inference; the LTM mechanism cannot detect the whole-sequence shrinkage for β_{3t} in this simulation example. Larger K_i can assure the shrinkage, but provide more uncertainty. To balance these two points, simulations in broader settings suggest $K_i = 3$, taken as a standard specification throughout this dissertation.

One remark is in order here in connection with sparsity and Bayesian computation. The computations for the LTM require additional components relative to those in the standard DLMS, although the cost of employing the LTM structure is small. Another simulation example fits non-threshold (NT; standard DLM) and latent threshold (LT) models to simulation data generated from DLMS, with the same setting as the previous simulation example except that $\mu_i = 3$ and $s_{it} = 0$ for all i and t . This reveals that posterior estimates are almost the same between the two models, even for β . The posterior distribution of d_i appears roughly flat over a region $(0, 2)$, yielding $\Pr[s_{it} = 0 | \mathbf{y}] \approx 1$ for all i and t . The inefficiency factor from the LT model is slightly higher due to the generation of $\mathbf{d}$ and the rejection step in the generation of β and Θ , while the difference is marginal. This implies that the cost of employing the LTM mechanism is negligible in terms of posterior inference; in contrast, the utility of inducing dynamic sparsity is evident as shown in the following applications.

2.5 Generalization

Subsequent chapters apply the LTM structure to multivariate time series models with real data applications. The analyses in the subsequent chapters use the basic formulation of the LTM defined in Section 2.1, while several generalizations can be considered for the LTM structure; this section discusses such extensions, which can be useful and important for some applications.

2.5.1 Threshold and shrinkage indicators

Asymmetric thresholds: The shrinkage region defined by eqn.(2.3) is $(-d_i, d_i)$, symmetric and centered at zero. We can assume generalized thresholds of a shrinkage region (d_i^L, d_i^U) , where $d_i^L < d_i^U$, with a shrinkage function

$$s_{it} = I(\beta_{it} \leq d_i^L, \text{ or } \beta_{it} \geq d_i^U).$$

This includes a one-sided threshold as a special case, i.e., $s_{it} = I(\beta_{it} > d_i^U)$ with $d_i^L = -\infty$, or $s_{it} = I(\beta_{it} < d_i^L)$ with $d_i^U = \infty$.

Joint indicator: We can specify a joint indicator,

$$\tilde{s}_t = I(|\beta_{1t}| \geq d_1, \dots, |\beta_{kt}| \geq d_k) = I(|\beta_{1t}| \geq d_1) \cdots I(|\beta_{kt}| \geq d_k),$$

with a link function $\mathbf{b}_t = \boldsymbol{\beta}_t \tilde{s}_t$. All the coefficients in $\mathbf{b}_t$ are simultaneously shrunk at zero when at least one of the latent processes $\boldsymbol{\beta}_t$ falls into the shrinkage region.

Exclusive variable selection: Consider a case where only or at most one variable (predictor) should be selected at each time point, excluding all the other variables. For example, consider a dynamic regression model with $k = 2$ covariates; a condition that at most one covariate is included in the measurement equation at each time is defined by

$$\begin{aligned} s_{1t} &= I(|\beta_{1t}| \geq d_1)I(|\beta_{2t}| < d_2), \\ s_{2t} &= I(|\beta_{1t}| < d_1)I(|\beta_{2t}| \geq d_2). \end{aligned}$$

This regression always includes one or no covariate, excluding the case of two covariates jointly predicting the response. A condition of including only one covariate requires a more complicated form on $(\boldsymbol{\beta}_t, \mathbf{d})$. These constraints can be generalized to the case where exactly m or at most m variables are included in predicting the

response at each time point. In either case, implementation of the posterior computation is straightforward in the single-move sampler used for β_t , where the additional constraint can be easily incorporated with the rejection sampling method as mentioned before. In practice, however, the constraint for selecting m covariates often appears intractable because the constraint limits the sample space of $(\beta, \mathbf{d})$, which leads to an unacceptably low acceptance rate; development of a more elaborate sampling scheme to address this issue is an open problem.

2.5.2 *Nonstationary latent process*

The basic LTM is based on the stationary latent process such as eqn.(2.4), which yields the latent threshold prior of eqn.(2.7); a nonstationary process can also be specified in the LTM mechanism. A nonstationary process does not have a marginal distribution; therefore, we need to specify a different prior for the latent threshold. An easy and pragmatic approach is to assume a uniform distribution, $d_i \sim U(0, D_i)$, where D_i is a pre-specified hyperparameter. For example, we first estimate the LTM without latent thresholds, assuming $\mathbf{s}_t = \mathbf{0}$ for all t , as a reference model to produce the posterior samples of the time-varying coefficient, $\hat{\beta}_{it}^{(j)}$, for $j = 1, \dots, J$, where J denotes the MCMC iteration size. Then, we set $D_i = \max_{j,t} (|\hat{\beta}_{it}^{(j)}|)$. When we have a wealth of historical data at hand, then an initial stretch of the data may be used for estimating the reference model in order to build the latent threshold prior for the remaining time series data.

2.5.3 *General covariance matrix for latent process*

The covariance matrix $\mathbf{V}$ of the innovation in the latent process is assumed as diagonal in the basic formulation of LTM. This can be relaxed; assuming a general (non-diagonal) covariance matrix $\mathbf{V}$ is a natural extension and easily implemented

based on the developed MCMC algorithm. Regarding the prior for latent thresholds, we can still exploit the marginal distribution of each latent process $\{\beta_{i,1:T}\}$ to form eqn.(2.7). The MCMC algorithm for the basic LTMs is directly applicable, except for sampling $\mathbf{V}$ from a truncated inverse Wishart distribution. Note that, as the dimension k of the regression variables gets large, posterior inference of $\mathbf{V}$ could face increasing uncertainty; some structure such as sparsity in $\mathbf{V}$ would be preferable, depending on context and data.

Latent Threshold Multivariate Models

This chapter applies the LTM structure to several multivariate time series models which have been studied for a wide range of applications. In the context of multivariate analysis, increasing dimension of responses and model structure drives a need for refined prior: model structuring for parsimony in the face of increasing uncertainty. The LTM mechanism addresses this issue in time-varying multivariate models such as time-varying VAR (TV-VAR), time-varying covariance matrix, and dynamic factor models; the purpose of modeling LTM-based time-varying parameter multivariate models is to induce effectively parsimonious structure where time-varying parameter processes may take practically non-negligible values for some epochs, but be effectively zero and practically irrelevant at other times. Details of the model specification and estimation scheme are described in the following sections.

3.1 Latent threshold time-varying VAR (LT-VAR) models

Traditional, constant coefficient VAR models are central to applied time series analysis (e.g., Prado and West 2010), and various approaches to TV-VAR modeling

are becoming increasingly prevalent in many fields. With increasing variables in the VAR, the number of coefficients in autoregressive coefficient matrices escalates as does the need for parameter constraints. Recent Bayesian VAR analysis addresses this using shrinkage and sparsity-inducing priors of various forms (Fox et al. 2008; George et al. 2008; Wang 2010) for traditional constant coefficient VAR models, but the induction of zeros into increasingly sparse time-varying coefficient matrices, with allowance for time-variation in the occurrence of non-zero values as well as local changes in coefficients when they are non-zero, has been challenging; the LTM ideas provide an approach.

3.1.1 Model structure

For the time series of $m \times 1$ vector responses $\mathbf{y}_t$, consider the basic TV-VAR model with autoregressive order p as

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{B}_{1t}\mathbf{y}_{t-1} + \cdots + \mathbf{B}_{pt}\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (3.1)$$

where $\mathbf{c}_t$ is the $m \times 1$ vector of time-varying intercept, $\mathbf{B}_{jt}$ is the $m \times m$ matrix of time-varying coefficients at lag j , ($j = 1, \dots, p$), and $\boldsymbol{\Sigma}_t$ is the $m \times m$ innovations covariance matrix that is also often time-varying. For each time t , define the $k \times 1$ vector $\mathbf{b}_t$, with $k = m(1 + pm)$ here, by stacking the set of $\mathbf{c}_t$ and $\mathbf{B}_{jt}$'s by rows and by order $j = 1, \dots, p$; and also define the corresponding $m \times k$ matrix $\mathbf{X}_t = \mathbf{I} \otimes (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$. Then the model can be written as a multivariate dynamic regression, viz.

$$\mathbf{y}_t = \mathbf{X}_t\mathbf{b}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t). \quad (3.2)$$

The time-varying coefficient vector $\mathbf{b}_t$ is often assumed to follow a VAR(1) process. This is generalized to the LTM framework by overlaying thresholds as in eqns.(2.2)-

(2.4). Specifically, with $k \times 1$ vectors β_t and s_t :

$$\mathbf{b}_t = \beta_t \circ \mathbf{s}_t, \quad (3.3)$$

$$\beta_t = \mu_\beta + \Phi_\beta(\beta_{t-1} - \mu_\beta) + \eta_{\beta_t}, \quad \eta_{\beta_t} \sim N(\mathbf{0}, \mathbf{V}_\beta), \quad (3.4)$$

and $s_{it} = I(|\beta_{it}| \geq d_i)$, for $i = 1, \dots, k$. This specification is referred to as the LT-VAR model; the LTM structure provides both whole-sequence and dynamic, adaptable variable selection for time-varying coefficients, with the ability to switch a specific coefficient, or set of coefficients, in/out of the model as defined by the threshold mechanism.

3.1.2 Time-varying covariance matrix

Modeling time-varying covariance matrices, both residual/error matrices in observation equations of dynamic models such as Σ_t in eqn.(3.2) and innovations/evolution variance matrices for latent processes, is key to analyzing multivariate series. To date, time-varying covariance matrix models have been widely developed (Quintana and West 1987; Bollerslev et al. 1992; Aguilar and West 2000; Shephard 2005; Carvalho and West 2007). Two popular and useful approaches are the discounting method and the Cholesky-type decomposition.

The discounting method is based on “locally smooth” discount factor-based models, where variances and covariances are allowed to change typically slowly and unpredictable in the short-term (Quintana and West 1987; Quintana et al. 2003). The sequential updating involves discounting of cumulated information continuously decaying over time. Define the precision matrix, $\Omega_t = \Sigma_t^{-1}$, and a set of disturbances, $\mathbf{D}_t = \{\varepsilon_1, \dots, \varepsilon_t\}$. Based on a discount factor λ_w ($0 \leq \lambda_w \leq 1$), the Wishart matrix discounting process is specified by the sequential updating of the precision matrix

with observed data, that is,

$$\begin{aligned}\boldsymbol{\Omega}_{t-1}|\boldsymbol{D}_{t-1} &\sim W(w_{t-1}, \boldsymbol{S}_{t-1}^{-1}), \\ \boldsymbol{\Omega}_t|\boldsymbol{D}_{t-1} &\sim W(\lambda_w w_{t-1}, (\lambda_w \boldsymbol{S}_{t-1})^{-1}),\end{aligned}$$

where W denotes the Wishart distribution. This evolution decreases the degrees of freedom discounted by λ_w , which implies an increase in uncertainty for the precision matrix. A higher value of λ_w , typically close to one, leads to a slow and steady change of variances and covariances.

The Cholesky-type decomposition method is based on a triangular reduction of $\boldsymbol{\Sigma}_t$, defined by $\boldsymbol{A}_t \boldsymbol{\Sigma}_t \boldsymbol{A}_t' = \boldsymbol{\Lambda}_t$, where $\boldsymbol{A}_t$ is the lower triangular matrix of covariance components with all diagonal elements equal to one and $\boldsymbol{\Lambda}_t$ is diagonal with positive elements. That is, $\boldsymbol{\Lambda}_t^{1/2}(\boldsymbol{A}_t')^{-1}$ is the Cholesky component of $\boldsymbol{\Sigma}_t$, viz.

$$\boldsymbol{\Sigma}_t = \boldsymbol{A}_t^{-1} \boldsymbol{\Lambda}_t (\boldsymbol{A}_t')^{-1}, \quad (3.5)$$

where

$$\boldsymbol{A}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{m1,t} & \cdots & a_{m,m-1,t} & 1 \end{pmatrix}, \quad \boldsymbol{\Lambda}_t = \begin{pmatrix} \sigma_{1t}^2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{mt}^2 \end{pmatrix}, \quad (3.6)$$

which implies $\boldsymbol{\varepsilon}_t = \boldsymbol{A}_t^{-1} \boldsymbol{\Lambda}_t^{1/2} \boldsymbol{e}_t$, where $\boldsymbol{e}_t \sim N(\mathbf{0}, \boldsymbol{I})$.

The basic time-varying model for the Cholesky parameters follows Primiceri (2005). Let $\boldsymbol{a}_t$ be the vector of the strictly lower-triangular elements of $\boldsymbol{A}_t$ (stacked by rows), and define $\boldsymbol{h}_t = (h_{1t}, \dots, h_{mt})'$, where $h_{jt} = \log \sigma_{jt}^2$, for $j = 1, \dots, m$. The dynamics of the covariances and variances are specified jointly with eqn.(3.4) as

$$\boldsymbol{a}_t = \boldsymbol{\mu}_a + \boldsymbol{\Phi}_a(\boldsymbol{a}_{t-1} - \boldsymbol{\mu}_a) + \boldsymbol{\eta}_{a_t}, \quad (3.7)$$

$$\boldsymbol{h}_t = \boldsymbol{\mu}_h + \boldsymbol{\Phi}_h(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_{h_t}, \quad (3.8)$$

where $(\mathbf{e}'_t, \boldsymbol{\eta}'_{\boldsymbol{\beta}}, \boldsymbol{\eta}'_{\mathbf{a}}, \boldsymbol{\eta}'_{\mathbf{h}})' \sim N[\mathbf{0}, \text{diag}(\mathbf{I}, \mathbf{V}_{\boldsymbol{\beta}}, \mathbf{V}_{\mathbf{a}}, \mathbf{V}_{\mathbf{h}})]$, and each of the matrices $(\boldsymbol{\Phi}_*, \mathbf{V}_*)$, for $*$ $\in \{\boldsymbol{\beta}, \mathbf{a}, \mathbf{h}\}$, is assumed diagonal. With $|\phi_*| < 1$ for all diagonal elements in $\boldsymbol{\Phi}_*$, all univariate time-varying parameters follow stationary AR(1) models. The specific cases of the log variances h_{it} define traditional univariate stochastic volatility (SV) models for which the MCMC strategies are standard and widely used both alone and as components of overall MCMC strategies for more complex models (Jacquier et al. 1994; Kim et al. 1998; Aguilar and West 2000; Omori et al. 2007; Prado and West 2010, Chapter. 7).

One key feature of the Cholesky-construction for time-varying covariance matrices is that we can translate the resulting dynamic model for $\boldsymbol{\Sigma}_t$ into a conditional DLM with $\mathbf{a}_t$ as the latent state vector. Define $\tilde{\mathbf{y}}_t = (\tilde{y}_{1t}, \dots, \tilde{y}_{mt})' = \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t$, and the $m \times m(m-1)/2$ matrix

$$\tilde{\mathbf{X}}_t = \begin{pmatrix} 0 & \dots & & & & & 0 \\ -\tilde{y}_{1t} & 0 & 0 & \dots & & & \vdots \\ 0 & -\tilde{y}_{1t} & -\tilde{y}_{2t} & 0 & \dots & & \\ 0 & 0 & 0 & -\tilde{y}_{1t} & \dots & & \\ \vdots & & & & \ddots & 0 & \dots & 0 \\ 0 & \dots & & & 0 & -\tilde{y}_{1t} & \dots & -\tilde{y}_{m-1,t} \end{pmatrix}. \quad (3.9)$$

From the model identity $\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{A}_t^{-1} \boldsymbol{\Lambda}_t \mathbf{e}_t$ and using the lower triangular form of $\mathbf{A}_t$, we deduce

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}_t \mathbf{a}_t + \boldsymbol{\Lambda}_t \mathbf{e}_t, \quad (3.10)$$

for all t . This couples with the state evolution of eqn.(3.7) to define a conditional DLM; the MCMC analysis will then extend to include a component to resample the $\mathbf{a}_{1:T}$ sequence at each iteration, using the efficient FFBS strategy for conditionally normal DLMs.

As mentioned by Primiceri (2005), it is not straightforward to theoretically explore compounded processes of the covariance/variance elements in the Cholesky-type covariance matrix. A simulation result is reported in Appendix B.

3.2 Latent threshold time-varying covariance matrix

Some recent innovations in sparsity modeling for covariance matrices are based on the discounting Wishart process in the context of Gaussian graphical modeling. Carvalho and West (2007) propose a model framework of normal hyper-inverse Wishart (NHIW) distributions for sequential learning via forward filtering and retrospective analysis to induce dynamic sparsity covariance matrices. Linked in part to developments in sparse factor modeling (e.g., West 2003; Carvalho et al. 2008), other approaches have adopted new priors on elements of triangular square-roots of covariance matrices such as $\mathbf{A}_t$ defined here. George et al. (2008) do this with priors allowing zero elements in Cholesky components of constant covariance matrices in a constant parameter VAR model context, building on previous, non-sparse approaches utilizing such constructions (Pinheiro and Bates 1996; Pourahmadi 1999; Smith and Kohn 2002). The construction has also appeared in time-varying covariance matrix modeling in VAR contexts (Cogley and Sargent 2005; Primiceri 2005; Lopes et al. 2012) which is one point of departure here; this section uses the above Cholesky structure and embeds it in a novel LTM framework, combining models for stochastic time-variation in variance matrices with the natural threshold-based sparsity inducing mechanism to shrink subsets of the lower-triangle of $\mathbf{A}_t$ to zero adaptively and dynamically.

The LTM structure is incorporated into time-varying covariance matrices, to enable shrinkage to zero of subsets of the elements of $\mathbf{a}_t$ over periods of time consistent with sparse structure. This simply directly adapts the LTM strategy from Chapter

2 into this context, applying the latent thresholding ideas now to the state vector $\mathbf{a}_t$ in the reformulated model above. That is, introduce a latent VAR process $\boldsymbol{\alpha}_t$ to substitute for $\mathbf{a}_t$ in the conditional model of the previous section. With $\boldsymbol{\alpha}_t$ having elements α_{it} 's stacked as are the elements of $\mathbf{a}_t$, the new idea formulates $\mathbf{a}_t = \boldsymbol{\alpha}_t \circ \mathbf{s}_{\mathbf{a}_t}$ with an indicator vector $\mathbf{s}_{\mathbf{a}_t}$ of the form of LTM: $s_{ait} = I(|\alpha_{it}| \geq d_{ai})$. The LTM extension of the conditional DLM with respect to $\mathbf{a}_t$ states is then

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}_t \mathbf{a}_t + \boldsymbol{\Lambda}_t \mathbf{e}_t, \quad (3.11)$$

$$\mathbf{a}_t = \boldsymbol{\alpha}_t \circ \mathbf{s}_{\mathbf{a}_t}, \quad (3.12)$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \boldsymbol{\eta}_{\boldsymbol{\alpha}t}, \quad \boldsymbol{\eta}_{\boldsymbol{\alpha}t} \sim N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\alpha}}), \quad (3.13)$$

where $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ and $\mathbf{V}_{\boldsymbol{\alpha}}$ are diagonal matrices. The MCMC estimation procedure developed in Section 2.3 can now be straightforwardly applied to this time-varying, sparsity-inducing LTM extension of the AR(1) Cholesky based time-varying covariance model. Computational details for the LT-VAR model with the LT time-varying covariance matrix are explained in Appendix A.2.

One point of interest is that a sparse $\mathbf{A}_t$ matrix can translate into a sparse precision matrix $\boldsymbol{\Omega}_t = \mathbf{A}_t' \boldsymbol{\Lambda}_t^{-1} \mathbf{A}_t$; the more zeros there are in the lower triangle of $\mathbf{A}_t$, the more zeros there will be in the precision matrix. Hence the LTM defines an approach to time-varying sparsity modeling for precision matrices, and therefore time-varying graphical models as a result. As mentioned before, graphical models characterize conditional independencies of multivariate series via graphs and zeros in the precision matrix of a normal distribution correspond to missing edges in the graph whose nodes are the variables (e.g., Jones et al. 2005). The LTM approach now clearly defines a new class of models for time-variation in the structure of the graphical model underlying $\boldsymbol{\Sigma}_t$, since the appearance of zeros in its inverse $\boldsymbol{\Omega}_t$ is driven by the latent stochastic thresholding structure; edges may come in/out the

graph over time, so extending previous time series graphical models that require a fixed graph (Carvalho and West 2007; Wang and West 2009) to a new class of dynamic graphs so induced. This issue is revisited in Chapter 7.

3.3 Latent threshold factor (LTF) models

Dynamic factor models have become standard tools for multivariate time series and for multivariate stochastic volatility in particular (Aguilar et al. 1999; Pitt and Shephard 1999; Aguilar and West 2000). This section adopts a general framework that builds on these prior approaches and extends to time-varying factor loadings matrices (Lopes and Carvalho 2007; Carvalho et al. 2011) as well as integrating short-term dependence structure into latent factor processes. Overlaying the resulting general class of dynamic factor models are the LTM mechanisms that define the ability for time-varying factor loadings to shrink to zero dynamically; i.e., dynamic sparsity models for loadings as well as possibly for other model components including dynamic regression coefficients.

The class of models for the $m \times 1$ vector response time series $\mathbf{y}_t$ is

$$\mathbf{y}_t = \mathbf{B}_t \mathbf{f}_t + \mathbf{c}_t + \mathbf{Z}_t \mathbf{x}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}_t), \quad (3.14)$$

$$\mathbf{f}_t = \mathbf{G}_t \mathbf{f}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (3.15)$$

where $\mathbf{f}_t$ is a $q \times 1$ vector of latent factors at time t , evolving according to a VAR(1) model with possibly time-varying AR coefficient matrix $\mathbf{G}_t$, and also possibly time-varying innovations covariance matrix $\mathbf{W}_t$ (both $q \times q$); $\mathbf{B}_t$ is the $m \times q$ time-varying loadings matrix of the dynamic latent factor component; $\mathbf{c}_t$ is a local trend term; $\mathbf{x}_t$ is a $\ell \times 1$ vector of covariates known at time t with an associated $m \times \ell$ time-varying parameter matrix $\mathbf{Z}_t$; $\boldsymbol{\nu}_t = (\nu_{1t}, \dots, \nu_{mt})'$ is a residual term with a diagonal time-varying volatility matrix $\boldsymbol{\Lambda}_t$; and the $\boldsymbol{\nu}_t$ and $\boldsymbol{\varepsilon}_\tau = (\varepsilon_{1\tau}, \dots, \varepsilon_{q\tau})'$ sequences

are independent and mutually independent. Following prior works, factor model identification is imposed by constraining $\mathbf{B}_t$ to have an upper right triangle of zeros and upper diagonal elements of unity (Aguilar and West 2000; Lopes and West 2004).

In the context of financial time series analysis, this model class represents a generalization of the standard factor models in a number of respects relevant to financial prices and returns. The elements of the factor loadings matrix (as well as regression parameter matrix) are allowed to change over time (Lopes and Carvalho 2007) while being subject to the LTM structure to shrink them to zero when data suggest they are irrelevant. Further, the latent factors follow an autoregressive process, so providing opportunity for improvements in short-term forecasting of underlying time-varying common movements in the multiple time series. Adding volatility to this latter component is one further extension of some practical value, discussed below.

