Bayesian Analysis of Latent Threshold Dynamic Models

Jouchi Nakajima & Mike West
Department of Statistical Science, Duke University, Durham, NC 27708-0251
{*jouchi.nakajima, mw*}*@stat.duke.edu*

Original report: January 2011
This final version: August 2012
Revised paper published: Journal of Business and Economic Statistics 31(2): 151-164, 2013
(Corrections to two minor typos in Supplementary Material/Appendix A: December 2013)

Abstract

We discuss a general approach to dynamic sparsity modeling in multivariate time series analysis. Time-varying parameters are linked to latent processes that are thresholded to induce zero values adaptively, providing natural mechanisms for dynamic variable inclusion/selection. We discuss Bayesian model specification, analysis and prediction in dynamic regressions, time-varying vector autoregressions and multivariate volatility models using latent thresholding. Application to a topical macroeconomic time series problem illustrates some of the benefits of the approach in terms of statistical and economic interpretations as well as improved predictions.

KEY WORDS: Dynamic graphical models; Macroeconomic time series; Multivariate volatility; Sparse time-varying VAR models; Time-varying variable selection.

The original version of this paper was presented at the 2011 Seminar on Bayesian Inference in Econometrics and Statistics (SBIES), dedicated to the memory of Arnold Zellner, at Washington University, St. Louis, April 29-30, 2011. As noted there: In addressing Bayesian modelling and analysis issues in multivariate time series for econometric and financial applications, this work speaks to areas that Arnold Zellner contributed substantially to over many years. Further, Zellner was always concerned that modelling innovations be squarely focused on improvements in predictive ability and validity, which is a prime motivation for the work we report.

The research reported here was partly supported by a grant from the National Science Foundation [DMS-1106516 to M.W.]. Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

1 Introduction

For analysis of increasingly high-dimensional time series in many areas, 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. We address these general questions with a new and broadly applicable approach 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. As we demonstrate through studies of topical macroeconomic time series, this *dynamic sparsity* modeling concept can be implemented in broad classes of multivariate time series models, and has the 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 in multivariate models. 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; Lopes et al. 2010; Yoshida and West 2010), graphical modeling (Jones et al. 2005), and traditional time series models (e.g. George et al. 2008; Chen et al. 2010). In both dynamic regression and multivariate volatility models, 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 (e.g. Carvalho and West 2007; George et al. 2008; Korobilis 2011; Wang 2010). We build on this idea but address 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.

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.

To introduce the LTM ideas we focus first on dynamic regression models. This context also illustrates how the LTM idea applies to richer classes of dynamic models (West and Harrison 1997; Prado and West 2010; West 2012). We then develop the methodology in time-varying vector autoregressive (TV-VAR) models, including models for multivariate volatility (Aguilar and West 2000). Examples in macroeconomic time series analysis illustrate the approach, as well as highlighting connections with other approaches, open questions and potential extensions.

Some notation: We use f, g, p for density functions, and the distributional notation $d \sim U(a, b)$, $p \sim B(a, b)$, $v \sim G(a, b)$, for the uniform, beta, and gamma distributions, respectively. Further notation includes the following: $\Phi(\cdot)$ is the standard normal cdf; $\otimes$ denotes Kronecker product; $\circ$

stands for element-wise product, e.g, the $k-$vector $\boldsymbol{x} \circ \boldsymbol{y}$ has elements $x_i y_i$; and $s : t$ stands for indices $s, s+1, \ldots, t$ when $s < t$.

2 Latent threshold modeling

2.1 Dynamic regression model: A key example context

The ideas are introduced and initially developed in the canonical class of dynamic regression models (e.g. West and Harrison 1997, chap. 2 & 4), a subclass of dynamic linear models (DLMs). Suppose a univariate time series $\{y_t,\ t = 1, 2, \ldots\}$ follows the model

$$y_t \;=\; \boldsymbol{x}_t' \boldsymbol{b}_t + \varepsilon_t, \quad \varepsilon_t \sim N(\varepsilon_t | 0, \sigma^2), \quad \boldsymbol{b}_t = (b_{1t}, \ldots, b_{kt})', \tag{1}$$

$$b_{it} \;=\; \beta_{it} s_{it} \quad \text{with} \quad s_{it} = I(|\beta_{it}| \geq d_i), \quad i = 1, \ldots, k, \tag{2}$$