A key feature of the proposed model is development on an underlying AR(1) process structure for each of the free elements of $\mathbf{B}_t$ as well as for all univariate entries in $\mathbf{Z}_t$. For the factor loadings, this is related to process models adopted for dynamic factor loadings in Lopes and Carvalho (2007). This basic idea is here adapted using the LTM mechanism as follows.

Dynamic factor loadings processes: The LTM structure is adopted for the free parameter processes in $\mathbf{B}_t$; i.e., those elements below the upper diagonal of the $m \times q$ matrix $\mathbf{B}_t$ (n.b. $m > q$). Write $\mathbf{b}_t$ for the column stack of those free elements, so that $\mathbf{b}_t$ is a $k \times 1$ vector where $k = mq - q(q + 1)/2$ here, with univariate elements b_{it} , $i = 1 : k$. The model assumes that the b_{it} independently follows the LTM-based

AR(1) process as in Sections 2.1 and 3.1,

$$\mathbf{b}_t = \boldsymbol{\beta}_t \circ \mathbf{s}_{\beta t}, \quad (3.16)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_\beta + \Phi_\beta(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_\beta) + \boldsymbol{\eta}_{\beta t}, \quad \boldsymbol{\eta}_{\beta t} \sim N(\mathbf{0}, \mathbf{V}_\beta), \quad (3.17)$$

with $s_{it} = I(|\beta_{it}| \geq d_i)$, for $i = 1, \dots, k$.

Dynamic regression parameter processes: Write $\mathbf{z}_t$ for the column stack of the elements in $\mathbf{Z}_t$ matrix; i.e., $m\ell \times 1$ vector with i -th element z_{it} . The LTM structure is employed with a corresponding latent process $\boldsymbol{\zeta}_t = (\zeta_{1t}, \dots, \zeta_{m\ell, t})'$ and thresholding indicator variable $\mathbf{s}_{zt} = (s_{z1t}, \dots, s_{z, m\ell, t})'$ as

$$\mathbf{z}_t = \boldsymbol{\zeta}_t \circ \mathbf{s}_{zt}, \quad (3.18)$$

$$\boldsymbol{\zeta}_t = \boldsymbol{\mu}_\zeta + \Phi_\zeta(\boldsymbol{\zeta}_{t-1} - \boldsymbol{\mu}_\zeta) + \boldsymbol{\eta}_{\zeta t}, \quad \boldsymbol{\eta}_{\zeta t} \sim N(\mathbf{0}, \mathbf{V}_\zeta), \quad (3.19)$$

with $s_{zit} = I(|\zeta_{it}| \geq d_{zi})$, for $i = 1, \dots, m\ell$.

The key idea of the latent threshold structure is that the value of each of the dynamic regression and time-varying loading parameters is shrunk to zero when irrelevant in the model. On the factor component, the relevance of each factor to the response is time-varying; a factor plays a role in predicting the response only when the corresponding β_{it} is “large enough”. For each time series response variable, the factor may have a non-zero loading in some time periods but zero in others, depending on the data and context. This mechanism leads to dynamic model uncertainty in the factor loading matrix by neatly embodying time-varying sparsity/shrinkage and parameter reduction. The other components in the model are specified as follows.

Local mean: The time-varying intercept, or local mean $\mathbf{c}_t = (c_{1t}, \dots, c_{mt})'$ evolves based on VAR(1) models

$$\mathbf{c}_t = \boldsymbol{\mu}_c + \Phi_c(\mathbf{c}_{t-1} - \boldsymbol{\mu}_c) + \boldsymbol{\eta}_{ct}, \quad \boldsymbol{\eta}_{ct} \sim N(\mathbf{0}, \mathbf{V}_c). \quad (3.20)$$

Residual volatility: With $\mathbf{\Lambda}_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{mt}^2)$, standard stochastic volatility process models are adopted as in eqn.(3.8). The log volatilities $h_{it} = \log \sigma_{it}^2$, follow a stationary VAR(1) model

$$\mathbf{h}_t = \boldsymbol{\mu}_h + \Phi_h(\mathbf{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_{h_t}, \quad \boldsymbol{\eta}_{h_t} \sim N(\mathbf{0}, \mathbf{V}_h). \quad (3.21)$$

Factor evolution: For the parameters of the factor evolution model in eqn.(3.15), a constant and diagonal autoregressive parameter matrix $\mathbf{G}_t \equiv \mathbf{G} = \text{diag}(g_1, \dots, g_q)$ is assumed for all t . For the factor innovation volatility matrix $\mathbf{W}_t$, a simple assumption is a diagonal structure with each element following to a volatility process similar to that for $\mathbf{\Lambda}_t$. More generally, departing from the diagonal assumption, we can specify a time-varying general covariance matrix. Following the discussion in Section 3.1.2, consider the Cholesky decomposition of $\mathbf{W}_t$ such that $\mathbf{A}_t \mathbf{W}_t \mathbf{A}_t' = \boldsymbol{\Psi}_t$, where $\mathbf{A}_t$ is the lower triangular matrix as defined in eqn.(3.6) and $\boldsymbol{\Psi}_t$ is diagonal, that is,

$$\mathbf{W}_t = \mathbf{A}_t^{-1} \boldsymbol{\Psi}_t (\mathbf{A}_t')^{-1}. \quad (3.22)$$

If we assume $\mathbf{A}_t = \mathbf{I}$ for all t , then the model reduces to the standard independent factor innovation processes.

Let $\mathbf{a}_t$ be the vector of the free elements of $\mathbf{A}_t$ (stacked by rows) as in Section 3.1.2 and specify the process

$$\mathbf{a}_t = \boldsymbol{\mu}_a + \Phi_a(\mathbf{a}_{t-1} - \boldsymbol{\mu}_a) + \boldsymbol{\eta}_{a_t}, \quad \boldsymbol{\eta}_{a_t} \sim N(\mathbf{0}, \mathbf{V}_a). \quad (3.23)$$

For the volatility process, defining $\boldsymbol{\Psi}_t = \text{diag}(\psi_{1t}^2, \dots, \psi_{qt}^2)$, with $\lambda_{it} = \log \psi_{it}^2$ for $i = 1 : q$, the model adopts stochastic volatility models

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu}_\lambda + \Phi_\lambda(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\mu}_\lambda) + \boldsymbol{\eta}_{\lambda_t}, \quad \boldsymbol{\eta}_{\lambda_t} \sim N(\mathbf{0}, \mathbf{V}_\lambda). \quad (3.24)$$

For all the latent processes, the following analyses assume that $(\Phi_*, \mathbf{V}_*)$ are diagonal for $* \in \{\beta, \zeta, \mathbf{c}, \mathbf{h}, \mathbf{a}, \boldsymbol{\lambda}\}$, and $|\phi_*| < 1$ for all diagonal elements in Φ_* . These

specifications are labeled as the LT-Factor (LTF) model. Posterior computations use standard samplers for the latent factor models (Aguilar and West 2000; Lopes and West 2004) and the proposed methods for the LTM components. Details of the MCMC algorithm for the LTF models are described in Appendix A.3.

Application to Macroeconomic Time Series

4.1 Context and literature

Bayesian TV-VAR modeling is increasingly common in macroeconomic studies. For example, Primiceri (2005), Benati (2008), Benati and Surico (2008), Koop et al. (2009) and Nakajima et al. (2011) involve studies that aim to assess dynamic relationships between monetary policy and economic variables, typically focusing on changes in the exercise of monetary policy and the resulting effect on the rest of the economy. Structural shocks hitting the economy and simultaneous interactions between macroeconomic variables are identified by TV-VAR models. This chapter uses the LTM strategy for TV-VAR models, namely the LT-VAR model developed in Section 3.1, in analysis of US and Japanese macroeconomic time series data.

Since Cogley and Sargent (2005) and Primiceri (2005) developed the nowadays standard TV-VAR approach to macroeconomic analysis, various structures have been examined for time-varying parameters. Koop et al. (2009) examined whether parameters in TV-VAR models are time-varying or not at each time by incorporating a mixture innovation structure for time-varying parameters, where innovation errors

can take either zero or non-zero value depending on Bernoulli random variables. Korobilis (2012) developed Bayesian variable selection for TV-VAR coefficients as well as structural breaks. Chan et al. (2012) exploited Markov switching indicator variables for innovation errors of time-varying regression coefficients to explore a temporal variable selection mechanism (see a broad literature survey in Nakajima (2011)). These works clearly relate to our LTM structure in some of their goals and also technically. However, the LTM is a general, natural framework where dynamic sparsity/variable selection occurs via gradual transitions of the underlying time-varying latent processes, applicable to a broad range of models as previously described. The analysis here focuses on developing and evaluating the LTM approach compared to the standard, non-thresholded model, which is the critical comparison of interest in the applied context where the TV-VAR models are accepted standards.

4.2 US macroeconomic data analysis

4.2.1 *Data and priors*

The study analyzes the $m = 3$ time series giving the quarterly inflation rate, unemployment rate and short-term nominal interest rate in the US economy, using data from 1963/Q1 to 2011/Q4, a context of topical interest (Cogley and Sargent 2005; Primiceri 2005; Koop et al. 2009). The inflation rate is the annual percentage change in a chain-weighted GDP price index, the unemployment rate is seasonally adjusted (all workers over 16), and the interest rate is the yield on three-month Treasury bills; see Figure 4.1. The variables are ordered this way in $\mathbf{y}_t$. Prior studies and especially that of Primiceri (2005) focused on the data over the period 1963/Q1–2001/Q3, which this analysis extends to the full data set up to more recent times, 1963/Q1–2011/Q4.

FIGURE 4.1: US macroeconomic time series (indices $\times 100$ for % basis).

The analysis fits and compares predictions using both the new LT-VAR and the standard, non-threshold (NT) TV-VAR models; in each case, the model includes the LT/NT time-varying covariance matrices with stochastic volatility described in Sections 3.1 and 3.2. The analysis uses the specifications of eqns.(3.2)-(3.4), (3.8), (3.12), and (3.13). Four TV-VAR models are considered with the following settings: (1) NT for both $\mathbf{b}_t$ and $\mathbf{a}_t$, (2) NT for $\mathbf{b}_t$ and LT for $\mathbf{a}_t$, (3) LT for $\mathbf{b}_t$ and NT for $\mathbf{a}_t$, and (4) LT for both $\mathbf{b}_t$ and $\mathbf{a}_t$. Model-(1) is almost equivalent to a commonly used TV-VAR of Primiceri (2005); the only difference is that Primiceri (2005) assumes random walk process for the time-varying parameters rather than the stationary AR processes adopted here. In posterior summaries below, Model-(1) is simply referred to as NT-VAR model, and Model-(4) as LT-VAR model.

The analysis uses VAR models with order $p = 3$, determined from the following pre-analysis. Based on the full MCMC analysis of the series over the period 1963/Q1–2000/Q3, out-of-sample forecasts were generated for each of the 4 quarters

in 2000/Q4–2001/Q3. This period was chosen specifically to align with the prior studies of Primiceri (2005) and others. As in the extensive forecasting study below, forecasting uses simulation from the implied posterior predictive distributions. Root mean squared forecast errors (RMSFE) across these out-of-sample forecasts and over all 4 horizons show that both threshold and non-threshold models perform best at $p = 3$, hence this value is used for the comparative analysis. Note that it is becoming increasingly understood in the forecasting literature that standard statistical model selection methods, being driven by model likelihoods that are inherently based on 1-step ahead forecast accuracy, can be suboptimal from the viewpoint of multi-step ahead forecasting (e.g., Xia and Tong 2011). Hence some focus on model choice based on out-of-sample predictive multi-step ahead is recommended when that is a primary goal, as it typically is in macroeconomic studies such as the example here.

For all results reported in this chapter, the analysis takes the following prior components: with v_{*i}^2 denoting the i^{th} diagonal elements of $\mathbf{V}_*$ for $*$ $\in \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}\}$, respectively, $v_{\beta i}^{-2} \sim G(20, 0.01)$, $v_{\alpha i}^{-2} \sim G(2, 0.01)$, $v_{hi}^{-2} \sim G(2, 0.01)$; and for $\boldsymbol{\mu}_h$, $\exp(-\mu_{hi}) \sim G(3, 0.03)$. The other prior specifications for the LTM models and the simulation size are as in Section 2.4.

4.2.2 Forecasting performance and comparisons

First of all, out-of-sample forecast performance is summarized to compare the LT- and NT-VAR models in predicting 1, 2, 3 and 4 quarters ahead using all the data periods. Consider the traditional recursive forecasting format that mirrors the reality facing forecasters; given data from the start of 1963/Q1 up to any quarter t , the analysis refits the full MCMC analysis of each model to define the posterior based on $\mathbf{y}_{1:t}$. The resulting out-of-sample predictive distributions are simulated over the following 4 quarters, times $t + 1 : t + 4$, to generate realized out-of-sample forecast

Table 4.1: Forecasting performance for US macroeconomic data over the 16 years (1996/Q4–2011/Q4): RMSFE for each horizon and each variable. NT and LT refer to the non-threshold and latent threshold models, respectively.

	Horizon (quarters)					
	1			2		
	LT	NT	LT/NT	LT	NT	LT/NT
p	0.264	0.244	1.08	0.393	0.434	0.90
u	0.308	0.336	0.91	0.539	0.560	0.96
r	0.477	0.604	0.78	0.839	1.010	0.83

	Horizon (quarters)					
	3			4		
	LT	NT	LT/NT	LT	NT	LT/NT
p	0.552	0.612	0.90	0.726	0.796	0.91
u	0.840	0.876	0.95	1.121	1.170	0.95
r	1.147	1.391	0.82	1.431	1.763	0.81

errors. Then, the analysis moves ahead one quarter to observe the next observation y_{t+1} , reruns the MCMC based on the updated data series $\mathbf{y}_{1:t+1}$, and forecasts the next 4 quarters $t+2 : t+5$. This is repeated up to time 2010/Q4, generating a series of out-of-sample forecasts for each horizon from each quarter over the 16 years. In addition to direct model comparisons, this allows us to explore performance over periods of very different economic circumstances and so study robustness to time periods of the predictive improvements.

Results are summarized in terms of RMSFE for each variable and each horizon, averaged over the 16 years of quarterly forecasting; see Table 4.1. With the exception of the 1-step forecasts of inflation, the LT-VAR model dominates NT-VAR across all variables and horizons, with larger improvements at longer horizons. Further investigation confirms that the dominance of the LTM is generally consistent across the full period of 16 years, across variables and horizons. The time series generally

evidences reduced predictability in the recessionary years of 2007–2011 under both LT- and NT-VAR models. This is reflected in increased forecast errors for all variables and horizons in this latter period relative to those in earlier years. Nevertheless, the RMSFEs for subsets of years computed to compare with the recessionary period indicates stability in terms of the improvements made under the LTM. Further, the reductions in RMSFE in the table are practically substantial. The one departure from this is the 1-step forecast of inflation, which is due to relatively less accurate point forecasts for inflation under the LTM in just two quarters: 2008/Q4 and 2010/Q2. If just these two times are removed from the full 16 year period of comparison, the LTM dominates. One possible explanation for this is that, having defined a posterior enabling a high degree of thresholding to zero for some of the TV-VAR parameter processes, the LTM may have been relatively “slow” to adapt to somewhat unusual changes in the recession. The NT model has more “free” parameters and so performed somewhat better at those two points, although this is over-whelmed in the broader comparison by being over-parametrized and lacking dynamic parsimony. These specific events do, however, relate to a possible extension of the LTM to allow the latent thresholds themselves to be time-varying; that is, if thresholds decreased in times of increased volatility and change, earlier “zeroed” parameters might more rapidly come back into play.

The results indicate dominance of the LTM in the empirical sense of multi-step forecasting, particularly relevant to policy uses of such models. The improvements in multi-step, out-of-sample predictions are practically significant and further investigation shows that they are robust and sustained over time. To make direct comparisons with prior studies and especially that of Primiceri (2005), forecasting result from data over the period 1963/Q1–2001/Q3 is reported; Table 4.2 shows summary RMSFE measures from four models, for each forecast horizon (averaged

Table 4.2: Forecasting performance for US macroeconomic data over the 5 years (1996/Q4–2001/Q3): RMSFE for each horizon and each variable.

	Model		Horizon (quarters)				Variable		
	$\mathbf{b}_t$	$\mathbf{a}_t$	1	2	3	4	p	u	r
RMSFE									
(1)	NT	NT	0.234	0.336	0.491	0.687	0.391	0.266	0.661
RMSFE relative to Model (1)									
(2)	NT	LT	0.995	0.995	0.996	0.993	0.985	0.980	0.999
(3)	LT	NT	0.957	0.979	0.984	0.916	0.887	0.915	0.970
(4)	LT	LT	0.946	0.912	0.911	0.864	0.777	0.983	0.909

across the three variables) and separately for each variable (averaged across the 4 horizons). Time-varying shrinkage under the LTM evidently contributes to improved predictions, robustly across variables and time, relative to the standard model. The LT- $\mathbf{b}_t$ models (Models 3 and 4) uniformly outperform the NT- $\mathbf{b}_t$ models (Models 1 and 2), indicating that the time-varying shrinkage structure on coefficients evidently contributes to the forecasting performance of time-varying VARs. The LT- $(\mathbf{b}_t, \mathbf{a}_t)$ model (Model 4) performs the best, with 14% improvement from the standard TV-VAR model (Model 1) at four-quarter horizon.

The RMSFE measures in the tables here are posterior means over the MCMC samples; Figure 4.2 shows posterior distributions of the RMSFE, based on the RMSFEs computed every MCMC iteration, which more formally substantiate the uniform consistency of improvements under the LT structure, the increased improvements as forecast horizon increases, and the increased ability of the LT-VAR models to maintain improved predictive ability in the more volatile second period of 2007–2011.

Additional formal model assessments are provided by out-of-sample predictive densities. In comparing LT-VAR with NT-VAR, the log predictive density ratio

(i) 1996/Q1–2001/Q3

(ii) 2006/Q2–2011/Q4

FIGURE 4.2: Posteriors of RMSFE from MCMC analysis of US macroeconomic data: (i) 1996/Q1–2001/Q3 and (ii) 2006/Q2–2011/Q4, using NT-VAR (black) and LT-VAR (light) models. Plots are by variable averaged across forecast horizons (left) and by horizon averaged across variables (right).

(LPDR) for forecasting τ quarters ahead from quarter t is

$$\text{LPDR}_t(\tau) = \log \left\{ \frac{p_{\text{LT-VAR}}(\mathbf{y}_{t+\tau} | \mathbf{y}_{1:t})}{p_{\text{NT-VAR}}(\mathbf{y}_{t+\tau} | \mathbf{y}_{1:t})} \right\},$$

where $p_M(\mathbf{y}_{t+\tau} | \mathbf{y}_{1:t})$ is the predictive density under model M . Relative forecasting accuracy is represented by this evaluated at the observed data, now comparing the uncertainty represented in predictions between the models, as well as point forecasts. Cumulative sums of $\text{LPDR}_t(1)$ define log model marginal likelihoods, and hence posterior probabilities, to formally evaluate evidence for LT-VAR versus NT-VAR via Bayesian testing (e.g., West and Harrison 1997, Chapters 10 & 12). Linking to broader applied interests and policy-related decision perspectives, the $\text{LPDR}_t(\tau)$ provide further insights into relative forecasting ability at longer horizons.

Figure 4.3 plots the log predictive density for LT-VAR and NT-VAR models: the per quarter p_M values are plotted at the times t when forecasts are made, and for each of the 4 forecast horizons τ . Values are computed for the full 5 years (20 quarters) of predictive evaluation in analyses of each of the two time periods. Figure 4.4 shows the log predictive density ratio. The LPDR values are all positive, indicating higher predictive density for all cases; the very high values per quarter indicate very strong support for LT-VAR relative to NT-VAR, confirming the earlier discussions on several other aspects of model fit and predictive performance. Cumulating values over each of the two 20-quarter periods with $\tau = 1$, reported in parentheses for each panel, give values that represent log Bayes factors (log model likelihood ratios) comparing LT-VAR to NT-VAR. For the analysis predicting over 16 years in total, the log Bayes factors is 415.17; this implies very substantial formal evidence in favor of the LT-VAR. Also, Figure 4.3 reflects the increased ability of LT-VAR to improve forecasts with increasing horizon, as the LPDR values are increasing with τ uniformly.

Some additional insights into contributions to the model fit and forecasting edge

FIGURE 4.3: Log predictive density $p_M(\mathbf{y}_{t+\tau}|\mathbf{y}_{1:t})$ for LT-VAR (bold) and NT-VAR (solid) models in analysis of the US time series.

FIGURE 4.4: Log predictive density ratios $LPDR_t(\tau)$ for LT-VAR over NT-VAR models in analysis of the US time series.

generated by the LTM strategies, for both predictive model parameters and volatility matrices, can be seen in the posterior summaries that follow.

4.2.3 *Summaries of posterior inferences*

Figures 4.5-4.7 show some selected summaries from the MCMC analysis of the LT-VAR model with full data time periods. In Figure 4.5, time-varying sparsity is observed for several of the $b_{ij,\ell,t}$ coefficients in $B_{\ell t}$, with estimated values shrinking to zero for some time periods but not others; some coefficients are shrunk to zero over the entire time period analysed. Figure 4.6 plots posterior means and the 95% credible intervals for the Cholesky elements $a_{ij,t}$, with posterior probabilities of $s_{aij,t} = 0$. Here $a_{21,t}$ is entirely shrunk to zero, while the other elements associated with interest rates are roughly 80-90% distinct from zero with stable time variation in their estimated values. As noted above, sparsity for only one covariance element has a lesser impact on improving predictive performance than does inferred dynamic sparsity in $\mathbf{b}_t$.

Constraining $s_{it} = 1$ for all (i, t) and $s_{aij,t} = 1$ for all (i, j, t) reduces the LT-VAR model to the standard NT-VAR model. Across the full set of MCMC iterations, in this analysis and in a more extensive MCMC of millions of iterates, such a state was never generated. This indicates significant lack of support for the NT-VAR model as a special case. This is also directly evident in the summaries of posterior sparsity probabilities of $s_{*,t} = 0$, which very strongly indicate high levels of sparsity within the full LT-VAR model and the irrelevance of a model with no thresholding at all.

Figure 4.7 graphs trajectories of posterior means and 95% credible intervals of the stochastic volatilities h_{it} and $\exp(h_{it}/2)$. Several volatile periods reflect non-systematic shocks hitting the economy. After volatile periods in the 1970s and early 1980s, volatilities decay, reflecting the so-called Great Moderation (Cogley and Sar-

FIGURE 4.5: Posterior probabilities of $s_{it} = 0$ for US macroeconomic data. The corresponding indices of $\mathbf{c}_t$ or $\mathbf{B}_{\ell t}$ are in parentheses.

FIGURE 4.6: Posterior trajectories of $a_{ij,t}$: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

FIGURE 4.7: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$: posterior means (solid) and 95% credible intervals (dotted).

gent 2005; Primiceri 2005). Another high-volatility shock arises in the late 2000s due to recent depression triggered by the US financial crisis. Broad patterns of change over time in inferred residual volatilities concord with prior analyses, while the LT-VAR reduces levels of variation attributed to noise with improved predictions as a result.

A final summary of practical interest relates to impulse response analysis. For models with time-varying coefficients, we need to compute impulse responses repeatedly for each quarter as we move the time window across the sample period. There are several ways to explore impulse responses; here consider the responses to shocks that are innovations to each of the three time series, with shock levels set at the average of the stochastic volatility level for each time series across the time frame. The impulse responses thus summarize the effects of average-sized structural shocks hitting the VAR system.

FIGURE 4.8: Impulse response trajectories for 1-, 2- and 3-year ahead horizons from NT-VAR (upper) and LT-VAR (lower) models. The symbols $\varepsilon_u \uparrow \rightarrow w$ refer to the response of the variable w to a shock to the innovation of variable u .

Figure 4.8 displays posterior means of the impulse response for 1-, 2- and 3-year ahead horizons from both models. The NT-VAR response profiles resemble those of Primiceri (2005); a slight difference arises due to different values of hyperparameters for priors and specification on time-varying parameter process, although the essence of economic interpretation remains unchanged. A focus here is in comparing the NT-VAR and LT-VAR models. First, the trajectory of the response from the LT-VAR model is smoother than that of the NT-VAR. Effective shrinkage in the TV-VAR coefficients and innovation covariance elements leads to less volatile estimates, and this plays a role in smoothing the time-variation of the projected economic dynamics. Second, the sizes of the responses from the LT-VAR model analysis are smaller than those from the NT-VAR analysis, being clearly shrunk towards zero due to the LTM structure. Based on the significant improvements in predictions discussed above, these findings indicate that the impulse responses from the NT-VAR model can represent over-estimates, at least to some extent, that are “corrected” via the induced time-varying shrinkage in the LTM analysis.

4.3 Japanese macroeconomic data analysis

4.3.1 Data

This section analyzes the $m = 3$ time series giving the quarterly inflation rate, national output gap and short-term interest rate gap in the Japanese economy during 1977/Q1–2007/Q4, following previous analyses of related time series data (Nakajima et al. 2010; Nakajima 2011); see Figure 4.9. The inflation rate gap is the log-difference from the previous year of the Consumer Price Index (CPI), excluding volatile components of perishable goods and adjusted for nominal impacts of changes in consumption taxes. The output gap is computed as deviations of real from nomi-

FIGURE 4.9: Japanese macroeconomic time series (indices $\times 100$ for % basis).

nal GDP, defined and provided by the Bank of Japan (BOJ). The interest rate gap is computed as log-deviation of the overnight call rate from its HP-filtered trend. One key and evident feature is that the interest rate gap stays at zero during 1999–2000, fixed by the BOJ zero interest rate policy, and again in 2001–2006 when the BOJ introduced a quantitative easing policy. Iwata and Wu (2006) proposed a constant parameter VAR model with a Tobit-type censored variable to estimate monetary policy effects including the zero interest rate periods. In contrast to that customized model, the LTM structure here offers a global, flexible framework to detecting and adapting to underlying structural changes induced by economic and policy activity, including such zero-value data periods. The same priors are taken as in the previous analysis of the US data.

Based on evaluation of RMSFE across the out-of-sample forecasts for the final 4 quarters, the LT-VAR models perform best when $p = 2$ is assumed, while the non-

Table 4.3: Forecasting performance for Japanese macroeconomic data over the 5 years (2003/Q1–2007/Q4): RMSFE for each horizon.

Model			Horizon (quarters)			
	$\mathbf{b}_t$	$\mathbf{a}_t$	1	2	3	4
RMSFE						
(1)	NT	NT	0.253	0.387	0.525	0.633
RMSFE relative to Model (1)						
(2)	NT	LT	1.016	1.018	1.006	1.008
(3)	LT	NT	0.949	0.721	0.611	0.594
(4)	LT	LT	0.889	0.680	0.592	0.507

threshold TV-VAR models perform best with more elaborate models, taking $p = 4$. The fact that the LTM strategy leads to improved short-term predictions based on reduced dimensional, and hence more parsimonious models is already an indication of the improved fit and statistical efficiency induced by latent thresholding.

4.3.2 Forecasting performance and comparisons

Some summaries of out-of-sample predictive accuracy appear in Table 4.3. RMSFEs are computed for ten different selections of subsets of data: beginning with the sample period from 1977/Q1–2004/Q3, the study fits the model and then forecasts 1- to 4-quarters ahead over 2004/Q4–2005/Q3, and then repeats the analysis rolling ahead 1 quarter at a time. Again, the LT-VAR model dominates, improving RMSFE measures quite substantially at all forecast horizons and quite dramatically so at the longer horizons. Improvement relative to the standard NT-VAR are much more distinctive than that of the US macroeconomic data, providing almost half of RMSFE at the 4-quarter horizon. The Japanese data include zero interest rate periods, therefore the benefit from time-varying shrinkage is perhaps expected to be larger than in the US study.

4.3.3 Summaries of posterior inferences

Figure 4.10 displays the posterior means over time of the time-varying coefficients, as well as the posterior probabilities of $s_{it} = 0$ for the LT-VAR model. Some marked patterns of time-varying sparsity are observed for several coefficients. Figure 4.6 plots the posterior means, the 95% credible intervals of $a_{ij,t}$ and the posterior probabilities of $s_{aij,t} = 0$. Here $a_{21,t}$ has a relatively distinctive shrinkage pattern, with a coefficient that varies slightly and is roughly 50% distinct from zero over most of the time frame, whereas the other two elements, which link directly to the interest rate series, are shrunk to zero with posterior probability close to one across the entire period. Figure 4.7 graphs the posterior means of the stochastic volatility, h_{it} and $\exp(h_{it}/2)$, together with their 95% credible intervals. Several volatile periods are observed for the inflation and interest rates series around 1980. It is quite understandable and appropriate that the volatility of the interest rate gap series is estimated close to zero during the zero interest rate periods.

Figure 4.8 displays posterior means of impulse response for one-, two- and three-year ahead horizons. This comparison fitted both the NT-VAR model and the LT-VAR using $p = 2$ lags. The LT-VAR model provides econometrically reasonable responses: the responses of inflation and output to an interest rate shock shrinks to zero during the zero interest rate periods for all horizons. This is not obtained from the NT-VAR model; there the associated time-varying coefficients and covariance components are fluctuating in non-zero values. The responses from the LT-VAR model indicate that the reactions of short-term interest rates to inflation and output decay after the beginning of the 1990s, and afterwards stay at zero due to the zero interest rates. Since the BOJ terminated the quantitative easing policy in 2006, small responses of interest rates are estimated after 2006. The LT-VAR model also suggests that the responses of inflation decay more dramatically to zero in the 1990's than the

FIGURE 4.10: Posterior means of β_t for Japanese macroeconomic data. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory. The corresponding indices of $\mathbf{c}_t$ or $\mathbf{B}_{\ell t}$ are in parentheses.

FIGURE 4.11: Posterior trajectories of $a_{ij,t}$: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

FIGURE 4.12: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$: posterior means (solid) and 95% credible intervals (dotted).

FIGURE 4.13: Impulse response trajectories for one-, two- and three-year ahead horizons from the NT-VAR (upper) and LT-VAR (lower) models. The symbols $\varepsilon_a \uparrow \rightarrow b$ refer to the response of the variable b to a shock to the innovation of variable a .

VAR model indicates. The responses of output to interest rates and to output itself decline more clearly in the LT-VAR model than in the VAR model. These differences obviously result from the LTM structure, which provides these plausible implications for the Japanese macroeconomic analysis as well as the improved multi-step-ahead predictions already discussed.

In addition, Figure 4.14 reports impulse response with credible intervals computed from posterior draws from NT-VAR and LT-VAR models. Trajectories of posterior median and 75% credible intervals are plotted for responses of the interest rate and inflation to a GDP shock. In both responses, the credible intervals from the LT-VAR model are narrower than that from the NT-VAR; further, the spread of intervals from the LT-VAR is more time-varying. The advantage of LTM structure is obvious particularly in the zero interest rate periods. In addition to the improvements in forecasting performance, these findings confirm that the posterior outputs from the LT-VAR provides more plausible evidences for the macroeconomic dynamics.

4.4 Summary

The extensive model evaluations in this chapter indicate substantial dominance of the latent thresholded model relative to the standard time-varying parameter VAR model that is increasingly used in econometrics. Comparisons in terms of evaluating the non-threshold model as a special case within the MCMC analysis, comparisons of out-of-sample predictions in two very different time periods (both per variable and across several forecast horizons), and evaluation of formal statistical log model likelihood measures all strongly support the use of latent thresholding overlying time-varying VAR and volatility models. Importantly for applications, the LT models are empirically seen to be increasingly preferred by measures of forecasting accuracy at longer horizons, which also supports the view that impulse response

(i) Response of interest rate to GDP shock ($\varepsilon_x \uparrow \rightarrow i$)

(ii) Response of inflation to GDP shock ($\varepsilon_x \uparrow \rightarrow p$)

FIGURE 4.14: Impulse response trajectories with credible intervals from the NT-VAR (left) and LT-VAR (right) models. Posterior median (solid) and 75% credible intervals (dotted).

analyses, as demonstrated in the analyses, will be more robust and reliable under latent thresholding. Specifically, in each of the US and Japanese macroeconomic data, the results revealed that: (i) there is strongly significant evidence for dynamic thresholding when compared to the models with no thresholding; (ii) the LTM analyses yield intuitive and interpretable results, particularly with respect to inferred impulse response functions; and (iii) again relative to the standard, non-thresholded models, the LTM analyses yield practically significant improvements in multi-step, out-of-sample predictions. These predictive improvements are robust and sustained over time, therefore the results are particularly relevant to policy uses of such models. The Japanese study includes a period of time including zero interest rates. The LTM structure naturally adapts to these zero-value periods by eliminating unnecessary fluctuations of time-varying parameters that arise in standard models. This is a nice example of the substantive interpretation of the LTM concept, in parallel to its role as an empirical statistical approach to inducing parsimony in dynamic models. The LT-VAR model provides plausible time-varying impulse response functions that uncover the changes in monetary policy and its effect on the rest of the economy.

Application to Financial Time Series

5.1 Context and literature

In the context of multivariate time series in financial econometrics, dynamic factor models have become standard tools to analyze dynamic comovements and relational structure with multivariate stochastic volatility (e.g., Aguilar et al. 1999; Pitt and Shephard 1999; Aguilar and West 2000; 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 and Carvalho 2007; Carvalho et al. 2011). As the number of assets in a time series study increases, the number of time-varying parameters defining factor loadings and innovations/residual volatility matrices grows quickly. Parsimony, practicality and predictive performance all argue in favor of reducing the number of parameters by constraining some to zero. With the LTM framework, this chapter illustrates the LTF models developed in Section 3.3 for four applications to financial time series data.

Central applied aspects of this chapter concern detailed study of the LTF models

in discussion of posterior estimates and portfolio allocation analyses, assessing the benefits they can provide in realistic decision making contexts across a range of portfolio strategies. This portfolio analysis follows previous Bayesian approaches to model evaluation using sequential portfolio decisions (Quintana 1992; Putnum and Quintana 1994; Quintana and Putnum 1996; Aguilar and West 2000; Carvalho and West 2007; Wang and West 2009; Carvalho et al. 2011). A main focus is the impact of the LTM structure on forecast accuracy relative to standard dynamic factor models, considering how the dynamic sparsity/shrinkage in time-varying factor loadings plays a relevant role in investment experiments.

The analysis uses traditional Bayesian mean-variance optimization (Markowitz 1959) subject to constraints. After observing the closing rates on the business day $t - 1$ the one-step ahead forecast mean vector and variance matrix of $\mathbf{y}_t$ are denoted by $\bar{\mathbf{y}}_t$ and $\bar{\mathbf{Q}}_t$; the implied precision matrix is then $\bar{\mathbf{K}}_t = \bar{\mathbf{Q}}_t^{-1}$. These are computed via the MCMC based on draws from one-step ahead predictive posterior distributions using the available data, $\mathbf{y}_{1:t-1}$. Investments are reallocated according to a vector of portfolio weights $\mathbf{w}_t$ optimized by a specific allocation rule, described below. The realized portfolio return at time t is $\mathbf{w}_t' \mathbf{y}_t$. The portfolio is reallocated on each business day based on one-step ahead forecasting computed via the MCMC given updated data. The total sum invested on each business day is fixed by restricting $\mathbf{w}_t' \mathbf{1} = 1$. Given a scalar daily return target r_t , we optimize the portfolio weights $\mathbf{w}_t$ by minimizing the one-step ahead variance of returns among the restricted portfolios whose one-step ahead expectation is equal to a target r_t . Specifically, at time t , an ex-ante portfolio variance $\mathbf{w}_t' \bar{\mathbf{Q}}_t \mathbf{w}_t$ is minimized subject to $\mathbf{w}_t' \bar{\mathbf{y}}_t = r_t$, and $\mathbf{w}_t' \mathbf{1} = 1$. The solution is

$$\mathbf{w}_t^*(r_t) = \bar{\mathbf{K}}_t(a_t \bar{\mathbf{y}}_t + b_t \mathbf{1}),$$

where

$$\begin{aligned} a_t &= \mathbf{1}' \bar{\mathbf{K}}_t \mathbf{e}_t, \quad b_t = -\bar{\mathbf{y}}_t' \bar{\mathbf{K}}_t \mathbf{e}_t, \quad \mathbf{e}_t = (\mathbf{1}r_t - \bar{\mathbf{y}}_t)/d_t, \\ d_t &= (\mathbf{1}' \bar{\mathbf{K}}_t \mathbf{1})(\bar{\mathbf{y}}_t' \bar{\mathbf{K}}_t \bar{\mathbf{y}}_t) - (\mathbf{1}' \bar{\mathbf{K}}_t \bar{\mathbf{y}}_t)^2. \end{aligned}$$

The analysis also considers the target-free minimum-variance portfolio whose solution is simply given by

$$\mathbf{w}_t^\diamond = \bar{\mathbf{K}}_t \mathbf{1} / (\mathbf{1}' \bar{\mathbf{K}}_t \mathbf{1}).$$

This portfolio allocation study implicitly assumes that we can freely reallocate the resources to arbitrary long or short positions across the currencies without any transaction cost.

5.2 Identification

Some comments on general questions of identification are in order. First, the overall latent threshold model strategy inherently promotes data-based shrinkage (completely to zero) of multiple factor loadings when it matters. Coupled with the assumed upper triangular form of the factor loadings matrix $\mathbf{B}_t$ to “name” and order factors according to the chosen first q time series variables, the results are inferences on, typically, quite sparse structures. The triangular form ensures mathematical identification, in terms of lack of an ability to “rotate” factors coupled with a restricted number of free elements in $\mathbf{B}_t$ (Aguilar and West 2000; Lopes and West 2004). Thresholding then typically reduces the number of non-zero elements more dramatically; in this sense, the model is automatically self-identifying.

One specific additional consideration is the use of models in which both factor processes and factor loadings are time-varying, and whether this introduces the potential for new identification questions. There are two points to note. First, again,

thresholding matters: a thresholded value of a factor loading wipes out the contribution of the factor itself over the period of time it is below threshold; so the actual values of the underlying latent factor process are irrelevant over that period. Second, and most importantly, careful specification of informative priors over the parameters of the time series models for the factor loadings, that is, the latent AR(1) process models of eqn.(3.17), provides critical control over where the resulting posteriors place mass. Models with rather vague priors that would allow for very widely fluctuating patterns of change over time in the factor loadings are irrelevant and dangerous, as they could engender models with wild variation in loadings and relatively insignificant factor processes themselves, distort inferences and certainly lead to weak identification problems. The same general comment applies to all hyper-parameters, of course, but is most particularly key in connection with the factor loadings model components. In reflection of this, the priors used in the analysis here favor highly persistent stationary AR(1) processes for β_{it} , strongly limiting the global fluctuations in loadings relative to potential fluctuation in the factors. This is attained with priors for the $\phi_{\beta i}$ of eqn.(3.17) that concentrate close to one, and with priors for the innovations variances $v_{\beta i}$ that strongly favor small values, again in the context of expected ranges of variation in the factor processes as specified in corresponding priors for their hyper-parameters.

5.3 Exchange rate data analysis – Major currencies

5.3.1 Data and priors

The first LTF analysis uses $m = 6$ daily international currency exchange rates relative to US dollar over a time period of 980 business days beginning in January 2006 and ending in December 2009. The returns are computed as $y_{it} = 100(p_{it}/p_{i,t-1} - 1)$,

Table 5.1: 6 international major currencies relative to US dollar.

1	GBP	British Pound Sterling
2	EUR	Euro
3	JPY	Japanese Yen
4	CAD	Canadian Dollar
5	AUD	Australian Dollar
6	CHF	Swiss Franc

where p_{it} denotes the daily closing spot rate. The series are listed in Table 5.1; the exchange rates and returns are plotted in Figure 5.1. The most striking feature in these sample periods is a quite volatile and turbulent movement triggered by the financial crisis around 2008; there are some superficial correlated flows and shifts among the series. Initial exploratory analysis using the univariate stochastic volatility fit to each series separately is summarized in Figure 5.2. There exist marked region-specific patterns. The peaks of stochastic volatility seem to coincide in 2008, but there exist several spikes common only to two or three currencies. From these exploratory analysis, the currencies are ordered in the $\mathbf{y}_t$ vectors as in Table 5.1.

This study considers a special case of the general LTF form based on eqns.(3.14)-(3.24) in which there are no covariates or no correlation in factor innovation; i.e., $\mathbf{x}_t = \mathbf{0}$, $\mathbf{Z}_t = \mathbf{O}$, and $\mathbf{A}_t = \mathbf{I}$, for all t . The following priors are used: $1/v_{ci} \sim G(40, 0.005)$, $1/v_{\beta i} \sim G(40, 0.02)$, $1/v_{*i} \sim G(2, 0.01)$ for $* \in \{h, \delta\}$; $(\phi_{*i} + 1)/2 \sim B(20, 1.5)$ for each $* \in \{c, \beta, h, \lambda\}$, and $(g_i + 1)/2 \sim B(1, 1)$; for each $* \in \{c, \beta\}$, $\mu_{*i} \sim N(0, 1)$; and for each $* \in \{h, \lambda\}$, $\exp(-\mu_{*i}) \sim G(3, 0.03)$. The MCMC analysis was run for a burn-in period of 20,000 samples prior to saving the following MCMC sample of size 100,000 for summary posterior inferences.

Determination of the number of factors q has been a challenging issue in the literature (e.g. Lopes and West 2004). The analysis here fit the LTF models to the

(i) Exchange rates

(ii) Returns

FIGURE 5.1: Daily exchange rates and returns for major currency data.

FIGURE 5.2: Estimated trajectories of univariate stochastic volatility ($\sigma_{it} = \exp(\delta_{it}/2)$) fit to each series separately.

FX data and repeated the analysis across models with different number of factors $q = 1, 2, 3$, or 4. Each of the resulting models was evaluated by forecasting over the final 5 business days using the first 930 observations, and this five-step ahead forecasting were repeated based on the first $930+5j$ observations for $j = 1, \dots, 9$, to obtain the 50-day forecasts. Based on the RMSFE for these out-of-sample forecasts, the LTF performs substantially better at $q = 3$ factors; this suggests a fourth factor is redundant while three are required. The study uses $q = 3$ and this issue is revisited below with some key aspects from the analysis of a model with $q = 4$.

5.3.2 Summaries of posterior inferences

Figure 5.3 displays posterior estimates of the latent time-varying factor loadings, as well as the posterior probabilities of $s_{it} = 0$ from the LTF model. The estimated latent process exhibits considerable time variation for several loadings; this cannot be captured by a constant-loading factor model. In particular, JPY-Factor1 and AUD-

FIGURE 5.3: Posterior means (solid) and 95% credible intervals (dotted) of time-varying factor loadings. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory.

Factor3 are estimated positive in 2006, but become negative after 2007, reflecting a shift of correlated dynamics among the FX series. The trajectories of posterior means of several loadings also evidently change over time. Time-varying sparsity is observed for the loadings of JPY-Factor1, AUD-Factor3 and CHF-Factor3, with estimated values shrinking to zero for some time periods but not others, whereas other loadings such as CAD-Factor2 and CHF-Factor3 are totally selected out over the entire time-frame.

The interesting observation of the shrinkage pattern of JPY-Factor1 in 2008 implies that the JPY returns turned to be negatively correlated with Factor1. The loadings on Factor1 are all positive in the currencies except JPY. In 2008, the financial crisis led investors to reduce their size of US dollar short position and the currencies in our analysis except JPY depreciated against US dollar, whereas the JPY temporarily appreciated.

Figure 5.4 shows the proportion of variation of the time series explained by each factor; the time variation in the contributions is evident. By construction, Factor1 is a GBP-leading factor, although its explanatory ratio decreases from 2006 to 2009, and there are a few downward spots in the crisis period, which indicates idiosyncratic variation of the GBP. Factor2 is recognized as a European factor that mainly explains the variation of EUR and CHF. The variation of JPY is of interest; Factor1 primarily explains its fluctuation in 2006, but later its role is shifted to Factor3, namely a JPY-leading factor. Clear shrinkage of the contribution in JPY-Factor1, CAD-Factor2, and CAD-Factor3 is detected by the LTF structure as observed in Figure 5.3; this data-induced shrinkage eliminates unnecessary fluctuations in what are negligible but non-zero values in time-varying loadings possibly found from a non-thresholded model, and leads to efficient discrimination and contextual interpretation of the factors.

FIGURE 5.4: Time series trajectories of proportion of variances explained by factors.

FIGURE 5.5: Left panels: posterior means of factors f_{it} . Right panels: posterior means (solid) and 95% credible intervals (dotted) of factor stochastic volatility $\psi_{it} = \exp(\lambda_{it}/2)$.

Figure 5.5 plots the posterior estimates of time trajectories of the factors and their innovation volatilities. The trajectories and peak of the stochastic volatility remarkably differ among the factors. The stochastic volatility of Factor1 has the peak around September 2008 corresponding to the major crash of the market, whereas that of Factor2 has its peak around December 2008, when EUR and CHF returns exhibit relatively higher volatility in reaction to the US Senate's rejection of the financial bailout for the automotive industry. The stochastic volatility of Factor1 exhibits a steeper hike towards the peak and a more moderate diminishing afterwards compared to that of Factor2. The result also shows that the posteriors for the AR(1) g_i parameters place substantial mass around zero, indicating lack of predictability in the latent factor processes, i.e., that the factor contributions are effectively unrelated over time, consistent with assumptions of factor volatility modeling of FX returns in

FIGURE 5.6: Posterior means of series-specific stochastic volatility ($\sigma_{it} = \exp(h_{it}/2)$).

prior work.

Figure 5.6 graphs posterior means of series-specific stochastic volatility, $\sigma_{it} = \exp(h_{it}/2)$, that is, volatility elements excluding the common factors, exhibiting quite a different picture from Figure 5.2. GBP has a unique hike of volatility in the second half of 2008 and CHF has several spikes in 2008 and 2009, which are not explained by the factors. Because most of the fluctuation in EUR returns are explained by Factor1 and Factor2 as can be seen in Figure 5.4, the remaining series-specific shocks are relatively smaller than the others as shown in Figure 5.6.

As mentioned above, the number of factors are determined as $q = 3$ based on out-of-sample, multi-day ahead forecasting performance. However, the LTF model provides some aspects of analysis results using $q = 4$ factors. As can be seen in Figure 5.7, the estimated trajectory of Factor4 is clearly shrunk with significantly smaller stochastic volatility than the other three relevant factors. Figure 5.8 shows

FIGURE 5.7: Results of FX analysis in model with $q = 4$ factors. Left panels: posterior means of factors f_{it} . Right panels: posterior means (solid) and 95% credible intervals (dotted) of factor stochastic volatility $\psi_{it} = \exp(\lambda_{it}/2)$.

results of time-varying loadings and sparsity, indicating that the loadings of Factor4 are mostly shrunk in the entire periods. This finding implies that the LTF structure can facilitate data-driven model search and selection of factor models via shrinkage of time-varying loadings.