where $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{kt})'$ is a $k \times 1$ vector of predictors, $\boldsymbol{d} = (d_1, \ldots, d_k)$ is a *latent threshold* vector with $d_i \geq 0$, for $i = 1, \ldots, k$, and $I(\cdot)$ denotes the indicator function. The time-varying coefficients $\boldsymbol{b}_t$ are governed by the underlying *latent time-varying parameters* $\boldsymbol{\beta}_t \equiv (\beta_{1t}, \ldots, \beta_{kt})'$ and the indicators $\boldsymbol{s}_t \equiv (s_{1t}, \ldots, s_{kt})'$ via $\boldsymbol{b}_t = \boldsymbol{\beta}_t \circ \boldsymbol{s}_t$, subject to some form time series model for the $\boldsymbol{\beta}_t$ process. The i^{th} variable x_{it} has time-varying coefficient whose value is shrunk to zero when it falls below a threshold, embodying sparsity/shrinkage and parameter reduction when relevant; see Figure 1. The shrinkage region $(-d_i, d_i)$ defines temporal variable selection; only when β_{it} is large enough does x_{it} play a role in predicting y_t. The relevance of variables is dynamic; x_{it} may have non-zero coefficient in some time periods but zero in others, depending on the data and context. The model structure is thus flexible in addressing *dynamic regression model uncertainty*.

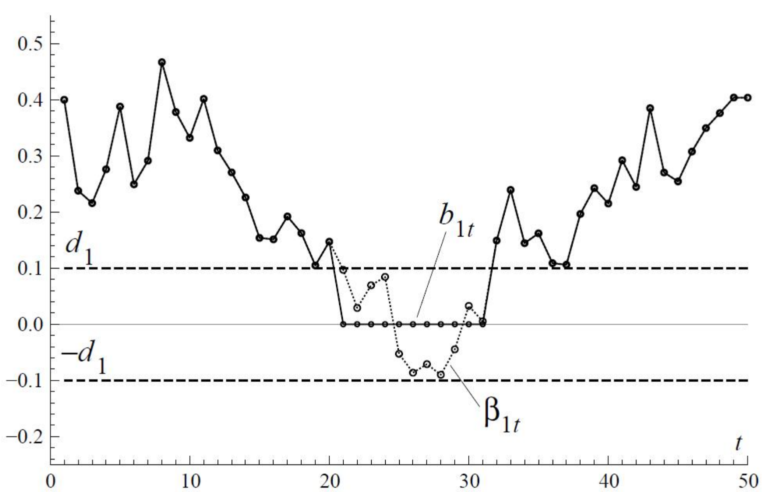


Figure 1: Illustration of LTM concept: The dynamic regression coefficient process β_{1t} arises as a thresholded version of an underlying dynamic coefficient time series.

Any form of model may be defined for $\boldsymbol{\beta}_t$; one of the simplest, and easily most widely useful in practice, is the vector autoregressive (VAR) model that we will use. That is,

$$\boldsymbol{\beta}_{t+1} \;=\; \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\beta}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \tag{3}$$

a VAR(1) model with individual AR parameters ϕ_i in the $k \times k$ matrix $\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_k)$, independent innovations $\boldsymbol{\eta}_t$ and innovation variance matrix $\boldsymbol{\Sigma}_\eta = \mathrm{diag}(\sigma_{1\eta}^2, \ldots, \sigma_{k\eta}^2)$. With $|\phi_i| < 1$ this defines stationary models with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)'$ and univariate margins

$$\beta_{it} \sim N(\mu_i, v_i^2), \quad v_i^2 = \sigma_{i\eta}^2/(1-\phi_i^2). \tag{4}$$

For hyper-parameter notation, we define $\boldsymbol{\theta}_i = (\mu_i, \phi_i, \sigma_i)$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\}$.

2.2 Threshold parameters and sparsity

We discuss the role and specification of priors on threshold parameters, using the dynamic regression model with $k = 1$ for clarity; the discussion applies to thresholding of time-varying parameters in other models by obvious extension. This simple model is now $y_t = b_{1t}x_{1t} + \varepsilon_t$ where b_{1t} is the thresholded version of a univariate AR(1) β_{1t} process.

2.2.1 Thresholding as Bayesian variable selection

Latent thresholding is a direct extension of standard Bayesian variable selection to the time series context. Standard Bayesian model selection methods assign non-zero prior probabilities 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 $\Pr(b_{1t} = 0)$ coupled with a continuous prior on non-zero values, but that also respects the need to induce dependencies in the b_{1t} being zero/non-zero over time. A highly positively dependent AR model for β_{1t} respects relatively smooth variation over time, while the thresholding mechanism induces persistence over periods of time in the occurrences of zero/non-zero values in the effective coefficients b_{1t}. 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_1 in defining the marginal probability $\Pr(b_{1t} = 0)$– the key *sparsity prior parameter* analogous to a prior variable exclusion probability in standard Bayesian variable selection.