Some experiments with various orderings of the FX series in $\mathbf{y}_t$ vectors are that the ordering reported here yields more reasonable factor estimates with plausible econometric interpretations than other ordering. The study continues with the model as specified here in the following portfolio analysis.

5.3.3 Portfolio analysis

The study reanalyzes the FX time series restricting to only the first $T = 880$ observations. The remaining 100 business days are then used for forecasting, portfolio

FIGURE 5.8: Results of FX analysis in model with $q = 4$ factors. Posterior means (solid) and 95% credible intervals (dotted) of time-varying factor loadings. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory.

reallocations and model comparisons. In addition to one-step ahead prediction, the study also examines the same analysis strategy but now focuses on five-step ahead forecasting and portfolio revisions. Every five business days, the posterior predictive distribution of five-step horizons, $(\mathbf{y}_t, \mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+4})$ is computed via MCMC based on the available data $\mathbf{y}_{1:t-1}$. This experiment assumes a possible situation that investors allocate their resource every business day based on weekly-updated forecasts.

For robust experiment design, the following three models from the class of LTF models are considered:

- **LM-AF**: Local autoregressive means $\mathbf{c}_t$ and autoregressive factors $\mathbf{f}_t$;
- **LM-IF**: Local autoregressive means $\mathbf{c}_t$ and time-independent factors $\mathbf{f}_t$; i.e., $\mathbf{G} = \mathbf{O}$;
- **CM-IF**: Constant levels $\mathbf{c}_t \equiv \mathbf{c}$ and time-independent factors.

For each model of these three specifications, our study analyzed two variants: the LT (latent threshold) and NT (non-threshold) versions, giving a total of six competing models analyzed via the forecasting and portfolio strategy over the final 100 business days. For the CM-IF model, an additional prior is assumed: $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I})$. The other prior specifications and simulation size are as in the previous analysis. Each model is repeatedly assessed using portfolios based on a range of daily target returns of $r_t = 0.02\%$, 0.04% , and 0.06% , corresponding to monthly (roughly 20 day) return of approximately 0.4% , 0.8% and 1.2% , respectively.

Table 5.2 reports cumulative returns of the portfolios resulting from the sequential investment over 100 business days. It is evident that the use of LT structure remarkably dominates the NT models regardless of other aspects of model specification or the portfolio allocation rules. Figure 5.9 displays the cumulative returns across time

Table 5.2: Cumulative returns over 100 business days for major currency data.

	LM-AF		LM-IF		CM-IF	
	NT	LT	NT	LT	NT	LT
(i) One-step ahead forecasts						
Target return: $r_t = 0.02$	-1.08	1.25	1.18	3.11	-3.03	2.06
0.04	-2.20	1.52	-0.33	2.55	-4.85	1.61
0.06	-3.31	1.78	-1.85	1.99	-6.67	1.15
Target-free	-1.42	2.11	-1.26	0.98	-1.12	-0.04
(ii) Five-step ahead forecasts						
Target return: $r_t = 0.02$	-1.26	2.09	-1.54	0.95	-0.55	1.20
0.04	-0.96	1.88	-1.33	1.03	-3.12	0.82
0.06	-0.66	1.68	-1.12	1.11	-5.69	0.45
Target-free	-0.98	0.52	-0.55	1.25	-0.53	0.50

periods from one-step ahead forecasting. LT models clearly outperform NT models and their difference is larger in the mean-target strategy of $r_t = 0.06\%$ than in the target-free minimum-variance portfolio. A sudden loss in the CM-IF-NT model in the late periods of the mean-target analysis indicates weakness of the constant mean, independent factor model; it cannot fully react to a rapid change of market circumstances. In contrast, it is remarkable that CM-IF-LT model avoids such a significant drop due to reasonably flexible shrinkage in the loadings matrix that also reduces uncertainty about the time-varying parameters, resulting in reduced forecast variances as well as improved forecast performance from short-term adaptability.

Additional insights into comparative performance come from examining levels of portfolio risk. From one-step ahead forecasting using the LM-AF models, Figure 5.10 plots ex-ante portfolio standard deviations $(\mathbf{w}_t' \bar{\mathbf{Q}}_t \mathbf{w}_t)^{1/2}$ at the optimizing weights $\mathbf{w}_t$. LT models clearly yield smaller variances of optimizing portfolios than NT models. The risk ratios based on these portfolio standard deviations for the NT model relative to the LT model are displayed in Figure 5.10; LT models exhibit roughly 20% lower

FIGURE 5.9: Cumulative returns over 100 business days based on one-step ahead forecasting for (a) portfolio under the target returns of 0.06% and (b) target-free minimum-variance portfolio.

Table 5.3: Portfolio risk ratio: Ex-ante ratio of portfolio standard deviation under the NT models relative to the LT models, averaged across 100 business days.

	LM-AF	LM-IF	CM-IF
(i) One-step ahead forecasts			
Target return: $r_t = 0.02$	1.033	1.007	1.185
0.04	1.149	1.187	1.062
0.06	1.232	1.407	0.969
Target-free	1.234	1.205	1.193
(ii) Five-step ahead forecasts			
Target return: $r_t = 0.02$	1.067	1.027	1.178
0.04	1.158	1.102	1.058
0.06	1.223	1.150	0.979
Target-free	1.249	1.224	1.172

FIGURE 5.10: Portfolio risk: Ex-ante portfolio standard deviation, $(\mathbf{w}_t' \bar{\mathbf{Q}}_t \mathbf{w}_t)^{1/2}$ based on one-step ahead forecasting of LM-AF models for (a) the portfolio under the target return of $r_t = 0.06\%$ (top), and (b) the target-free minimum-variance portfolio (bottom). Time trajectory (left panels) and histogram of the risk ratio for NT model relative to LT model (right panels). The vertical line in the histogram refers to the ratio of one.

risk in their portfolios. Table 5.3 reports the average risk ratios from the full analyses, indicating that the portfolio variances are likely smaller in using LT models, and that the ratios tend to be larger in rather more flexible specifications, LM-AF and LM-IF models, than in CM-IF models. Overall, this implies that the LTF structure plays a considerably beneficial role in reducing predicted investment risk in the portfolio allocations.

One further aspect of interest is the trajectories of optimized portfolio weights based on one-step ahead and five-step ahead forecasting, displayed in Figure 5.11 and 5.12, respectively. In the NT model, the allocation is mainly led by a long position

FIGURE 5.11: Trajectories of portfolio weights based on one-step ahead forecasts of the LM-AF models for (a) the portfolio under the target return of $r_t = 0.06\%$ (upper), and (b) the target-free minimum-variance portfolio (lower).

FIGURE 5.12: Trajectories of portfolio weights based on five-step ahead forecasts of the LM-AF models for (a) the portfolio under the target return of $r_t = 0.06\%$ (upper), and (b) the target-free minimum-variance portfolio (lower).

on EUR and a short position on CHF. This reflects quite a high correlation between these two currencies in the estimated forecast variance matrix. By contrast, LT models provide milder degrees of correlation between those two currencies, yielding a smaller proportion of resources allocated to the entire currencies. The lower degree of variation in portfolio weights resulting from the LT models leads to more stable portfolios, which would also reduce transaction costs in reality.

5.3.4 Diagnostics

The model analysis and mean-variance portfolio strategy rely on a number of modeling assumptions, among which is the assumed conditional normality of the error terms; that for the observation errors $\boldsymbol{\nu}_t$ and for the latent factor innovations $\boldsymbol{\varepsilon}_t$ in eqns.(3.14) and (3.15), respectively. Non-normality would suggest consideration of other forms of utility function in the portfolio strategy, such as power utility functions that go beyond the mean-variance based forms to weigh risks in terms of other distributional aspects.

Consistency of the data with the assumed normal forms is examined using a number of traditional methods applied to posterior samples of the standardised errors: $\tilde{\nu}_{it} = \nu_{it}/\sigma_{it}$, and $\tilde{\varepsilon}_{jt} = \varepsilon_{jt}/\psi_{jt}$ for $i = 1 : m$, $j = 1 : k$. At each MCMC iteration the current sampled values of each $\tilde{\nu}_{it}$ and $\tilde{\varepsilon}_{jt}$ are saved, building up a full posterior sample of all $(m + k) \times T$ error terms. Using these, the study here utilized qq-plots and two traditional non-parametric and parametric tests. Figure 5.13 graphs qq-plots of the standardised errors against the standard normal, with approximate posterior 99% credible intervals, using some of the residual values at one randomly selected MCMC iterate. This is very typical; viewing many such plots across the MCMC iterates, we see close concordance with normality as a general rule, with only a few posterior samples observed showing slight deviations from the theoretical quantiles

FIGURE 5.13: QQ-plots for standardized errors $\tilde{\nu}_{it}$ and $\tilde{\varepsilon}_{jt}$ against the standard normal (theoretical quantiles on the horizontal axis) with approximate posterior 99% credible intervals (dashed lines). Here the residuals used represent just one random selection from the full posterior sample of residuals, as an example of a typical case.

in the tails. In simply viewing hundreds of repeat plots, in no cases do we see significant deviation of sampled residuals outside the intervals. These findings provide initial, exploratory graphical support for the assumed normality.

The study considers more formal and quantify the assessment of normality using traditional test statistics, namely the non-parametric Pearson's chi-square goodness-of-fit test and the Shapiro-Wilk parametric test. For each of the two sets of residuals sampled from their posterior at each MCMC iterate, notional p -values are computed for each of these statistics. This yields a posterior distribution for the p -values as one global numerical summary of extent to which the data support the normality assumed. Figure 5.14 shows box plots of these posteriors for each series; it is clear that all the distributions are spread and favor larger values, strongly confirming lack of evidence against normality. Importantly, this applies to both the observation residual/error terms ν_{it} and the innovation terms ε_{jt} impacting on the latent factor

FIGURE 5.14: Box-plots for posterior distributions of p -values in the normality tests using full posterior samples of standardised errors $\tilde{\nu}_{it}$ and $\tilde{\varepsilon}_{jt}$: (1) Pearson’s chi-square normality test, and (2) Shapiro-Wilk parametric normality test. F_i refers to Factor i .

processes.

This concordance with the normality assumption is not, perhaps, so surprising given that the overall LTF model allows for volatility patterns overlaying the residuals, and the models have dynamic adaptability in the latent factor components. The results indicate that the model assumptions checked seem to be quite reasonable based on these detailed diagnostics.

5.4 Exchange rate and commodity price data analysis

5.4.1 Data and priors

A second study extends the FX time series above with 3 additional currencies as listed in Table 5.4 and explores analysis now including a dynamic regression component based on commodity prices, oil and gold. This analysis now directly models the logged prices of currencies in US dollars, rather than the returns analyzed above;

Table 5.4: Additional international currencies relative to US dollar.

7	NOK	Norwegian Krone
8	NZD	New Zealand Dollar
9	ZAR	South African Rand

that is, $\mathbf{y}_t$ is the vector of (natural logs of) exchange rates at 4:30pm on day t . As covariates, the analysis uses prices of Brent Crude Oil futures and Comex Gold futures. The LTF model is employed with dynamic regression and autoregressive latent factors, that is, eqns.(3.14)-(3.24) in which here the specification includes $\mathbf{x}_t$ and $\mathbf{Z}_t$ for the covariates, though it still assumes $\mathbf{A}_t = \mathbf{I}$, for all t . The analysis defines $\mathbf{x}_t$ as the 2-vector of logged commodity prices on day $t - 1$, and $\mathbf{y}_t$ as the 9-vector of logged exchange rates on day t . Thus the model has $\mathbf{m} = 9$ daily currency exchange rates, $\ell = 2$ covariates and $q = 3$ latent factor processes with constant means ($\mathbf{c}_t \equiv \mathbf{c}$) in view of the predictive component of the model. The time period is $T = 125$ business days from the beginning of January to the end of June in 2009, a turbulent period for the financial markets. Priors are specified as in the previous analysis with additional priors assumed here: $1/v_{\zeta i} \sim G(40, 0.005)$, $(\phi_{\zeta i} + 1)/2 \sim B(20, 1.5)$ and $\mu_{\zeta i} \sim N(0, 1)$ for $i = 1 : 18$.

5.4.2 Posterior inferences and portfolio analysis

Figure 5.15 plots the trajectories of posterior probabilities of $s_{it} = 0$ for the factor loadings. Both global and local sparsity patterns found across the currencies indicate adequately distinguished factors oriented by GBP, EUR and JPY. Figure 5.16 shows the posterior means of trajectories of the factors and posterior distributions of the latent factor AR parameters g_i . The latter indicate values consistent with highly persistent AR(1) processes for the factors. This suggests that modeling exchange

FIGURE 5.15: Posterior probability of $s_{it} = 0$ associated with the loadings for exchange rate and commodity data.

FIGURE 5.16: Left panels: posterior means of factors f_{it} and log of commodity prices (x -axis refers to the month in 2009). Right panels: estimated posterior distribution of the g_i .

rates themselves can yield an utility in short-term forecasting; taking returns from the prices may lose some information about short-term changes in patterns of covariation.

Figure 5.17 graphs posterior estimates of the dynamic regression coefficients ζ_t and the posterior probabilities of $s_{zit} = 0$. For several currencies, the posterior means of the ζ_{it} are positive throughout the sample period; consistent with context, positive movements in oil and gold tend to predict marginal appreciation of these currencies against the US dollar, all other things being equal. The inferences further indicate effective whole-sequence sparsity for the major currencies GBP, EUR and JPY; the dynamic regressions on commodities are playing no role in short-term prediction of these prices, with inferred coefficients zero or close to zero across the entire time frame. This indicates that there is little room for commodity prices to

FIGURE 5.17: Posterior means (solid) and 95% credible intervals (dotted) of time-varying regression coefficients ζ_t for exchange rate and commodity data. Posterior probabilities of $s_{zit} = 0$ are plotted below each trajectory.

explain meaningful residual changes in these currencies in the context of the dynamic factors estimated, which is further interpreted by examining inferences on the factors themselves. Figure 5.16 plots the trajectories of the commodity prices as well as estimated factors. We see that the posterior mean trajectories of Factor1 and Factor2 are primarily correlated with the fluctuations of oil price; in fact these correlations are high as 0.92 for Factor1 and 0.84 for Factor2. The correlation between the factors and gold prices is relatively mild, around 0.2. The GBP and EUR leading factors essentially capture the predictive information in oil price returns and, as a result, the latent threshold mechanism shrinks dynamic regression effects very substantially.

Finally, some comments are provided on a further, detailed comparison using portfolio analysis, focused on comparing NT and LT models. The portfolio allocation rules and estimation strategy are as in Section 5.3.3 (see also Section 5.1). One-day ahead forecasts and the optimal portfolios are computed over 25 business days beginning with the first $T = 100$ observations. The study assumes that the commodity returns in $\mathbf{x}_{n+1}$ are available when the model forecasts the one-day ahead exchange rate $\mathbf{y}_{n+1}$. Figure 5.18 graphs the resulting cumulative returns from this using both the NT and LT models; here the analysis uses a target return of $r_t = 0.06\%$ and compared also based on the target-free minimum-variance strategy. It is evident that the LT model dominates the NT model for both the strategies. The most striking evidence is that the optimal weights allocated to the 9 FX currencies are considerably shrunk towards zero in the LT model as can be seen in Figure 5.19. This effective shrinkage yields higher performance of the portfolios under the LTM structure.

FIGURE 5.18: Cumulative returns over 25 business days based on one-step ahead forecasting for (a) portfolio under the target returns of 0.06% and (b) target-free minimum-variance portfolio using exchange rate and commodity data.

FIGURE 5.19: Trajectories of portfolio weights based on the portfolio under the target return of $r_t = 0.06\%$ for exchange rate and commodity data.

5.5 Exchange rate and stock price data analysis

5.5.1 Data and priors

A third data set serves to demonstrate a substantial, higher-dimensional application. The study analyzes $m = 40$ financial variables: daily FX time series of 20 currencies relative to the US dollar and stock price indices from 20 countries; see Table 5.5. The returns are computed over a time period of 1,000 business days beginning in January 2008 and ending in October 2011. This data set includes severely high volatility periods of the financial crisis around 2008 and the European sovereign debt crisis from 2010 to 2011. There exist similar volatility dynamics across countries as well as several regionally or industrially common fluctuations. Exploratory analysis suggests an initial model using $q = 7$ factors, ordering the first seven variables in $\mathbf{y}_t$ as GBP, US-stock, EUR, JPY, IDR, Germany-stock, and Brazil-stock. In addition to GBP, EUR, and JPY used in the previous analysis, IDR (Indonesian Rupiah) is selected to capture exchange rate flows of emerging countries against the US dollar. US-stock is intended to represent a global stock price measure, Germany-stock describes common stock movements of European countries, and Brazil-stock traces stock markets in the emerging countries. The analysis uses the same LTF specification and priors as in Section 5.3.

5.5.2 Posterior inferences and portfolio analysis

Figure 5.20 plots the trajectories of estimated latent factors and their innovation volatilities. Factor2 and Factor6 have levels of volatility higher than other factors as they are led by stock prices of US and Germany, respectively, and the latter exhibit higher levels of variability than the exchange rate series. As observed in the first example, all the factor stochastic volatilities have peaks around the last three months of 2008 due to the financial crisis. The volatilities of GBP-leading Factor1

Table 5.5: 20 international currencies for FX and 20 countries for stock price index.

FX			Stock price	
1	GBP	British Pound Sterling	1	US
2	EUR	Euro	2	Germany
3	JPY	Japanese Yen	3	Brazil
4	IDR	Indonesian Rupiah	4	Canada
5	CAD	Canadian Dollar	5	Mexico
6	AUD	Australian Dollar	6	UK
7	NZD	New Zealand Dollar	7	France
8	CHF	Swiss Franc	8	Spain
9	NOK	Norwegian Krone	9	Italy
10	SEK	Swedish Krona	10	Netherlands
11	RUB	Russian Ruble	11	Sweden
12	INR	Indian Rupee	12	Swiss
13	PHP	Philippine Peso	13	Japan
14	SGD	Singapore Dollar	14	China
15	KRW	South Korean Won	15	Hong Kong
16	TWD	Taiwanese Dollar	16	Taiwan
17	THB	Thai Baht	17	Korea
18	ZAR	South African Rand	18	India
19	BRL	Brazilian Real	19	Russia
20	CLP	Chilean Peso	20	Australia

and JPY-leading Factor4 decline after that crisis to the end of the sample, while the EUR-leading Factor3 shows considerable increases in volatility around mid-2010 and during 2011 corresponding to the European sovereign debt crisis triggered by the Greek government debt crisis arising in April 2010. At that time, The US-stock-leading Factor2, IDR-leading Factor5, and Germany-stock-leading Factor6 commonly capture a sudden hike in volatility. These factors also exhibit another significantly volatile period beginning in August 2011 when Standard & Poor's cut the long-term US credit rating from triple-A to AA-plus.

Figure 5.21 shows the trajectories of estimated posterior probabilities of $s_{it} = 0$ for selected loadings. The first row indicates that the US stock market (Factor2) is

FIGURE 5.20: Posterior means of factors f_{it} , posterior means (solid) and 95% credible intervals (dotted) of factor stochastic volatility $\psi_{it} = \exp(\lambda_{it}/2)$ for FX and stock price data.

relevant to describing the Brazilian stock market in addition to the leading-factor itself (Factor7). A drop of shrinkage probability on Brazil-Factor5 from the beginning of 2009 presumably reflects the fact that exchange rate fluctuations in emerging countries led by IDR affect some proportion of Brazilian stock market volatility. Several factors associated with NZD (in the second row of the figure) show clearly time-varying shrinkage patterns; in contrast, the model suggests the Spain stock price index (in the third row) is primarily explained by the Germany-stock-leading Factor6. The shrinkage pattern of the India stock price index (in the forth row) exhibits some temporal changes on Factor2 and Factor5, while German-stock Factor6 and Brazil-stock Factor7 are evidently relevant in explaining the volatility dynamics.

FIGURE 5.21: Posterior probability of $s_{it} = 0$ associated with selected loadings. Brazil-stock, New Zealand Dollar FX, Spain-stock, and India-stock.

FIGURE 5.22: Heat map of posterior probabilities of $s_{it} = 0$.

Table 5.6: Cumulative returns (%) and average transaction (in weights per day) over 50 business days.

	NT	LT
(1) Cumulative returns		
Target return: $r_t = 0.02$	2.20	2.50
0.04	2.03	2.29
0.06	1.86	2.08
Target-free	2.39	2.73
(2) Average transaction		
Target return: $r_t = 0.02$	0.27	0.18
0.04	0.38	0.25
0.06	0.51	0.36
Target-free	0.16	0.15

A variance decomposition analysis (not graphed here) also indicates that each factor evidently well identifies the characteristics of commonly observed fluctuations across the high-dimensional financial responses. In addition, Figure 5.22 graphs a heat map of posterior probabilities of $s_{it} = 0$ for all i and t . To make the image clear, the rows corresponding to factor loadings have been ordered to highlight patterns of difference. Red areas show high levels of LT shrinkage, while blue areas imply relevant factor loadings beyond the threshold. There exist red bands indicating whole-sequence shrinkage as well as some patterns of temporal shrinkage where the color is clearly changing through time.

In line with the preceding examples, further analysis evaluated the LTF models based on forecasting in portfolio allocation experiments. Forecast means and covariance matrices of the 40 dimensional variables from one-step to five-step ahead were computed every five business days with the first subset of data up to time $n = 950$. The optimal weights of the portfolio strategy were obtained for target returns of $r_t = 0.02\%$, 0.04% , and 0.06% , as well as the target-free minimum variance strategy.

Table 5.6 reports cumulative returns of the portfolios resulting from the sequential investment over 50 business days. Evidently, the LT structure leads to outcomes that dominate those of NT models regardless of the portfolio allocation rules. The table also reports average transaction volume in weights per day for each model; this indicates that the LT models entail lower levels of transaction than the NT models. The investor benefit is essentially cumulative return minus transaction cost, so that this is a very clear additional attraction of the threshold model approach for increasingly high-dimensional problems: the sparsity induced by the LT structure more effectively decreases overall uncertainties in the portfolio decision-making process as dimension grows.

5.6 Exchange rate data analysis – World-wide currencies

5.6.1 Data and priors

The last analysis for exchange rate return time series uses $m = 20$ international currencies relative to US dollar, as listed in Table 5.7. The time period is $T = 1,043$ business days beginning in January 2008 and ending in December 2011. The study considers the LTF form based on eqns.(3.14)-(3.24) in which the specification includes

Table 5.7: World-wide 20 international currencies.

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

no covariates ($\mathbf{x}_t = \mathbf{0}$, $\mathbf{Z}_t = \mathbf{O}$), but does include correlation in factor innovation, apart from independent assumption of the previous analyses. In the prior literature, little has been discussed about correlations between factor processes. Exploratory studies with a number of financial data sets have convinced us of the need to allow dependencies among factors and of the forecasting benefits this can induce.

For the number of factors and ordering of the variable in the $\mathbf{y}_t$, some pre-analysis suggests $q = 4$ factors (using the same general ideas as in the previous sections), and an appropriate choice of the first four series in $\mathbf{y}_t$ 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, the same priors are specified as in Section 5.3 with additional components: $1/v_{\alpha i} \sim G(40, 0.005)$, $(\phi_{\alpha i} + 1)/2 \sim B(20, 1.5)$, and $\mu_{\alpha i} \sim N(0, 1)$.

5.6.2 Summaries of posterior inferences

Figure 5.23 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 factors show a high volatility period in late 2008 tracing major turbulent fluctuations of the financial crisis triggered by the Lehman brothers shock. Focusing further around mid-2008, Factor1 (GBP-leading)

FIGURE 5.23: Left column: posterior means of trajectories of factors f_{it} for world-wide FX data. Central column: posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of trajectories of factor stochastic volatilities $\tilde{w}_{it}$, which is the square root of diagonal elements of $\mathbf{W}_t$. Right column: posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of trajectories of underlying components of factor volatilities $\psi_{it} = \exp(\lambda_{it}/2)$.

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 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 5.24 graphs the poster estimates of trajectories of some of the factor cor-

FIGURE 5.24: Posterior means and ± 1 standard deviation credible intervals (dotted) of trajectories of several correlations from $\mathbf{W}_t$. Factors are labeled by leading currency.

relations. 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, a sudden change is commonly 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 5.25 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

FIGURE 5.25: Posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of trajectories of selected time-varying factor loadings for world-wide FX data. The dashed lines show posterior means of the corresponding latent thresholds. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory.

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

5.6.3 Portfolio analysis

Portfolio studies are provided following similar themes to those of Section 5.1, while now using more realistic and structured portfolio strategies extending the target return, minimum variance portfolio. The total sum of investment is again fixed by restricting $\mathbf{w}_t' \mathbf{1} = 1$, where $\mathbf{w}_t$ 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. The study assumes 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.

An explicit form for the optimized weights is derived for the decoupling strategy. Define R_t as the return of the benchmark asset at time t , and write the mean and covariance structure of $\mathbf{y}_t$ and z_t jointly as follows:

$$E \begin{pmatrix} \mathbf{y}_t \\ R_t \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{y}}_t \\ \bar{R}_t \end{pmatrix}, \quad V \begin{pmatrix} \mathbf{y}_t \\ R_t \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{Q}}_t & \bar{\mathbf{q}}_t \\ \bar{\mathbf{q}}_t' & \bar{\xi}_t \end{pmatrix},$$

with a $m \times 1$ vector of covariances $\bar{\mathbf{q}}_t = Cov(\mathbf{y}_t, R_t)$ between asset returns and the benchmark under our “current” forecast distribution. At time t , we minimize the ex-ante portfolio variance $\mathbf{w}_t' \bar{\mathbf{Q}}_t \mathbf{w}_t$, subject to (i) $\mathbf{w}_t' \bar{\mathbf{y}}_t = \bar{R}_t + r_t$, (ii) $Cov(\mathbf{w}_t' \mathbf{y}_t, R_t) = 0$, and (iii) $\mathbf{w}_t' \mathbf{1} = 1$, where r_t is the expected margin gained over the forecast mean of the baseline asset. This quadratic optimization yields an explicit solution $\mathbf{w}_t = \bar{\mathbf{K}}_t \mathbf{C}_t' (\mathbf{C}_t \bar{\mathbf{K}}_t \mathbf{C}_t')^{-1} \mathbf{a}_t$, where $\bar{\mathbf{K}}_t = \bar{\mathbf{Q}}_t^{-1}$, $\mathbf{C}_t = [\bar{\mathbf{y}}_t; \bar{\mathbf{q}}_t; \mathbf{1}]'$, and $\mathbf{a}_t = (r_t, 0, 1)'$.

In this portfolio allocation study, three models are compared:

- **NT**: Non-threshold and independent factor innovations, i.e., $\mathbf{A}_t = \mathbf{I}$;
- **NTC**: Non-threshold and correlated factor innovations;
- **LTC**: Latent threshold and correlated factor innovations.

For the baseline asset, the study uses S&P500 index from the US stock market. To obtain the covariance structure between the 20 currencies and the baseline asset, the study reanalyzed the LTF models fit to the $m = 21$ vector $(R_t, \mathbf{y}_t)'$. with $q = 5$ factors and the ordering of the first five variables in $\mathbf{y}_t$ as (US-stock, GBP, EUR,

JPY, INR), a natural extension of the earlier FX 4-factor model. In addition to the restricted condition, $\mathbf{w}'_t \mathbf{1} = 1$, the study examines 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% .

Table 5.8 reports realized cumulative returns of the portfolio allocation from the sequential investment over 50 business days. The use of correlated factor innovations and the latent threshold structure evidently and uniformly dominates the other models regardless of the target margins. Figure 5.26 shows the cumulative returns across time periods using a target margin of $r_t = 0.06\%$. The LTC model performs the best among three candidate models, and the NTC model clearly outperforms the NT model, which implies both 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 5.26(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; this evidently provides a beneficially, more stable portfolio. Portfolios based on the standard portfolio allocation are also computed, which excludes the constraint related to the benchmark asset (i.e., setting $\bar{R}_t = 0$ and $\bar{\mathbf{q}}_t = \mathbf{0}$). Figure 5.27(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 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.

Table 5.8: Cumulative returns (%) for (i) restricted ($\mathbf{w}'_t \mathbf{1} = 1$) and (ii) unrestricted allocations over 50 business days for world-wide FX data.

	NT	NTC	LTC
(i) Restricted allocation			
Target margin: $r_t = 0.02$	-1.676	-0.678	1.551
0.04	-1.851	-0.884	1.441
0.06	-2.027	-1.089	1.330
(ii) Unrestricted allocation			
Target margin: $r_t = 0.02$	-2.828	-2.483	-0.110
0.04	-2.974	-2.575	-0.106
0.06	-3.121	-2.667	-0.101

FIGURE 5.26: Cumulative returns (%) over 50 business days for target margin $r_t = 0.06\%$ with restricted (top) and unrestricted (bottom) allocations.

FIGURE 5.27: Analysis of world-wide 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 $r_t = 0.06\%$.

5.7 Summary

The results reveal that the LTF models can identify naturally emerging patterns of remarkably correlated dynamics in factors, coupled with dynamic shrinkage in factor loadings. Dynamic sparsity patterns induced by the LTM structure underlie plausible parameter estimates for complicated flow and dynamics in the exchange rate time series, 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.

6

Application to EEG Time Series

6.1 Context and literature

Electroencephalographic (EEG) recordings are time series of electrical potential fluctuations at various scalp locations of a human subject, reflecting physiological behavior underlying brain activity and connectivity. Analysis of multichannel EEG traces has become a standard tool to understand mechanisms of electroconvulsive therapy (ECT), one of the most effective treatments known for major depression with electrically induced seizures in patients (Weiner and Krystal 1994). The convulsive seizure activity drives multichannel EEG traces and a statistical interest here is to model such multivariate time series processes that reflect dynamic relations across EEG channels and time in order to reveal underlying characteristics and effects of ECT. Various classes of dynamic time series models have been studied to explore 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). Among them, time-varying parameter autoregressive (TV-AR) models and decompositions (West et al. 1999; Prado et al. 2001) reveal considerable changes in the patterns of evolution of time-

frequency structure with spectral content, differences and changes in relationships among the channels. However, for a further multivariate modeling and application with dynamic latent components, the increasing dimension of parameters requires refined prior and model structuring.

In this chapter, an analysis applies the LTF model developed in Section 3.3 to multivariate times series of 19 channels of EEG recordings. The model is built to address the following potential features underlying the EEG traces:

- all 19 EEG series have a common pattern of frequency content, i.e., a dominant “seizure” process;
- departures from the common pattern differ across the channels and time;
- the latent seizure process is dominated by quasiperiodic behavior that dynamically and globally changes over time; and
- some series exhibit signals delayed from that in other series during some time periods.

A key is modeling a TV-AR structure for the latent seizure process in the dynamic factor model, where multi-factor loadings are associated with the current value or the most recent values of the seizure process. This is overlaid with the latent threshold approach that allows the time-varying factor loadings to shrink to zero dynamically. The temporal sparsity structure provides practical and interpretable outcomes to more clearly understand mechanisms by which the latent seizure process drives activities in each area monitored by the EEG probes, as well as channel-specific local sub-activities induced by stochastic variances of idiosyncratic errors.

6.2 Latent threshold factor process (LTFP) model

Following the LTF model in Section 3.3, consider the model

$$\mathbf{y}_t = \mathbf{B}_t \mathbf{f}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}_t), \quad (6.1)$$

where $\mathbf{B}_t$ is the $m \times q$ time-varying loadings matrix, $\mathbf{f}_t = (x_t, x_{t-1}, \dots, x_{t-q+1})'$ here is a $q \times 1$ vector of latent factor that consists of the most recent q values of the underlying driving latent process $\{x_t\}$, and $\boldsymbol{\Lambda}_t$ is the time-varying volatility matrix with the diagonal structure $\boldsymbol{\Lambda}_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{mt}^2)$. The latent process $\{x_t\}$ follows the univariate TV-AR model

$$x_t = \sum_{j=1}^p \delta_{jt} x_{t-j} + \varepsilon_t, \quad \varepsilon_t \sim N(0, w_t), \quad (6.2)$$

$$\boldsymbol{\delta}_t = \boldsymbol{\delta}_{t-1} + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim N(\mathbf{0}, \boldsymbol{\Psi}_t), \quad (6.3)$$

where $\boldsymbol{\delta}_t = (\delta_{1t}, \dots, \delta_{pt})'$ is the time-varying AR coefficients. The study assumes that the $\boldsymbol{\nu}_t$, ε_t , $\boldsymbol{\xi}_t$ sequences are mutually independent.

A key feature here is a single, focal latent process but multivariate parallel factor loadings in terms of distributed lags. This factor modeling is aimed at capturing lag-lead structures across the EEG channels (Prado et al. 2001). For the 19 EEG series, we observe the signals on some EEG channels are delayed with respect to other channels, and such tendency depends on the location and parts of the seizure course. To identify the single latent factor process and to capture the dynamics of lag-lead structure, the first row of $\mathbf{B}_t$ is constrained to explicitly contain one unity element and zeros elsewhere. For example, when the first row of $\mathbf{B}_t$ is assumed $(0, 0, 1, 0, 0)$ with $q = 5$, the first EEG channel in $\mathbf{y}_t$ loads only x_{t-2} as an identification basis. Then, for the other channels, the first (last) two columns of $\mathbf{B}_t$ load the lead (lagged) values of the process.

To specify the stochastic variances $\{\sigma_{1:m,t}^2, w_t\}$, a single discount factor approach is adopted for modeling slow and steady random changing variances with discount learning (West and Harrison 1997; Prado and West 2010, Chapter. 4). The study assumes $\Psi_t \equiv \Psi$, a time-invariant covariance matrix for innovation of the time-varying AR coefficients. A key feature of the model here is again the LTM structure implemented in the factor loadings $\mathbf{B}_t$. This structured model is labeled as a latent threshold factor process (LTFP) model. The computational methodology for the LTFP model is mostly described in Chapter 3; additional posterior computations for the latent factor process states and the volatility process are provided as follows.

Latent factor process states $\{x_t\}$: The model of eqns.(6.1)-(6.3) can be regarded as a conditionally linear, Gaussian dynamic model for a state vector $\tilde{\mathbf{f}}_t = (x_t, \dots, x_{t-p+1})'$, with a state transition

$$\tilde{\mathbf{f}}_t = \mathbf{G}_t \tilde{\mathbf{f}}_{t-1} + \mathbf{e}_t,$$

$$\mathbf{G}_t = \begin{pmatrix} \delta_{1t} & \delta_{2t} & \cdots & \delta_{p-1,t} & \delta_{pt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \mathbf{e}_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Generation of the latent states is implemented by the FFBS algorithm.

Volatility processes: The following filtering method for $\{w_t\}$ yields sequential updating of conditional posterior of its precision $a_t = 1/w_t$ that discounts historical information based on a discount factor $\lambda_w \in (0, 1)$, typically close to one. Specifically, with accumulated data $D_t = \{\varepsilon_1, \dots, \varepsilon_t\}$ at time t , write the conditional posterior for the precision at time $t-1$ as $a_{t-1}|D_{t-1} \sim G(n_{t-1}/2, S_{t-1}/2)$. The discount learning reads $a_t|D_{t-1} \sim G(\lambda_w n_{t-1}/2, \lambda_w S_{t-1}/2)$, which is updated by the time

t error ε_t to the time t posterior $a_t|D_t \sim G(n_t/2, S_t/2)$, where $n_t = \lambda_w n_{t-1} + 1$, and $S_t = \lambda_w S_{t-1} + \varepsilon_t^2$. The past information is discounted to reflect changes in volatility with sequentially updated degrees of freedom (n_t) and sum-of-squared product (S_t). This procedure is also implemented for updating $\{\sigma_{it}^2\}$ with a discount factor $\lambda_{\sigma,i}$, for $i = 1, \dots, m$.

6.3 EEG data analysis

6.3.1 Data and priors

The analysis uses 19-channel EEG traces recorded in one seizure of one patient, analyzed in West et al. (1999), Prado et al. (2001), Prado and West (2010). 19 parallel EEG series are recorded from 19 electrodes of the International 10-20 EEG System, located around and over the patient's scalp. 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. The analysis uses 3,000 observations from the sub-sampled series.

Figure 6.1 illustrates a representation of the approximate 19-electrodes placement over the scalp. From the results of Prado et al. (2001), the vertex channel Cz is ordered in the first element of $\mathbf{y}_t$ to identify the latent factor process, and the first row of $\mathbf{B}_t$ is specified as $(0, 0, 1, 0, 0)$, so that the lead/lag structure can be captured for the range $t+2, t+1, \dots, t-2$. This implies that timing of the loading on the latent

FIGURE 6.1: Representation of the approximate 19-electrodes placement over the scalp.

factor process is determined with respect to the signal of the channel Cz. Preliminary analyses fitting the LTFP model to the EEG data show that the identification for the latent factor process requires a time-invariant variance for the idiosyncratic shocks for the channel Cz, i.e, $\sigma_{1t}^2 \equiv \sigma_1^2$, which avoids undesirable sample correlations between the processes $\{\sigma_{1t}^2\}$ and $\{w_t\}$; the prior is assumed as $\sigma_1^{-2} \sim G(500, 10^4)$. The other stochastic variances use the discount factor $\lambda_w = \lambda_{\sigma,i} = 0.99$, for $i = 2 : m$. The following priors are used for the hyperparameters: $v_{\beta i}^{-2} \sim G(50, 0.01)$, $(\phi_{\beta i} + 1)/2 \sim B(20, 1.5)$, $\mu_{\beta i} \sim N(0, 1)$; and $\Psi^{-1} \sim W(100, 10^{-3}I)$. The MCMC analysis was run for a burn-in period of 5,000 samples prior to saving the following MCMC sample of size $J = 20,000$ for summary posterior inferences.

Determination of the number of AR lags (p) has been a challenging issue for univariate time-varying autoregressive processes, discussed in many studies (e.g., Huerta and West 1999; Prado and Huerta 2002). This application estimated the LTFP model for the EEG data with $p = 6, \dots, 12$, and examined a decomposition of the latent process $\{x_t\}$ with the time-frequency structure based on the eigenstructure

of the TV-AR evolution matrix $\mathbf{G}_t$. The basic result of West et al. (1999) states that

$$x_t = \sum_{j=1}^{p_{z,t}} x_{t,j}^{(z)} + \sum_{j=1}^{p_{y,t}} x_{t,j}^{(y)}, \quad (6.4)$$

where $p_{z,t}$ is the number of pairs of complex characteristic roots of the time t instantaneous AR characteristic polynomial based on $\boldsymbol{\delta}_t$, and $p_{y,t}$ is the number of pairs of real characteristic roots (i.e., $2p_{z,t} + p_{y,t} = p$). Under certain conditions in West et al. (1999), it is shown that the processes $\{x_{t,j}^{(z)}\}$ and $\{x_{t,j}^{(y)}\}$ are dominated by time-varying autoregressive moving average, TV-ARMA(2,1) and TV-AR(1) processes, respectively (see also Prado 1998; Prado and West 2010). The TV-ARMA(2,1) component is quasiperiodic with time-varying characteristic frequency and modulus. Fitting the LTFP model to the EEG data reveals three substantial latent components of the quasiperiodic processes with each moduli values higher than 0.5 during whole the course of the seizure. In higher-order AR lags, additional quasiperiodic components are found, while the trajectory of their frequency and modulus are relatively unstable. From these experiments, the analysis uses $p = 6$ in the following posterior summaries.

6.3.2 Summaries of posterior inferences

Figure 6.2 displays the posterior means of the factor latent process $\{x_t\}$ and stochastic variance $\{w_t\}$ for the latent process innovation, as well as the trajectories of estimated characteristic frequency and modulus for the quasiperiodic components. In Figure 6.2(ii), the innovation variance driving the factor latent process is evidently variable over time, rising to high values after the take-off of the seizure and then eventually considerably decaying toward zero. This evidence implies that the dominating higher-frequency seizure oscillation appears in the early period driven by the impact of the ECT; then it gradually diminishes toward the end of the sample period, con-

FIGURE 6.2: Estimated latent factor process for EEG data: posterior means of (i) the factor (x_t) and (ii) the innovation standard deviation ($w_t^{1/2}$); the characteristic (iii) frequencies and (iv) moduli of three quasiperiodic components; (v) frequency and (vi) modulus of the lowest-frequency component (solid) with 95% credible intervals (dotted).

sistent with the result of Prado et al. (2001). Figure 6.2(iii) graphs the frequency trajectories of three quasiperiodic components, showing that each component lies in one of the neuropsychiatric categories, so-called “delta” (roughly 0-4Hz), “theta” (4-8Hz), and “alpha” (8-13Hz) ranges (see e.g., Dyro 1989). The lowest-frequency component stays in the delta range and gradually decreases over time. The corresponding modulus is close to one (solid line in Figure 6.2(iv)), which indicates a considerably persistent component referring to the so-called delta-slow wave that dominates the factor process during the course of the seizure (West et al. 1999). The other quasi-periodic components lie in the theta and alpha ranges, while their moduli and amplitudes are relatively lower than the dominating lowest-frequency component over the whole seizure course.

FIGURE 6.3: Trajectories of posterior means (solid) and 95% credible intervals (dotted) for the TV-AR coefficients $\{\delta_{1:p,t}\}$.

Figure 6.3 provides the trajectories of the posterior mean and 95% credible intervals for the TV-AR coefficients $\{\delta_{1:p,t}\}$. The estimates suggest that the coefficients are clearly time-varying and mostly relevant during the sample periods. The 95% credible intervals are slightly wider during the late time periods, which leads to a change in uncertainty of the quasiperiodic component. Figures 6.2(v) and (vi), displaying the posterior means and 95% credible intervals of the frequency and modulus for the lowest-frequency component respectively, indicate an increase in the uncertainty of the frequency and the modulus around the end of the seizure, as pointed out by Prado et al. (2001).

Figure 6.4 plots trajectories of posterior probability of $s_{it} = 1$ for the channel F7. Time variation of the shrinkage is evident: temporal shrinkage is found in loadings of Lead(+2), Lead(+1) and Lag(−2). The synchronized loading with respect to the channel Cz are relevant, whereas the Lag(−1) loading is totally selected out over the entire time-frame.

FIGURE 6.4: Trajectories of posterior probability of $s_{it} = 1$ for the channel F7 .

Figure 6.5 provides the posterior probability of $s_{it} = 1$ across the channels drawn at selected time points. The images are created by linearly interpolating the estimates at the electrode locations. Note that this figure is not obtained from any spatial dependent structure model, but marked patterns of the shrinkage are found and seem to evolve through the time. For example, loadings of Lead(+2) are commonly and globally shrunk to zero from left frontal to right occipital sites. The Lead(+2) loadings around right frontal and prefrontal areas exhibit evolving degrees of shrinkage. Similar changes are found in the parietal and occipital regions of Lag(−2) loadings. Meanwhile, almost no shrinkage is found in the synchronized loadings except for the channel T3 (left temporal).

Figure 6.6 illustrates estimated factor loadings with the lag-lead structure drawn at selected time points. The displayed images are based on the posterior mean ($\hat{\beta}_{it}$) when the posterior probability of $s_{it} = 1$ is larger than 0.5, and zero otherwise;

i.e., the value of $\hat{\beta}_{it} \times I[\widehat{\Pr}(s_{it} = 1 | \mathbf{y}_{1:T}) > 0.5]$. Note that the factor loading of the channel Cz is 1 for the basis and 0 for lagged/leaded times. The estimates show a considerable relationship between neighbors of the channel. Overall, the synchronized loadings clearly exhibit a round contour, taking higher values in the central area of the scalp and getting lower towards the edges with some temporal local activities. The center of the high intensity contour moves from front to back through the course of the seizure, which is consistent with Prado et al. (2001). In the first row of the figure ($t = 500$), i.e., an early stage of the seizure, the loadings exhibit marked intensity around right prefrontal sites in Lead(+2), right frontal sites in Lead(+1), left temporal and parietal to occipital regions in Lag(-1), and the left occipital area in Lag(-2). The marked intensity in Lag(-2) diminishes in the second row of the figure ($t = 1,400$, a middle period of the seizure), and instead the intensity of Lag(-1) loadings increases around the occipital region.

In the third row of the figure ($t = 2,200$), the highest intensity appears in the channel Pz, and the parietal region exhibits rising intensity. Another higher intensity is found around the right temporal area in Lead(+2) and in the channel C4 in Lead(+1). This suggests dynamics of the driving latent process with the lead effect of the right temporal/central areas and with the lag effect of the occipital region, which leads to an asymmetric pattern of the spatial relation in loadings, as indicated by Prado et al. (2001). In the last row of the figure ($t = 2,700$, a late stage of the seizure), the major lead/lag intensity diminishes, and the higher intensity stays around the central to parietal areas in the synchronized loadings. Through the focal driving process $\{x_t\}$ driven from a neural region around some EEG probes, the signal is passed along to other probe regions, exhibiting the time-varying spatial relations in the estimated factor loadings.

Figure 6.7 plots posterior means of the standard deviation for idiosyncratic in-

novation shocks, $\{\sigma_{it}\}$. Note that $\{\sigma_{1t}\}$, the innovation standard deviation for the channel Cz, is assumed time-invariant, as stated before. The figure shows spatial relations in a pattern of trajectory across channels. Trajectories of the neighboring channels F4 and F8 are clearly similar, exhibiting a major hike in the middle of the seizure. It is evident that some parietal and occipital sites (T3, P3, O1, O2 and P4) share a common trajectory, which marks a peak in an early stage of the seizure and then gradually decreases toward the end of the seizure. As seen in Figures 6.5 and 6.6, these sites also share common patterns in the shrinkage and loadings of $\text{Lag}(-1)$. Further, the estimate shows a similarity among the channels Pz, C4 and T6, which differs from the occipital region. This suggests an intrinsic difference between the central sites (Pz, C4 and T6) and the occipital areas (T3, P3, O1, O2 and P4), which is also observed in Figures 6.5 and 6.6. These idiosyncratic innovation shocks primarily consist of the variation of some local sub-activity of the seizure that is not explained by the latent factor process as well as some measurement errors such as power line noise and scalp-electrode impedance, which presumably cause the spatial patterns in the idiosyncratic innovation variance.

FIGURE 6.5: Posterior probability of $s_{it} = 1$ with the lag-lead structure at selected time points.

FIGURE 6.6: Estimated factor loadings with the lag-lead structure at selected time points. The displayed values are the posterior means multiplied by the posterior shrinkage indicator, $I[\Pr(s_{it} = 1) > 0.5]$.

FIGURE 6.7: Posterior means of time-varying standard deviation for idiosyncratic shocks, $\{\sigma_{it}\}$. Each graph is roughly located at the corresponding electrode placement. The x -axis refers to time $t = 1 : 3,000$.

6.3.3 An extension and impulse response analysis

The analysis examines an extension of the LTFP model with an additional lagged variable, that is,

$$\mathbf{y}_t = \sum_{j=1}^r \mathbf{\Gamma}_{jt} \mathbf{y}_{t-j} + \mathbf{B}_t \mathbf{f}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}_t), \quad (6.5)$$

where $\mathbf{\Gamma}_{jt}$ is the $m \times m$ matrix of TV-VAR coefficients; the LTM structure is adopted for each element of $\mathbf{\Gamma}_{jt}$. The baseline LTFP model reasonably explained the EEG traces with the underlying factor process driving the multivariate responses, and the other idiosyncratic shocks and local activities are described by the innovations $\boldsymbol{\nu}_t$. The introduction of the LT-VAR term in eqn.(6.5) is aimed to represent the spill-over effect of the EEG cascade for such local activities that are not explained by the underlying factor process. Typically, $\mathbf{\Gamma}_{jt}$'s are expected to be sparse; they represent local instantaneous interconnections around the surface of the brain, which are presumably sparse and also time-varying. The study assumes parameters in the first row and in the first column are equal to zero, reflected from the identification structure in $\mathbf{B}_t$ in relation to the Cz channel. In this setting with $r = 1$, posterior estimates of $\mathbf{\Gamma}_{1t}$ reveal some relevance of lagged coefficients in predicting the responses. This extended model, labeled as a LTFP-VAR model, and the baseline LTFP model are compared in the following impulse response analysis.

Based on posterior means for all latent variables at each time point, impulse responses are computed at each channel from an experimental shock to the underlying factor process. Specifically, the analysis examines paths of impulse response with the shock to the innovation of the process in eqn.(6.2), where the shock size is the average of posterior means of w_t across time. Figure 6.8 plots the impulse responses of the 19 EEG channels to underlying factor process shock, obtained from the baseline LTFP

FIGURE 6.8: Impulse responses of 19 EEG channels to underlying factor process shock, obtained from (i) LTFP model and (ii) LTFP-VAR model. In the upper row, the graphs are simply truncated on the horizontal axis (time horizon from the shock) for visual clarity.

model and the LTFP-VAR model. Patterns of the impulse response are clearly time-varying across exhibited time points with respect to wave frequency, magnitude, decay speed, and variation across the channels. In early periods of the seizure, the response decays slowly with a high-frequency cyclical wave, while in later periods it decays quickly with a low-frequency after a higher-magnitude oscillation. The responses from the LTFP-VAR model exhibit more variation across the channels than those from the LTFP model, which probably results from some spill-over effects of idiosyncratic shock and local activities revealed by the LT-VAR structure. Further, Figure 6.9 plots heat maps of the impulse response obtained from the LTFP-VAR model at selected time points for horizons $h = 2$ and 8, which clearly shows the time variation of the responses spreading over the channels.

Movies for some summaries of the posteriors in this chapter are provided in Nakajima and West (2012c).

FIGURE 6.9: Impulse responses of 19 EEG channels to underlying factor process shock from LTFP-VAR model: Heat map at selected time points for horizons $h = 2$ and 8.

Dynamic Network Modeling

7.1 Context and literature

A variety of network analyses have been studied in many scientific fields from neuroscience to financial econometrics. For statistical estimation of networks, VAR models have become a standard tool (e.g. Valdés-Sosa 2004; Valdés-Sosa et al. 2005). A typical approach to exploring relationships between responses is to examine the relevance of VAR coefficients, relying on Granger (1969) causality; a considerable number of works in network modeling have been based on Granger causality. Further, TV-VAR models have become popular in analysis of dynamic network data (e.g. Fujita et al. 2007). From another perspective, graphical modeling for multivariate responses addresses network associations induced by conditional independencies via graphs (Whittaker 1990; Lauritzen 1996; Dobra et al. 2004; Jones et al. 2005; Carvalho and West 2007).

This chapter introduces a general framework that builds on the LTM technique. As related literature, Valdés-Sosa et al. (2005) employ sparsity modeling for VAR coefficients with standard penalized regression techniques in analysis of brain func-

tional connectivity. In another approach, Yoshida et al. (2005) propose a dynamic linear model with Markov switching for estimating time-dependent gene network structure. The LTM approach differs from these approaches in that threshold mechanisms operate continuously in the parameter space and over time, which induces temporal sparsity in time-varying parameters. The proposed network modeling is the dynamic sparsity pattern in time-varying VAR coefficients of lagged predictors as well as in time-varying precision matrices. A salient feature of the LTM approach is the Cholesky-type triangular decomposition of covariance matrices defined by eqns.(3.5)-(3.6) that allows parallel computation of univariate dynamic regressions (Lopes et al. 2012). This work responds to increasingly pressing needs to develop network detection method in high-dimensional problems. This chapter illustrates this technique with application to time series of foreign currency exchange rates (FX) and electroencephalography (EEG) recordings.

In the context of Gaussian graphical modeling, zeros in the off-diagonal elements of the precision matrix $\mathbf{\Omega}_t$ imply lack of edges between responses in an undirected graph representation (Whittaker 1990; Lauritzen 1996; Jones et al. 2005). These conditional independencies are key idea to reducing parameter dimension and to exploring effectively parsimonious structure in high-dimensional problems. A goal here is to explore dynamic networks of undirected association based on the graph induced from the time-varying precision matrix, as well as those of directed association based on coefficients of lagged predictors in the TV-VAR models using the LTM mechanism that thresholds these time-varying parameters. The study also addresses prior works that provide an efficient and fast computational scheme which arises from conditionally independent univariate regressions estimated in parallel.

7.2 LT-VAR model revisited

Consider the TV-VAR(p) model

$$\mathbf{y}_t = \sum_{j=1}^p \mathbf{\Gamma}_{jt} \mathbf{y}_{t-j} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t), \quad (7.1)$$

where $\mathbf{\Gamma}_{jt}$ is the $m \times m$ matrix of time-varying coefficients at lag j , ($j = 1, \dots, p$). The model can include time-varying intercepts and regression for other predictors known at time $t - 1$, although these components are omitted in this chapter for simplicity.

The Cholesky decomposition of $\mathbf{\Sigma}_t$ defined by eqns.(3.5)-(3.6) yields

$$\mathbf{A}_t \mathbf{y}_t = \sum_{j=1}^p \mathbf{B}_{jt} \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{\Lambda}_t), \quad (7.2)$$

where $\mathbf{B}_{jt} = \mathbf{A}_t^{-1} \mathbf{\Gamma}_{jt}$, for $j = 1, \dots, p$. The (i, j) off-diagonal element of $\mathbf{A}_t$ is defined as $-a_{ij,t}$ for $i > j$ in this chapter, apart from the definition of eqn.(3.6). With the diagonal structure of $\mathbf{\Lambda}_t$, the model can then be recast as univariate triangular regressions:

$$\begin{aligned} y_{1t} &= \mathbf{B}_t^{(1)} \mathbf{z}_{t-1} + \varepsilon_{1t}, \\ y_{2t} &= \mathbf{B}_t^{(2)} \mathbf{z}_{t-1} + a_{21,t} y_{1t} + \varepsilon_{2t}, \\ y_{3t} &= \mathbf{B}_t^{(3)} \mathbf{z}_{t-1} + a_{31,t} y_{1t} + a_{32,t} y_{2t} + \varepsilon_{3t}, \\ &\vdots \\ y_{mt} &= \mathbf{B}_t^{(m)} \mathbf{z}_{t-1} + a_{m1,t} y_{1t} + \dots + a_{m,m-1,t} y_{m-1,t} + \varepsilon_{mt}, \end{aligned} \quad (7.3)$$

where $\mathbf{z}_{t-1} = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$; $\mathbf{B}_t^{(i)}$ is the $1 \times mp$ vector of corresponding coefficients in the i -th rows of $\mathbf{B}_{jt}$'s; $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$, for $i = 1, \dots, p$; and $\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$,

for $i \neq j$. Under conditionally independent priors as specified below, inference in the original TV-VAR model reduces to simply m univariate DLMS, which leads to efficient and fast parallel computation. This structure is of great advantage for analysis and prediction in higher-dimensional problems.

In the literature on network analysis, a typical approach examines the relevance of VAR coefficients, following Granger (1969) causality. This suggests that a response y_{1t} is *causal* for another response y_{2t} if $y_{1,t-1}$ can improve prediction of y_{2t} conditional on $y_{2,t-1}$ and other all information available at time $t-1$ except $y_{1,t-1}$. Following this concept, the analysis here explores sparsity in the TV-VAR coefficients to identify directed association, as well as sparsity in the precision matrix for undirected association based on the Gaussian graphical modeling strategy. A key approach to address these connectivity in a dynamic way is to employ the LTM ideas for time-varying parameters.

Following the discussion in Chapter 2, the LTM is defined here for the univariate time series y_{it} , ($t = 1, 2, \dots$), as following:

$$y_{it} = \mathbf{x}_{it}' \mathbf{b}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_{it}^2), \quad \mathbf{b}_{it} = (b_{i1t}, \dots, b_{ikt})', \quad (7.4)$$

$$b_{ijt} = \beta_{ijt} s_{ijt} \quad \text{with} \quad s_{ijt} = I(|\beta_{ijt}| \geq d_{ij}), \quad j = 1, \dots, k, \quad (7.5)$$

where

$$\beta_{ijt} = \mu_{ij} + \phi_{ij}(\beta_{ij,t-1} - \mu_{ij}) + \eta_{ijt}, \quad \eta_{ijt} \sim N(0, v_{ij}), \quad (7.6)$$

with $|\phi_{ij}| < 1$, for $j = 1, \dots, k$. The LTM is applied to the univariate triangular regressions developed above, where $\mathbf{x}_{1t} = \mathbf{z}_{t-1}$, $\mathbf{b}_{1t} = (\mathbf{B}_t^{(1)})'$; $\mathbf{x}_{it} = (\mathbf{z}_{t-1}', y_{1t}, \dots, y_{i-1,t})'$, and $\mathbf{b}_{it} = (\mathbf{B}_t^{(i)}, a_{i1,t}, \dots, a_{i,i-1,t})'$, for $i = 2, \dots, m$. Posterior estimates of these reduced-form parameters and sparsity patterns yields those of the original matrices $(\mathbf{\Gamma}_{jt}, \mathbf{\Omega}_t)$. The triangular regressions and LTM mechanism jointly lead to dynamic

model uncertainty in the original TV-VAR model by embodying time-varying sparsity in the lag coefficients as well as the precision matrix. This salient modeling strategy induces dynamic networks of directed/undirected association between multivariate responses. For the residual volatility σ_{it}^2 , ($i = 1, \dots, m$), the analysis adopts a discounting time-varying variance approach (West and Harrison 1997; Prado and West 2010, Chapter.4) or stochastic volatility process models as in the previous chapters.

One remark is in order, related to efficient and fast computation using parallelization of the MCMC. The assumption of prior independence for $\boldsymbol{\theta}_i$ and $\mathbf{d}_i = (d_{i1}, \dots, d_{ik})$ across i leads to conditionally independent posteriors across the triangular univariate regression of eqn. (7.3), as discussed by Lopes et al. (2012) in the context of the multivariate Cholesky stochastic volatility models. Specifically, given completed data $\mathbf{y}_{1:T}$, the sample of latent process states and hyper-parameters can be drawn based on the i -th triangular regression equation independently of the other regressions, which leads to m parallel processing of the MCMC sequences. As mentioned above, the goal in this chapter is to make inference on dynamic networks based on posteriors of $(\boldsymbol{\Gamma}_{jt}, \boldsymbol{\Omega}_t)$; this is obtained from the collection of posterior samples from m triangular regressions, explored in parallel. If we are interested in forecasting analysis, posterior predictive distributions can be sampled via the MCMC, now requiring recursive assessment of $y_{i,T+h}$ through the triangular regression for $i = 1, \dots, m$, and for horizons $h = 1, 2, \dots$, sequentially.

7.3 Exchange rate data analysis

The first study of dynamic network modeling applies the LTM-based network model to a series of daily foreign exchange (FX) rate returns from world-wide international financial markets. A focus here is sparsity patterns on time-varying relation-

Table 7.1: FX return data: 20 international currencies, ordered by the higher turnover in the global foreign exchange market in April 2007 reported by BIS (2010).

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

ships underlying the FX series and induced dynamic networks among international financial markets. As related literature 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. In the context of Granger (1969) causality, VAR or cointegration models have been well studied. Awokuse et al. (2009) use rolling cointegration methods to investigate evolving pattern of interdependence among international stock markets. In another context, Carvalho and West (2007) introduce dynamic Gaussian graphical models and provide analysis of FX and stock price index time series (see also Prado and West 2010, Chapter. 10).

7.3.1 Data and priors

The data are $m = 20$ daily international currency exchange rates relative to US dollar over a time period of $T = 1,304$ business days beginning in January

2007 and ending in December 2011; a set of the currencies is the same as Section 5.6, reproduced in Table 7.1. One important aspect of the Cholesky-type time-varying covariance/precision matrix in real data analysis is ordering of responses. As well discussed by Primiceri (2005), the order matters because the lower triangular structure is imposed on the $\mathbf{A}_t$. A practical approach to determining the ordering is to consider identification of structural shock based on the triangular reduction of the covariance matrix, i.e., $\mathbf{u}_t = \mathbf{A}_t^{-1}\boldsymbol{\varepsilon}_t$ in eqns. (7.1) and (7.2), depending on context and some prior theory. In this FX analysis, the currencies are ordered by the turnover in the global foreign exchange market in April 2007 reported by BIS (2010), as in Table 7.1. This implies that an idiosyncratic structural shock to one currency affects other currencies whose trade volume is less than that currency in one business day as a simultaneous relation, but not the other way around. One can argue such effects both ways, although this analysis relies on relevancy of the effect from higher-turnover currencies. Another ordering is examined and this issue is revisited later.

From previous works, the analysis uses $p = 1$ for the VAR lag. The stochastic volatility defined by eqn.(3.8) is adopted for $h_{it} = \exp(\sigma_{it}/2)$. The following priors are used: $1/v_{ij} \sim G(100, 0.001)$; $1/v_{hi} \sim G(3, 0.03)$; $(\phi_{ij} + 1)/2 \sim B(20, 1.5)$, $(\phi_{hi} + 1)/2 \sim B(20, 1.5)$; $\mu_{ij} \sim N(0, 1)$, and $\exp(-\mu_{hi}) \sim G(3, 0.03)$. 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.

7.3.2 Summaries of posterior inferences

Figure 7.1 displays posterior probabilities of $s_{ijt} = 1$ for the covariance components in $\mathbf{A}_t$ and induced posterior probabilities of dynamic network in the precision matrix $\boldsymbol{\Omega}_t = \mathbf{A}_t' \boldsymbol{\Lambda}_t^{-1} \mathbf{A}_t$, plotted for $t = 651$ (2009/Jun/30), the mid-time point of

FIGURE 7.1: Results of FX network analysis: Posterior probabilities of $s_{ijt} = 1$ for the covariance components in $\mathbf{A}_t$, induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$, and posterior probabilities $s_{ijt} = 1$ for the VAR coefficients in $\mathbf{B}_{1t}$. The panels are plotted for $t = 2009/6/30$.

the sample period. At this time, the volatility of the world market was temporarily low after the periods of the US financial crisis. The panel of $\mathbf{\Omega}_t$ exhibits interesting patterns of sub-networks among international currencies. The posterior probabilities for a group of the first 10 currencies are evidently high, which implies considerable correlations in the simultaneous shocks to major industrialized countries. The second sub-group consists of the last 10 currencies from emerging countries, exhibiting relatively more sparse structure than the 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 connectivity for the simultaneous shocks.

Posterior estimates of the shrinkage probability reveal a considerable stability

FIGURE 7.2: Evolution of networks for FX analysis: Induced posterior probabilities of dynamic network in Ω_t for select time points. The graph includes only edges having posterior inclusion probability exceeding 0.9.

of the network connections among the international currencies. Figure 7.2 graphs induced posterior probabilities of dynamic network in Ω_t plotted for selected time points. The graphs here are in a binary scale: they include only edges (elements in black) having posterior inclusion probability exceeding a cutoff value 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 7.1. 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 correlated in those periods, although the network model here provides the parsimonious structure which identifies some sub-groups of currencies. The third panel shows a further evolving network during the European crisis. One interesting change is that CHF has fewer 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.

FIGURE 7.3: Posterior means (solid) and ± 2 standard deviation credible intervals (dotted) of a_{ijt} as well as posterior means of latent threshold (dashed) for selected predictors. The title line “A–B” refers to a simultaneous coefficient of currency B in predicting A. Posterior probabilities of $s_{ijt} = 0$ are plotted below each trajectory

These posterior probabilities indicate relevance of edges in dynamic network; the model also provides posterior estimates of time-varying parameters. Figure 7.3 plots the posterior means and ± 2 standard deviation credible intervals of selected a_{ijt} 's, as well as posterior probabilities of $s_{ijt} = 0$ plotted below each trajectory. The first panel shows the posteriors of the EUR's simultaneous coefficient to predict JPY. After the outbreak of US financial crisis in September 2008, the coefficient rapidly declines. It keeps moving up and down afterward with temporal changes in posterior shrinkage probabilities. These behaviors probably reflect a market sentiment that JPY was regarded as a safer currency like CHF, 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 EUR's coefficient to predict GBP, implying a clear drop during the European crisis from 2010 to 2011. The third panel shows the GBP's

FIGURE 7.4: Results of FX network analysis: Posterior probabilities of $s_{ijt} = 1$ for induced dynamic network in Ω_t (top) and $\mathbf{B}_{1t}$ (bottom) from baseline ordering (left) and reverse ordering (right). The panels are plotted for $t = 2011/12/13$. See Figure 7.1 for the heat-map scale.

coefficient to predict CHF, which exhibits sharp declines in 2008 and 2010 related to the crisis periods. As can be seen in these results, even if the posterior probability of edges is stable, the model yields time-varying values in the latent parameters.

Posterior probabilities of $s_{ijt} = 1$ for the VAR coefficients in $\mathbf{B}_{1t}$ are plotted in Figure 7.1. The first lag coefficients are considerably sparse as expected in modeling returns of financial variables, except for some relevant coefficients in the KRW currency. Estimation results reveal that these sparsity patterns are quite stable through the sample periods.

Regarding the ordering in the triangular regressions, the study examines the reverse ordering: the currencies are ordered by lower turnovers. Figure 7.4 shows the

posterior probabilities in $\mathbf{\Omega}_t$ and $\mathbf{B}_{1t}$ for the time point of 2010/Oct/19 from the baseline ordering and the reverse ordering. 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 more sparse pattern in 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 covariance model. An analysis using reversible-jump MCMC methods can address prior works on the different orderings: this may be anticipated in future work.

7.4 EEG data analysis

7.4.1 Data and priors

The second analysis applies the LTM-based dynamic network models to the multivariate time series of EEG recordings used in Chapter 6 with $k = 19$ and $T = 3,000$. For the Cholesky-type triangular equations, the analysis uses a baseline ordering listed in Table 7.2(i), starting from the vertex channel Cz, known as a key factor

Table 7.2: 19 EEG channels: (i) Baseline and (ii) factor-based ordering.

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

that potentially drives the other EEG traces in previous literature, going around inside channels clockwise and then outside channels as well. The baseline ordering is arranged so that the responses have the spatial dependence structure, which is discussed as relevant in previous works. Another ordering is also assessed later in this section.

From preliminary analyses, the study uses $p = 2$ for the VAR lag. The model with some higher lags were examined up to $p = 6$, which typically yields overfitting tendency in posterior estimates and considerably sparse patterns in coefficients of higher lagged predictors. The following priors are used: $1/v_{ij} \sim G(200, 10^{-4})$, $(\phi_{ij} + 1)/2 \sim B(20, 1.5)$, and $\mu_{ij} \sim N(0, 1)$. For the time-varying volatility process, the analysis here adopts the discount factor approach as in Chapter 6. The discount factor is set as $\lambda = 0.99$. The iteration size of the MCMC analysis is the same as in the previous section.

7.4.2 Summaries of posterior inferences

Figure 7.5 displays posterior probabilities of $s_{ijt} = 1$ for the covariance components in $\mathbf{A}_t$ and induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$, plotted for $t = 1,501$. The panels exhibit a remarkable band pattern, indicating a spatial dependence among EEG channels. Figure 7.6 shows heat maps for the induced dynamic networks based on the posterior probabilities of the $\mathbf{\Omega}_t$ in Figure 7.5. 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 $\mathbf{\Omega}_t$ indicates relevant spatial relation between channels in the inside

FIGURE 7.5: Results of EEG network analysis: Posterior probabilities of $s_{ijt} = 1$ for the covariance components in $\mathbf{A}_t$, induced posterior probabilities of dynamic network in $\mathbf{\Omega}_t$, and posterior probabilities $s_{ijt} = 1$ for the VAR coefficients in $\mathbf{B}_{jt}$, $j = 1, 2$. The panels are plotted for $t = 1,501$. See Figure 7.1 for the heat-map scale.

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). 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 frontal region. In contrast, T4 exhibits relations with the 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.

FIGURE 7.6: Results of EEG network analysis: Posterior probabilities of $s_{ijt} = 1$ for the induced posterior probabilities of dynamic network in Ω_t , for $t = 1,501$. Each sub-map shows posterior probabilities that the channel of its plotted location (marked by $\otimes$) has the edge in the induced network. See Figure 7.1 for the heat-map scale.

These characteristics of networks among the EEG channels are dynamic in part; Figure 7.7 shows evolutions of the posterior probabilities in Ω_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 the 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.

The ordering issue is again addressed by examining alternative ordering labeled as “factor-based,” listed in Table 7.2. Prado et al. (2001) estimate a dynamic factor model for the same EEG time series and provide estimates of factor weights relating the responses of 19 channels to the underlying factor process. In the analysis here, the channels are arranged by their estimated posterior means of the factor weight from high to low. This weight can be regarded as a proxy of relations between each channel and the underlying factor process, which can be implicitly linked to the idea of associations among the channels. Figure 7.8 exhibits the posterior probabilities in Ω_t plotted for $t = 2,001$, from the baseline and the factor-based orderings. The network patterns are mostly similar to each other, indicating that the above estimates are reasonably robust with respect to the ordering in this analysis.

Movies for some summaries of the posteriors in this chapter are provided in Nakajima and West (2012d).

FIGURE 7.7: Evolution of networks for EEG analysis: Induced posterior probabilities of dynamic network in Ω_t and selected network maps for select time points. $\otimes$ refers to the channel location. See Figure 7.1 for the heat-map scale.

FIGURE 7.8: Results of EEG network analysis: Posterior probabilities of $s_{ijt} = 1$ for induced dynamic network in Ω_t from baseline ordering (left) and factor-based ordering (right). The panels are plotted for $t = 2,001$. See Figure 7.1 for the heat-map scale.

Concluding Remarks

Dynamic sparsity patterns induced by the latent threshold structure underlie plausible parameter estimates for complicated flows and dynamics in the time series data analyzed in this dissertation. The analyses using the LTM-based multivariate time series models illustrate the key ideas of the latent threshold strategy: reducing estimation uncertainty and improving prediction as well as model interpretation. These benefits of using LTMs are evidently general, as can be seen in the analyses using different time series data. The LTM has great potential to be a standard strategy for inducing dynamic sparsity in high-dimensional time series modeling, applicable to a wide range of real data analyses.

There are a number of methodological and computational areas for further investigation. In terms of modeling strategy, several generalizations of the LTM form can be considered as discussed in Section 2.5; modeling asymmetric latent thresholds and building structured shrinkage indicators are of interest. From another perspective, as mentioned in Section 4.1, Koop et al. (2009), Korobilis (2012), Chan et al. (2012) propose several techniques to explore parsimonious TV-VAR models. Though such

models share similar technical aspects with the LTM approach, they are inherently focused on different questions of identifying regime changes or “sudden breaks” at discrete times. The LTM approach is not focused on that at all; its goal is dynamic sparsity for model parameter reduction and the improved precision and predictions that can yield. It is certainly of interest to explore commonalities and differences between the approaches, and possibly direct comparisons in specific data analyses, and this may be anticipated in future work as the novel LTM approach becomes more widely explored.

Another possible future work is time change/shift in latent thresholds as mentioned in Section 4.2.2. From the forecasting results in the US macroeconomic data analysis, modeling a time-varying latent threshold is of interest. As related extensions, we can model the latent threshold or the shrinkage indicator variable as a function of covariates. In terms of modeling strategies in macroeconomic analysis, the LTM can be incorporated into more elaborate factor models, such as factor-augmented VAR (FAVAR) models (Bernanke et al. 2005; Baumeister et al. 2010), where the LTM strategy could play an important role as the model dimension grows.

One important open question is a potential computational strategy for sequential model learning, including sequential particle learning algorithms (e.g., Carvalho et al. 2010) for LTMs, as implementation of sequential updating and forecasting is central to investment decision making. In such a context, the computations involved in repeat MCMC analyses with a moving time window may become prohibitive and sequential Monte Carlo methods become increasingly relevant. For example, the examples in Chapter 5 show that scaling of the MCMC to moderately high dimensional series will be accessible. The study in Section 5.3 uses $m = 6$ series and $k = 3$ factors (with $p = 12$ loadings), while the $m = 40$ dimensional example in Section 5.5 has $k = 7$ factors ($p = 252$), with the time length almost equal. Regarding com-

putational time, using a dual-core 3.3GHz CPU computer, the first example takes 0.3 second per one iteration of MCMC, while this high-dimensional setting takes only 3.7 seconds. That is, compute time increases just over 12-fold in moving to the larger mode context, whereas the increase in the number of loadings is 21-fold. The analysis did not formally benchmark performance in scaling to increasingly high-dimensional settings, but this first experiment in scaling is encouraging. Moving to higher-dimensions (hundreds of series) will require different coding and computational strategies. For both MCMC and particle learning methods, scaling will have to explore parallelization. Distributing core computational tasks within each iterate of the demanding MCMC is one immediate consideration, and most sequential Monte Carlo methods are naturally parallelizable. Future studies will investigate the opportunities for parallelization that can exploit CUDA/GPU implementations (Suchard et al. 2010a,b) for massively parallel desktop analyses as well as traditional multi-core approaches.

Appendix A

Posterior Computation and MCMC Algorithm

A.1 MCMC algorithm for LTMs

This appendix details the MCMC algorithm for the basic form of LTM, defined by eqns.(2.1)-(2.3). The following prior forms are assumed: $\mu_i \sim N(\mu_{i0}, \sigma_{i0}^2)$, $(\phi_i + 1)/2 \sim \pi(\phi_i)$, $v_i^{-2} \sim G(v_{0i}/2, V_{0i}/2)$, $\sigma^{-2} \sim G(n_0/2, S_0/2)$, $\beta_{i1}|\boldsymbol{\theta}_i \sim N(\mu_i, w_i^2)$, and $d_i|\boldsymbol{\theta}_i \sim U(0, |\mu_i| + K_i w_i)$, where $w_i^2 = v_i^2/(1 - \phi_i^2)$. The MCMC algorithm is implemented for simulation of the full joint posterior $p(\boldsymbol{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \mathbf{d}|\mathbf{y}_{1:T})$ by the following sequence of sample generation.

A.1.1 Sampling $\boldsymbol{\Theta}$

Conditional on $(\boldsymbol{\beta}_{1:T}, \mathbf{d})$, sampling of the VAR parameters $\boldsymbol{\Theta}$ reduces to generation from conditionally independent posterior $p(\boldsymbol{\theta}_i|\beta_{i,1:T}, d_i)$, for $i = 1 : k$, due to the diagonal assumption for $\boldsymbol{\Phi}$ and $\mathbf{V}$. First, the conditional posterior distribution of μ_i is given by

$$\mu_i|\phi_i, v_i, \beta_{i,1:T}, d_i \sim TN_{D_i}(\hat{\mu}_i, \hat{\sigma}_i^2),$$

where TN_{D_i} denotes a truncated normal distribution that has a positive density in a region $D_i = \{\mu_i : d_i < |\mu_i| + K_i w_i\}$, and

$$\begin{aligned}\hat{\sigma}_i^2 &= \left\{ \frac{1}{\sigma_{i0}^2} + \frac{(1 - \phi_i^2) + (T-1)(1 - \phi_i)^2}{v_i^2} \right\}^{-1}, \\ \hat{\mu}_i &= \hat{\sigma}_i^2 \left\{ \frac{\mu_{i0}}{\sigma_{i0}^2} + \frac{(1 - \phi_i^2)\beta_{i1} + (1 - \phi_i) \sum_{t=1}^{T-1} (\beta_{i,t+1} - \phi_i \beta_{it})}{v_i^2} \right\}.\end{aligned}$$

Second, the conditional posterior density of ϕ_i is

$$\pi(\phi_i | \mu_i, v_i, \beta_{i,1:T}, d_i) \propto \pi(\phi_i) \sqrt{1 - \phi_i^2} \exp \left\{ -\frac{(\phi_i - \hat{\phi}_i)^2}{2\sigma_{\phi_i}^2} \right\} I(D_i),$$

where $\hat{\phi}_i = \sum_{t=1}^{T-1} \bar{\beta}_{i,t+1} \bar{\beta}_{it} / \sum_{t=2}^{T-1} \bar{\beta}_{it}^2$, $\sigma_{\phi_i}^2 = v_i^2 / \sum_{t=2}^{T-1} \bar{\beta}_{it}^2$ with $\bar{\beta}_{it} = \beta_{it} - \mu_i$, and $I(D_i)$ is an indicator function for $D_i = \{\phi_i : d_i < |\mu_i| + K_i v_i / (1 - \phi_i^2)^{1/2}\}$. The MH algorithm is implemented with a candidate generated as $\phi_i^* \sim TN_{(-1,1) \times D_i}(\hat{\phi}_i, \sigma_{\phi_i}^2)$; the corresponding acceptance probability is

$$\alpha(\phi_i, \phi_i^*) = \min \left\{ 1, \frac{\pi(\phi_i^*) \sqrt{1 - \phi_i^{*2}}}{\pi(\phi_i) \sqrt{1 - \phi_i^2}} \right\}.$$

Third, the conditional posterior distribution of v_i is obtained as $v_i^{-2} | \mu_i, \phi_i, \beta_{i,1:T}, d_i \sim G_{D_i}(\hat{v}_i/2, \hat{V}_i/2)$, where the gamma distribution is truncated to $D_i = \{v_i : d_i < |\mu_i| + K_i v_i / (1 - \phi_i^2)^{1/2}\}$, and

$$\hat{v}_i = v_{0i} + T, \quad \hat{V}_i = V_{0i} + (1 - \phi_i^2) \bar{\beta}_{i1}^2 + \sum_{t=1}^{T-1} (\bar{\beta}_{i,t+1} - \phi_i \bar{\beta}_{it})^2.$$

Note that these sample generations slightly differ from that of standard AR model parameters. Because the latent threshold prior uses the uniform prior distribution whose range depends on the AR parameters, the generation requires truncation of the standard conditional posterior distribution. Truncation is easily built in using a simple rejection step; acceptance rates are practically high in real data analyses.

A.1.2 Sampling $\beta_{1:T}$

Conditional on $(\Theta, \sigma, \mathbf{d}, \mathbf{y}_{1:T})$, the algorithm samples the conditional posterior distribution $p(\beta_t | \beta_{-t})$, sequentially for $t = 1 : T$ using a MH algorithm. The MH proposals come from a non-thresholded version of the dynamic linear model by fixing $\mathbf{s}_t = \mathbf{1}$; the proposal distribution forms $N(\beta_t | \mathbf{m}_t, \mathbf{M}_t)$, where

$$\begin{aligned}\mathbf{M}_t^{-1} &= \sigma^{-2} \mathbf{x}_t \mathbf{x}_t' + \mathbf{V}^{-1} (\mathbf{I} + \Phi' \Phi), \\ \mathbf{m}_t &= \mathbf{M}_t [\sigma^{-2} \mathbf{x}_t y_t + \mathbf{V}^{-1} \{ \Phi(\beta_{t-1} + \beta_{t+1}) + (\mathbf{I} - 2\Phi + \Phi' \Phi) \boldsymbol{\mu} \}],\end{aligned}$$

for $t = 2 : T - 1$. For $t = 1$ and $t = T$, a slight modification is required as follows:

$$\begin{aligned}\mathbf{M}_1^{-1} &= \sigma^{-2} \mathbf{x}_1 \mathbf{x}_1' + \mathbf{V}_0^{-1} + \mathbf{V}^{-1} \Phi' \Phi, \\ \mathbf{m}_1 &= \mathbf{M}_1 [\sigma^{-2} \mathbf{x}_1 y_1 + \mathbf{V}_0^{-1} \boldsymbol{\mu} + \mathbf{V}^{-1} \Phi \{ \beta_2 - (\mathbf{I} - \Phi) \boldsymbol{\mu} \}], \\ \mathbf{M}_T^{-1} &= \sigma^{-2} \mathbf{x}_T \mathbf{x}_T' + \mathbf{V}^{-1}, \\ \mathbf{m}_T &= \mathbf{M}_T [\sigma^{-2} \mathbf{x}_T y_T + \mathbf{V}^{-1} \{ \Phi \beta_{T-1} + (\mathbf{I} - \Phi) \boldsymbol{\mu} \}],\end{aligned}$$

where $\mathbf{V}_0 = \text{diag}(w_1^2, \dots, w_k^2)$. The candidate is accepted with probability

$$\alpha(\beta_t, \beta_t^*) = \min \left\{ 1, \frac{N(y_t | \mathbf{x}_t' \mathbf{b}_t^*, \sigma^2) N(\beta_t | \mathbf{m}_t, \mathbf{M}_t)}{N(y_t | \mathbf{x}_t' \mathbf{b}_t, \sigma^2) N(\beta_t^* | \mathbf{m}_t, \mathbf{M}_t)} \right\},$$

where $\mathbf{b}_t = \beta_t \circ \mathbf{s}_t$ is the current LTM state, and $\mathbf{b}_t^* = \beta_t^* \circ \mathbf{s}_t^*$ the candidate.

A.1.3 Sampling $\mathbf{d}$

A simple MH algorithm is adopted to sample the conditional posterior distribution of d_i , conditional on $(\Theta, \sigma, \beta_{1:T}, \mathbf{d}_{-i}, \mathbf{y}_{1:T})$, where $\mathbf{d}_{-i} = d_{1:k} \setminus d_i$. A candidate is drawn from the current conditional prior, $d_i^* \sim U(0, |\mu_i| + K_i w_i)$, and accepted with MH probability

$$\alpha(d_i, d_i^*) = \min \left\{ 1, \frac{N(y_t | \mathbf{x}_t' \mathbf{b}_t^*, \sigma^2)}{N(y_t | \mathbf{x}_t' \mathbf{b}_t, \sigma^2)} \right\},$$

where $\mathbf{b}_t$ is the state based on the current thresholds $(d_i, \mathbf{d}_{-i})$, and $\mathbf{b}_t^*$ the candidate based on $(d_i^*, \mathbf{d}_{-i})$.

A.1.4 Sampling σ

Finally, σ is drawn from $\sigma^{-2} | \boldsymbol{\beta}_{1:T}, \mathbf{d}, \mathbf{y}_{1:T} \sim G(\hat{n}/2, \hat{S}/2)$, where $\hat{n} = n_0 + T$, and $\hat{S} = S_0 + \sum_{t=1}^T (y_t - \mathbf{x}_t' \mathbf{b}_t)^2$.

A.2 MCMC algorithm for LT-VAR models

This section details posterior computations for the LT-VAR model where both the VAR coefficients and covariance components of Cholesky-decomposed variance matrices follow the LTM; the model comprises eqns.(3.2)-(3.5), (3.8), (3.12), and (3.13). Define $\boldsymbol{\Theta}_* = (\boldsymbol{\mu}_*, \boldsymbol{\Phi}_*, \mathbf{V}_*)$, where $*$ $\in \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{h}\}$, and let $\mathbf{d}_a$ denote a vector of $\{d_{ai}\}$. The MCMC algorithm is developed for the full joint posterior $p(\boldsymbol{\Theta}_o, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{d}, \mathbf{d}_a | \mathbf{y}_{1:T})$, where $\boldsymbol{\Theta}_o = (\boldsymbol{\Theta}_\beta, \boldsymbol{\Theta}_\alpha, \boldsymbol{\Theta}_h)$. Standard MCMC algorithms for TV-VAR models are well documented; see, for example, Primiceri (2005), Koop and Korobilis (2010), and Nakajima (2011). The algorithm is based on these standard methods for the new MCMC sampler with the LTM structure as described in the following sampling steps.

A.2.1 Sampling $\boldsymbol{\beta}_{1:T}$ and $\boldsymbol{\alpha}_{1:T}$

Conditional on $(\boldsymbol{\Theta}_\beta, \boldsymbol{\alpha}_{1:T}, \mathbf{h}_{1:T}, \mathbf{d}, \mathbf{y}_{1:T})$, $\boldsymbol{\beta}_t$ is generated using the MH sampler implemented in Appendix A.1.2. Note that the response here is multivariate; the ingredients in the proposal distribution are generalized to

$$\begin{aligned} \mathbf{M}_t^{-1} &= \mathbf{X}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{X}_t + \mathbf{V}_*^{-1} (\mathbf{I} + \boldsymbol{\Phi}_*' \boldsymbol{\Phi}_*), \\ \mathbf{m}_t &= \mathbf{M}_t \left[\mathbf{X}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{y}_t + \mathbf{V}_*^{-1} \left\{ \boldsymbol{\Phi}_* (\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}) + (\mathbf{I} - 2\boldsymbol{\Phi}_* + \boldsymbol{\Phi}_*' \boldsymbol{\Phi}_*) \boldsymbol{\mu}_* \right\} \right], \end{aligned}$$

for $t = 2 : T - 1$, where $*$ = β . For $t = 1$ and $t = T$,

$$\begin{aligned} \mathbf{M}_1^{-1} &= \mathbf{X}_1' \Sigma_1^{-1} \mathbf{X}_1 + \mathbf{V}_0^{-1} + \mathbf{V}_*^{-1} \Phi_*' \Phi_*, \\ \mathbf{m}_1 &= \mathbf{M}_1 [\mathbf{X}_1' \Sigma_1^{-1} \mathbf{y}_1 + \mathbf{V}_0^{-1} \boldsymbol{\mu}_* + \mathbf{V}_*^{-1} \Phi_* \{\beta_2 - (\mathbf{I} - \Phi_*) \boldsymbol{\mu}_*\}], \\ \mathbf{M}_T^{-1} &= \mathbf{X}_T' \Sigma_T^{-1} \mathbf{X}_T + \mathbf{V}_*^{-1}, \\ \mathbf{m}_T &= \mathbf{M}_T [\mathbf{X}_T' \Sigma_T^{-1} \mathbf{y}_T + \mathbf{V}_*^{-1} \{\Phi_* \beta_{T-1} + (\mathbf{I} - \Phi_*) \boldsymbol{\mu}_*\}]. \end{aligned}$$

The MH acceptance probability is given by

$$\alpha(\beta_t, \beta_t^*) = \min \left\{ 1, \frac{N(\mathbf{y}_t | \mathbf{X}_t \mathbf{b}_t^*, \Sigma_t) N(\beta_t | \mathbf{m}_t, \mathbf{M}_t)}{N(\mathbf{y}_t | \mathbf{X}_t \mathbf{b}_t, \Sigma_t) N(\beta_t^* | \mathbf{m}_t, \mathbf{M}_t)} \right\}.$$

The procedure sequentially generates the sample of β_t for $t = 1 : T$ using this MH algorithm.

Conditional on $(\Theta \boldsymbol{\alpha}, \beta_{1:T}, \mathbf{h}_{1:T}, \mathbf{d}_a, \mathbf{y}_{1:T})$, sampling $\boldsymbol{\alpha}_{1:T}$ requires the same MH sampling strategy as $\beta_{1:T}$ based on the model (3.11)-(3.13), with $*$ = $\boldsymbol{\alpha}$, $\mathbf{y}_t = \tilde{\mathbf{y}}_t$, $\mathbf{X}_t = \tilde{\mathbf{X}}_t$, and $\Sigma_t = \Lambda_t$.

A.2.2 Sampling $\mathbf{h}_{1:T}$

Conditional on $(\Theta \mathbf{h}, \beta_{1:T}, \boldsymbol{\alpha}_{1:T}, \mathbf{y}_{1:T})$, defining $\mathbf{y}_t^* = \mathbf{A}_t(\mathbf{y}_t - \mathbf{X}_t \beta_t)$ and $\mathbf{y}_t^* = (y_{1t}^*, \dots, y_{mt}^*)'$ yields a form of univariate stochastic volatility:

$$y_{it}^* = \exp(h_{it}/2) e_{it}, \quad (\text{A.1})$$

$$h_{it} = \mu_{hi} + \phi_{hi}(h_{i,t-1} - \mu_{hi}) + \eta_{hit}, \quad (\text{A.2})$$

$$\begin{pmatrix} e_{it} \\ \eta_{hit} \end{pmatrix} \sim N \left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & v_{hi}^2 \end{pmatrix} \right), \quad (\text{A.3})$$

where μ_{hi} , ϕ_{hi} and v_{hi}^2 are the i -th (diagonal) element of $\boldsymbol{\mu}_h$, Φ_h and $\mathbf{V}_h$, respectively. As in Primiceri (2005) and Nakajima (2011), we can adopt the standard, efficient algorithm for stochastic volatility models (e.g., Kim et al. 1998; Omori et al.

2007; Shephard and Pitt 1997; Watanabe and Omori 2004) for this step. See sampling details in Appendix of Nakajima (2011).

A.2.3 Sampling $\Theta_\circ$ and $(\mathbf{d}, \mathbf{d}_a)$

Conditional on $(\beta_{1:T}, \mathbf{d})$ and $(\alpha_{1:T}, \mathbf{d}_a)$, sampling Θ_β and Θ_α is implemented as in Appendix A.1.1. Conditional on $\mathbf{h}_{1:T}$, sampling Θ_h also follows the same sampling strategy, but it does not require the rejection step associated with the thresholds. Conditional on all other parameters and data, the algorithm generates the latent thresholds $\mathbf{d}$ and $\mathbf{d}_a$ using the sampler described in Appendix A.1.3.

A.3 MCMC algorithm for LTF models

This section describes the MCMC algorithm for the LTF model defined by eqns. (3.14)-(3.24). Define $\Theta_* = (\mu_*, \Phi_*, \mathbf{V}_*)$, where $*$ $\in \{\beta, \zeta, c, h, a, \lambda\}$, and $\mathbf{dz} = (d_{z1}, \dots, d_{z,m\ell})$. The Bayesian computation explores the full posterior distribution

$$p(\Theta_\circ, \beta_{1:T}, \zeta_{1:T}, c_{1:T}, h_{1:T}, a_{1:T}, \lambda_{1:T}, f_{1:T}, g, \mathbf{d}, \mathbf{d}_z | \mathbf{y}_{1:T}),$$

where $\Theta_\circ$ denotes all the Θ_* , and $\mathbf{g} = (g_1, \dots, g_q)'$. The MCMC algorithm consists of the following sampling sequence.

A.3.1 Sampling $\beta_{1:T}$ and $\zeta_{1:T}$

Define $\mathbf{B}_t^{(i)}$ and $\mathbf{Z}_t^{(i)}$ as the i -th row vector of $\mathbf{B}_t$ and $\mathbf{Z}_t$; and define $\beta_t^{(i)}$ and $\zeta_t^{(i)}$ as the corresponding vector of free parameters in $\mathbf{B}_t^{(i)}$ and $\mathbf{Z}_t^{(i)}$, respectively. Conditional on other parameters and data, the diagonal structure of Λ_t allows generation of $\beta_t^{(i)}$ and $\zeta_t^{(i)}$ in parallel for $i = 1, \dots, m$, based on the single LTM-based regression:

$$y_{it} = \mathbf{B}_t^{(i)} \mathbf{f}_t + c_{it} + \mathbf{Z}_t^{(i)} \mathbf{x}_t + \nu_{it}, \quad \nu_{it} \sim N(0, \sigma_{it}^2).$$

The algorithm generates $\beta_t^{(i)}$ and $\zeta_t^{(i)}$ (separately or jointly) using the sampler in Section A.1.2, for $i = 1, \dots, m$. Note that the sequence omits $\beta_t^{(1)}$, because $\mathbf{B}_t^{(i)}$ does not include any free parameter of factor loadings.

A.3.2 Sampling $\mathbf{f}_{1:T}$, $\mathbf{c}_{1:T}$ and $\mathbf{a}_{1:T}$

Conditional on other parameters and data, eqns.(3.14) and (3.15) form a DLM with respect to the latent process $\mathbf{f}_t$, which enables a forward-filtering, backward sampling (FFBS) algorithm (e.g., de Jong and Shephard 1995; Durbin and Koopman 2002; Prado and West 2010). Sampling $\mathbf{c}_{1:T}$ uses the same strategy based on eqns.(3.14) and (3.20). For sampling $\mathbf{a}_{1:T}$, as in eqn.(3.10), the FFBS bases on eqns.(3.15) and (3.22), which yields

$$\tilde{\mathbf{f}}_t = \tilde{\mathbf{H}}_t \mathbf{a}_t + \Psi_t \mathbf{e}_t, \quad \mathbf{e}_t \sim N(\mathbf{0}, \mathbf{I}_q),$$

where $\tilde{\mathbf{f}}_t = (\tilde{f}_{1t}, \dots, \tilde{f}_{qt})' = \mathbf{f}_t - \mathbf{G}_t \mathbf{f}_{t-1}$, and $\tilde{\mathbf{H}}_t$ is $q \times q(q-1)/2$ matrix filled with $\tilde{f}_{it}$ in the same manner as eqn.(3.9). Then, the model forms a DLM with respect to $\mathbf{a}_t$, where the FFBS algorithm can be employed.

A.3.3 Sampling $\mathbf{h}_{1:T}$ and $\boldsymbol{\lambda}_{1:T}$

Conditional on other parameters and data, the generation of $\mathbf{h}_{1:T}$ is based on the univariate stochastic volatility model as in Appendix A.2.2. To derive the form of eqns.(A.1)-(A.3), define $\mathbf{y}_t^* = (y_{1t}^*, \dots, y_{mt}^*)' = \mathbf{y}_t - \mathbf{B}_t \mathbf{f}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{x}_t$. For the factor innovation volatility $\boldsymbol{\lambda}_{1:T}$, the algorithm uses the same sampling strategy, defining $\mathbf{y}_t^* = (y_{1t}^*, \dots, y_{qt}^*)' = \mathbf{A}_t(\mathbf{f}_t - \mathbf{G}_t \mathbf{f}_{t-1})$. The diagonal structure of $(\Phi_*, \mathbf{V}_*)$, $*$ $\in \{\mathbf{h}, \boldsymbol{\lambda}\}$, and conditionally independent priors allow the generation of each univariate stochastic volatility sequence in parallel.

A.3.4 Sampling Θ_o and $(\mathbf{d}, \mathbf{dz})$

Conditional on $(\beta_{1:T}, \mathbf{d})$ and $(\zeta_{1:T}, \mathbf{dz})$, sampling Θ_β and Θ_ζ is implemented as in Appendix A.1.1. Conditional on $\mathbf{c}_{1:T}$, $\mathbf{h}_{1:T}$, $\mathbf{a}_{1:T}$, $\boldsymbol{\lambda}_{1:T}$, sampling Θ_* follows the sampling strategy of the AR model parameters for $* = \mathbf{c}, \mathbf{h}, \mathbf{a}, \boldsymbol{\lambda}$, respectively.

Conditional on all other parameters and data, the algorithm generates the latent thresholds $\mathbf{d}$ and $\mathbf{dz}$ using the sampler described in Appendix A.1.3.

A.3.5 Sampling $\mathbf{g}$

The analysis assumes independent priors for the factor AR coefficient g_i using shifted beta priors to constrain $|g_i| < 1$, for $i = 1, \dots, q$. The algorithm then samples the conditional posteriors via Metropolis-Hastings algorithm.

A.4 Software

Computations were performed using custom code in Ox (Doornik 2006). The code is available to interested readers.

Appendix B

Cholesky-type Time-varying Covariance Matrix

Several analyses in this dissertation rely on the Cholesky decomposition of the time-varying covariance matrix, defined by eqn.(3.5). However, it is not straightforward to theoretically explore compound processes of the covariance/variance elements in the Cholesky-type covariance matrix. This appendix provides some results from a simulation analysis.

Consider the lower triangular matrix $\mathbf{A}_t$ in a simple case, $k = 3$; define $\mathbf{A}_t$ with free (time-varying) parameters $a_{1:3,t}$, and derive $\mathbf{A}_t^{-1}$ as

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & 0 \\ a_{1t} & 1 & 0 \\ a_{2t} & a_{3t} & 1 \end{pmatrix}, \quad \mathbf{A}_t^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -a_{1t} & 1 & 0 \\ a_{1t}a_{3t} - a_{2t} & -a_{3t} & 1 \end{pmatrix}.$$

In the case of $k = 4$, we have

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_{1t} & 1 & 0 & 0 \\ a_{2t} & a_{3t} & 1 & 0 \\ a_{4t} & a_{5t} & a_{6t} & 1 \end{pmatrix}, \quad \mathbf{A}_t^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -a_{1t} & 1 & 0 & 0 \\ a_{1t}a_{3t} - a_{2t} & -a_{3t} & 1 & 0 \\ a_{7t}^* & a_{3t}a_{6t} - a_{5t} & -a_{6t} & 1 \end{pmatrix},$$

where $a_{7t}^* = a_{1t}a_{5t} + a_{2t}a_{6t} - a_{1t}a_{3t}a_{6t} - a_{4t}$. The strictly lower triangular elements of $\mathbf{A}_t^{-1}$ include multiplication and their linear combinations of the process $\{a_{it}\}$. Further, define $\mathbf{\Lambda}_t = \text{diag}(\sigma_{1:3,t}^2)$ in the case of $k = 3$, and obtain the covariance matrix as

$$\mathbf{\Sigma}_t = \mathbf{A}_t^{-1} \mathbf{\Lambda}_t (\mathbf{A}_t')^{-1} = \begin{pmatrix} \sigma_{1t}^2 & \sigma_{12t} & \sigma_{13t} \\ \sigma_{12t} & \sigma_{1t}^2 a_{1t}^2 + \sigma_{2t}^2 & \sigma_{23t} \\ \sigma_{13t} & \sigma_{23t} & \sigma_{1t}^2 (a_{1t}a_{3t} - a_{2t})^2 + \sigma_{2t}^2 a_{3t}^2 + \sigma_{3t}^2 \end{pmatrix},$$

where $\sigma_{12t} = -\sigma_{1t}^2 a_{1t}$, $\sigma_{13t} = \sigma_{1t}^2 (a_{1t}a_{3t} - a_{2t})$, and $\sigma_{23t} = -\sigma_{1t}^2 a_{1t} (a_{1t}a_{3t} - a_{2t}) - \sigma_{2t}^2 a_{3t}$. This implies that both resulting variance and covariance elements are linear combination of multiplied AR and AR-SV processes; complexity increases from top/left to bottom/right of the covariance matrix. As there is no explicit form of the time-series process for these compound variance and covariance elements, a simulation analysis follows.

The simulation generates $\mathbf{a}_t$ and $\mathbf{h}_t$ with for $k = 4$, following eqns.(3.7) and (3.8) with parameters generated from standard priors: $(\phi_{*i} + 1)/2 \sim B(20, 1.5)$, $v_{*i}^{-2} \sim G(2, 0.01)$, for $* \in \{a, h\}$, $\mu_{ai} \sim N(0, 1)$, and $\exp(-\mu_{hi}) \sim G(3, 0.03)$. Figure B.1 plots trajectories of simulated $\mathbf{a}_t$ and $\mathbf{h}_t$ as well as the resulting elements in the covariance matrix $\mathbf{\Sigma}_t$: variances (denoted by S_{it}^2), covariances (S_{ijt}), and correlations (r_{ijt}). The compound processes appear close to the original processes a_{it} and h_{jt} in terms of their time series characteristics.

To explore the characteristics of the compound process, the simulation generates 1,000 sets of the above process for $k = 10$ and $T = 1,000$. Figure B.2 plots autocorrelation functions (ACF) of S_{it}^2 , S_{ijt} and r_{ijt} , averaged across the sets of processes. All ACFs look similar to that of AR(1) process with $\phi_i \approx 0.9$. This finding and results of partial ACFs indicate that the Cholesky-type covariance matrix consists of the time-varying elements whose time series process is similar to the stationary AR(1)

FIGURE B.1: Trajectories of simulated $\mathbf{a}_t$ and $\mathbf{h}_t$ (left), as well as resulting elements in the covariance matrix (right): S_{ijt} (covariance), S_{it}^2 (variance), and r_{ijt} (correlation).

FIGURE B.2: Autocorrelation function of S_{ijt} (covariance), S_{it}^2 (variance), and r_{ijt} (correlation) in the covariance matrix.

process. The prior used here usually yields the time-varying parameter process such that $|a_{it}| < 1$ and $\sigma_{it}^2 < 1$, which implies that the multiplication of the processes shrinks its term, and one single term of the AR(1) process dominates the compound process. Note that the characteristics of the compound process can change when $|a_{it}| \geq 1$ and/or $\sigma_{it}^2 \geq 1$ due to the multiplication of the process.

Bibliography

- Aguilar, O., Prado, R., Huerta, G., and West, M. (1999), “Bayesian inference on latent structure in time series (with discussion),” in *Bayesian Statistics 6*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford: Oxford University Press, pp. 3–26.
- Aguilar, O. and West, M. (2000), “Bayesian dynamic factor models and portfolio allocation,” *Journal of Business and Economic Statistics*, 18, 338–357.
- Awokuse, T. O., Chopra, A., and Bessler, D. A. (2009), “Structural change and international stock market interdependence: Evidence from Asian emerging markets,” *Economic Modelling*, 26, 549–559.
- Baumeister, C., Liu, P., and Mumtaz, H. (2010), “Changes in the transmission of monetary policy: Evidence from a time-varying factor-augmented VAR,” Working Paper No.401, Bank of England.
- Benati, L. (2008), “The “Great Moderation” in the United Kingdom,” *Journal of Money, Credit and Banking*, 40, 121–147.
- Benati, L. and Surico, P. (2008), “Evolving U.S. monetary policy and the decline of inflation predictability,” *Journal of the European Economic Association*, 6, 643–646.
- Bernanke, B., Boivin, J., and Elias, P. (2005), “Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach,” *Quarterly Journal of Economics*, 120, 387–422.
- BIS (2010), *Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in 2010*, Bank for International Settlements.
- Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), “ARCH modeling in finance: A review of the theory and empirical evidence,” *Journal of Royal Statistical Society, Series B*, 52, 5–59.

- Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and 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., and Polson, N. G. (2010), “Particle learning and smoothing,” *Statistical Science*, 25, 88–106.
- Carvalho, C. M., Lopes, H. F., and Aguilar, O. (2011), “Dynamic stock selection strategies: A structured factor model framework (with discussion),” in *Bayesian Statistics 9*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., Oxford University Press, pp. 69–90.
- Carvalho, C. M. and West, M. (2007), “Dynamic matrix-variate graphical models,” *Bayesian Analysis*, 2, 69–98.
- Chan, J. C. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012), “Time varying dimension models,” *Journal of Business and Economic Statistics*, forthcoming.
- Chen, C. W. S., Liu, F., and Gerlach, R. (2011), “Bayesian subset selection for threshold autoregressive moving-average models,” *Computational Statistics*, 26, 1–30.
- Chen, C. W. S., McCulloch, R. E., and Tsay, R. S. (1997), “A unified approach to estimating and modeling linear and nonlinear time-series,” *Statistica Sinica*, 7, 451–472.
- Clyde, M. and George, E. I. (2004), “Model uncertainty,” *Statistical Science*, 19, 81–94.
- Cogley, T. and Sargent, T. J. (2005), “Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.” *Review of Economic Dynamics*, 8, 262–302.
- de Jong, P. and Shephard, N. (1995), “The simulation smoother for time series models,” *Biometrika*, 82, 339–350.
- Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G., and West, M. (2004), “Sparse graphical models for exploring gene expression data,” *Journal of Multivariate Analysis*, 90, 196–212.
- Doornik, J. (2006), *Ox: Object Oriented Matrix Programming*, London: Timberlake Consultants Press.
- Durbin, J. and Koopman, S. J. (2002), “Simple and efficient simulation smoother for state space time series analysis,” *Biometrika*, 89, 603–616.

- Dyro, F. M. (1989), *The EEG Handbook*, Boston: Little, Brown and Co.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2008), “Nonparametric Bayesian learning of switching linear dynamical systems,” in *Proceeding of Neural Information Processing Systems*.
- Frauendorfer, K., Jacoby, U., and Schwendener, A. (2007), “Regime switching based portfolio selection for pension funds,” *Journal of Banking and Finance*, 31, 2265–2280.
- Fujita, A., Sato, J. R., Garay-Malpartida, H. M., Morettin, P. A., Sogayar, M. C., and Ferreira, C. E. (2007), “Time-varying modeling of gene expression regulatory networks using the wavelet dynamic vector autoregressive method,” *Bioinformatics*, 23, 1623–1630.
- Galvao, A. B. and Marcellino, M. (2010), “Endogenous monetary policy regimes and the Great Moderation,” Working Paper, 2010/22, European University Institute.
- George, E. I. and McCulloch, R. E. (1993), “Variable selection via Gibbs sampling,” *Journal of the American Statistical Association*, 88, 881–889.
- (1997), “Approaches for Bayesian variable selection,” *Statistica Sinica*, 7, 339–373.
- George, E. I., Sun, D., and Ni, S. (2008), “Bayesian stochastic search for VAR model restrictions,” *Journal of Econometrics*, 142, 553–580.
- Granger, C. W. J. (1969), “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, 37, 424–438.
- Gulko, L. (2002), “Decoupling,” *Journal of Portfolio Management*, 28, 59–66.
- Harper, R. B. (2003), “Asset allocation, decoupling, and the opportunity cost of cash,” *Journal of Portfolio Management*, 29, 25–35.
- Huerta, G. and West, M. (1999), “Priors and component structures in autoregressive time series models,” *Journal of the Royal Statistical Society, Series B*, 61, 881–899.
- Iwata, S. and Wu, S. (2006), “Estimating monetary policy effects when interest rates are close to zero,” *Journal of Monetary Economics*, 53, 1395–1408.
- Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), “Bayesian analysis of stochastic volatility models,” *Journal of Business and Economic Statistics*, 12, 371–389.
- Jones, B., Carvalho, C. M., Dobra, A., Hans, C., Carter, C., and West, M. (2005), “Experiments in stochastic computation for high-dimensional graphical models,” *Statistical Science*, 20, 388–400.

- Kim, C.-J. and Nelson, C. (1999), *State-Space Models with Regime Switching*, Cambridge, MA: MIT Press.
- Kim, J. and Cheon, S. (2010), “A Bayesian regime-switching time series model,” *Journal of Time Series Analysis*, 31, 365–378.
- Kim, S., Shephard, N., and Chib, S. (1998), “Stochastic volatility: Likelihood inference and comparison with ARCH models,” *Review of Economic Studies*, 65, 361–393.
- Kitagawa, G. and Gersch, W. (1996), *Smoothness priors analysis of time series*, Lecture Notes in Statistics, Vol.116, New York: Springer-Verlag.
- Koop, G. and Korobilis, D. (2010), “Bayesian multivariate time series methods for empirical macroeconomics,” *Foundations and Trends in Econometrics*, 3, 267–358.
- Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2009), “On the evolution of the monetary policy transmission mechanism,” *Journal of Economic Dynamics and Control*, 33, 997–1017.
- Korobilis, D. (2012), “VAR forecasting using Bayesian variable selection,” *Journal of Applied Econometrics*, forthcoming.
- Lauritzen, S. L. (1996), *Graphical Models*, Oxford: Clarendon Press.
- Lopes, H. F. and 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., and Tsay, R. (2012), “Cholesky stochastic volatility models for high-dimensional time series,” Tech. rep., University of Chicago, Booth Business School.
- Lopes, H. F. and West, M. (2004), “Bayesian model assessment in factor analysis,” *Statistica Sinica*, 14, 41–67.
- Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, New York: John Wiley and Sons.
- Nakajima, J. (2011), “Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications,” *Monetary and Economic Studies*, 29, 107–142.
- Nakajima, J., Kasuya, M., and Watanabe, T. (2011), “Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy,” *Journal of the Japanese and International Economies*, 25, 225–245.

- Nakajima, J., Shiratsuka, S., and Teranishi, Y. (2010), “The effects of monetary policy commitment: Evidence from time-varying parameter VAR analysis,” IMES Discussion Paper, 2010-E-6, Bank of Japan.
- Nakajima, J. and West, M. (2012a), “Bayesian analysis of latent threshold dynamic models,” *Journal of Business and Economic Statistics*, forthcoming.
- (2012b), “Dynamic factor volatility modeling: A Bayesian latent threshold approach,” *Journal of Financial Econometrics*, doi:10.1093/jjfines/nbs013.
- (2012c), “Evaluating EEG dynamics: An approach using latent threshold factor modelling,” Manuscript, Department of Statistical Science, Duke University.
- (2012d), “Latent threshold models: Multivariate time series and dynamic networks,” Manuscript, Department of Statistical Science, Duke University.
- Niemi, J. B. and West, M. (2010), “Adaptive mixture modelling Metropolis methods for Bayesian analysis of non-linear state-space models,” *Journal of Computational and Graphical Statistics*, 19, 260–280.
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), “Stochastic volatility with leverage: Fast likelihood inference,” *Journal of Econometrics*, 140, 425–449.
- Onnela, J.-P., Kaski, K., and Kertesz, J. (2004), “Clustering and information in correlation based financial networks,” *European Physical Journal B*, 38, 353–362.
- Pinheiro, J. C. and Bates, D. M. (1996), “Unconstrained parametrizations for variance-covariance matrices,” *Statistics and Computing*, 6, 289–296.
- Pitt, M. and Shephard, N. (1999), “Filtering via simulation: Auxiliary particle filter,” *Journal of the American Statistical Association*, 94, 590–599.
- Polasek, W. and Krause, A. (1994), “The hierarchical Tobit model: A case study in Bayesian computing,” *OR Spectrum*, 16, 145–154.
- Pourahmadi, M. (1999), “Joint mean-covariance models with application to longitudinal data: Unconstrained parametrisation,” *Biometrika*, 86, 677–690.
- Prado, R. (1998), “Latent Structure in Non-stationary Time Series,” Ph.D. thesis, Duke University, Durham, N.C.
- (2010a), “Characterization of latent structure in brain signals,” in *Statistical Methods for Modeling Human Dynamics*, eds. Chow, S., Ferrer, E., and Hsieh, F., New York: Routledge, Taylor and Francis, pp. 123–153.

- (2010b), “Multi-state models for mental fatigue,” in *The Handbook of Applied Bayesian Analysis*, eds. O’Hagan, A. and West, M., Oxford University Press, pp. 845–874.
- Prado, R. and Huerta, G. (2002), “Time-varying autoregressions with model order uncertainty,” *Journal of Time Series Analysis*, 23, 599–618.
- Prado, R. and West, M. (2010), *Time Series Modeling, Computation, and Inference*, New York: Chapman & Hall/CRC.
- Prado, R., West, M., and Krystal, A. D. (2001), “Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag-lead structure,” *Journal of the Royal Statistical Society, Series C*, 50, 95–109.
- Primiceri, G. E. (2005), “Time varying structural vector autoregressions and monetary policy,” *Review of Economic Studies*, 72, 821–852.
- Putnum, B. H. and 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 *Bayesian Statistics IV*, eds. Berger, J. O., Bernardo, J. M., Dawid, A. P., and Smith, A. F. M., Oxford University Press, pp. 753–762.
- Quintana, J. M., Carvalho, C. M., Scott, J., and Costigliola, T. (2010), “Futures markets, Bayesian forecasting and risk modeling,” in *The Handbook of Applied Bayesian Analysis*, eds. O’Hagan, A. and West, M., Oxford: Oxford University Press, pp. 343–365.
- Quintana, J. M., Lourdes, V., Aguilar, O., and Liu, J. (2003), “Global gambling,” in *Bayesian Statistics VII*, eds. Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., Oxford University Press, pp. 349–368.
- Quintana, J. M. and Putnum, B. H. (1996), “Debating currency markets efficiency using dynamic multiple-factor models,” in *Proceedings of the ASA Section on Bayesian Statistical Science*, American Statistical Association, pp. 55–60.
- Quintana, J. M. and West, M. (1987), “An analysis of international exchange rates using multivariate DLM’s,” *Statistician*, 36, 275–281.
- Shephard, N. (2005), *Stochastic Volatility: Selected Readings*, Oxford: Oxford University Press.

- Shephard, N. and Pitt, M. (1997), “Likelihood analysis of non-Gaussian measurement time series,” *Biometrika*, 84, 653–667.
- Smith, M. and Kohn, R. (2002), “Parsimonious covariance matrix estimation for longitudinal data,” *Journal of the American Statistical Association*, 97, 1141–1153.
- Suchard, M. A., Holmes, C., and West, M. (2010a), “Some of the *What?*, *Why?*, *How?*, *Who?* and *Where?* of graphics processing unit computing for Bayesian analysis,” *Bulletin of the International Society for Bayesian Analysis*, 17, 12–16.
- Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A. J., and West, M. (2010b), “Understanding GPU programming for statistical computation: Studies in massively parallel massive mixtures,” *Journal of Computational and Graphical Statistics*, 19, 419–438.
- Tse, C. K., Liu, J., and Lau, F. C. M. (2010), “A network perspective of the stock market,” *Journal of Empirical Finance*, 17, 659–667.
- Valdés-Sosa, P. A. (2004), “Spatio-temporal autoregressive models defined over brain manifolds,” *Neuroinformatics*, 2, 239–250.
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., and Canales-Rodríguez, E. (2005), “Estimating brain functional connectivity with sparse multivariate autoregression,” *Philosophical Transactions of the Royal Society B*, 360, 969–981.
- Vandewalle, N., Brisbois, F., and Tordoir, X. (2001), “Self-organized critical topology of stock markets,” *Quantitative Finance*, 1, 372–375.
- Wang, H. (2010), “Sparse seemingly unrelated regression modelling: Applications in finance and econometrics,” *Computational Statistics and Data Analysis*, 54, 2866–2877.
- Wang, H. and West, M. (2009), “Bayesian analysis of matrix normal graphical models,” *Biometrika*, 96, 821–834.
- Watanabe, T. and Omori, Y. (2004), “A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard and Pitt (1997),” *Biometrika*, 91, 246–248.
- Wei, S. X. (1999), “A Bayesian approach to dynamic Tobit models,” *Econometric Reviews*, 18, 417–439.
- Weiner, R. D. and Krystal, A. D. (1994), “The present use of electroconvulsive therapy,” *Annual Review of Medicine*, 45, 273–281.

- West, M. (2003), “Bayesian factor regression models in the “large p, small n” paradigm,” in *Bayesian Statistics 7*, eds. Bernardo, J., Bayarri, M., Berger, J., David, A., Heckerman, D., Smith, A., and West, M., Oxford, pp. 723–732.
- (2013), “Bayesian dynamic modelling,” in *Bayesian Theory and Applications*, eds. Damien, P., Dellaportes, P., Polson, N. G., and Stephens, D. A., Clarendon: Oxford University Press, chap. 8.
- West, M. and Harrison, P. J. (1997), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag, 2nd ed.
- West, M. and Mortera, J. (1987), “Bayesian models and methods for binary time series,” in *Probability and Bayesian Statistics*, ed. Viertl, R., Elsevier, Amsterdam, pp. 487–496.
- West, M., Prado, R., and Krystal, A. D. (1999), “Evaluation and comparison of EEG traces: Latent structure in nonstationary time series,” *Journal of the American Statistical Association*, 94, 375–387.
- Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, Chichester, United Kingdom: John Wiley and Sons.
- Xia, Y. and Tong, H. (2011), “Feature matching in time series modeling,” *Statistical Science*, 26, 21–46.
- Yoshida, R., Imoto, S., and Higuchi, T. (2005), “Estimating time-dependent gene networks from time series microarray data by dynamic linear models with Markov switching,” in *Proceedings of IEEE 4th Computational Systems Bioinformatics Conference*, pp. 289–298.
- Yoshida, R. and West, M. (2010), “Bayesian learning in sparse graphical factor models via annealed entropy,” *Journal of Machine Learning Research*, 11, 1771–1798.
- Zhou, X., Nakajima, J., and West, M. (2012), “Dynamic dependent factor models: Improving forecasts and portfolio decisions in financial time series,” Working Paper 12-09, Department of Statistical Science, Duke University.

Biography

Jouchi Nakajima was born in Tokyo, Japan. He received his bachelor degree in Economics from University of Tokyo in 2004. He received a M.A. degree in Economics from University of Tokyo in 2006, and a M.S. degree in Statistical Science from Duke University in 2011. Before coming to Duke University, Jouchi was an economist at the Bank of Japan.