2.2.2 Prior sparsity probabilities and thresholds

In the dynamic regression model with $k = 1$ and $\mu_1 = 0$, reflecting a prior centered at the "null hypothesis" of no regression relationship between y_t and x_{1t}, set $\sigma = 1$ with no loss of generality. At each instant t, marginalizing over β_{1t} under eqn. (4) yields $\pi_1 = \Pr(|\beta_{1t}| \geq d_1) = 2\Phi(-d_1/v_1)$ where Φ is the standard normal cdf. The fundamental sparsity probability is $\Pr(b_{1t} = 0) = 1 - \pi_1 = 2\Phi(d_1/v_1) - 1$. This indicates the importance of the scale v_1 in considering relevant values of the threshold d_1; standardizing on this scale, we have $\Pr(b_{1t} = 0) = 2\Phi(k_1) - 1$ where $d_1 = k_1 v_1$. For example, a context where we expect about 5% of the values to be thresholded to zero implies $k_1 = 0.063$ and $d_1 = 0.063v_1$; a context where we expect much higher dynamic sparsity with, say, 90% thresholding implies $k_1 = 1.65$ and a far higher threshold d_1; and a value of $k_1 = 3$ or above leads to a marginal sparsity probability exceeding 0.99. In practice, we assign priors over the threshold based on this reasoning. A neutral prior will support a range of sparsity values in order to allow the data to inform on relevant values; the above indicates a relevant range $0 < d_1 = k_1 v_1 < K v_1$ for some K value well into the upper tail of the standard normal. Unless a context involves substantive information to suggest favoring smaller or larger degrees of expected sparsity, a uniform prior across this range is the natural default, i.e., $k_1 \sim U(0, K)$ for specified K.

In the general model with multiple regression parameter processes β_{it} and mean parameters μ_i, this prior specification extends to the thresholds d_i as follows and assuming independence across thresholds. Conditional on the hyper-parameters $\boldsymbol{\theta}_i = (\mu_i, \phi_i, \sigma_i)$ underlying the stationary margins of eqn. (4),

$$d_i|\boldsymbol{\theta}_i \sim U(0, |\mu_i| + K v_i)$$

for a given upper level K. As noted above, taking $K = 3$ or higher spans essentially the full range of prior sparsity probabilities implied, and in a range of studies we find no practical differences in results based on $K = 3$ relative to higher values; hence $K = 3$ is recommended as a default.

Importantly, we note that when combined with priors over the model hyper-parameters $\boldsymbol{\theta}_i$ the implied *marginal prior* for each threshold will not be uniform, reflecting the inherent relevance of the scale of variation of the β_{it} processes in constraining the priors. The marginal prior on each threshold is

$$p(d_i) = \int U(d_i|0, |\mu_i| + Kv_i)p(\mu_i, v_i)d\mu_i dv_i \tag{5}$$

with, typically, normal and inverse gamma priors on the μ_i and v_i^2, respectively. The first example on simulated data below provides illustration with priors for each of three thresholds displayed in Figure 3. In each case the underlying scales of the β_{it} processes are $v_i = 0.7$, we have $K = 3$ and see how the induced marginal priors have non-zero density values at zero and then decay slowing over larger values of d_i, supporting the full range of potential levels of sparsity. The corresponding prior means of the resulting marginal sparsity probabilities are $\Pr(b_{it} = 0) \approx 0.85$ in this example.

2.2.3 Posteriors on thresholds and inferences on sparsity

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 for primary time-varying parameters and for the resulting implications for predictive distributions. Correspondingly, there is no interest in the underlying values of the latent β_{it} when they are below threshold. Depending on the model and data context, any one of the threshold parameters may be well-estimated, with the posterior being precise relative to the prior; in contrast, the data may be uninformative about other thresholds, with posteriors that are close to their priors. In the dynamic regression example, a data set in which there is little evidence of a regression relationship between y_t and x_{it}, in the context of other regressors, will lead to high levels of thresholding on β_{it}; the posterior will suggest larger values of d_i, providing no information about the β_{it} process but full, informative inference on the effective coefficient process b_{it}. At the other extreme, a strong relationship sustained over time is consistent with a low threshold and the posterior will indicate such.

Figure 3 from the simulated data example illustrates this. For 2 of the regressors, the latent coefficient processes exceed thresholds for reasonable periods of time, and the posterior shows strong evidence for low threshold values. The third coefficient process stays below threshold, and the posterior on the threshold d_3 is appropriately close to the prior, indicating the lack of information in the data about the actual value of d_3. The further details of that example demonstrate the key point that the model analysis appropriately identifies the periods of non-zero effective coefficients and their values, in the posterior for the b_{it} processes and the corresponding posterior estimates of sparsity probabilities at each time point. These inferences, and follow-on predictions, are marginal with respect to thresholds; the posteriors for primary model parameters integrate over the posterior uncertainties about thresholds, their likely values being otherwise of no interest. This viewpoint is echoed in our macroeconomic studies, below, and other applications. The applied interests lie in the practical benefits of the threshold mechanism as a technical device to induce dynamic sparsity/parsimony, hence avoid over-fitting and, as a general result, improve predictions.

2.3 Outline of Bayesian computation

Model fitting using Markov chain Monte Carlo (MCMC) methods involves extending traditional 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 $\boldsymbol{y}_{1:T} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$ over a given time period of T intervals, we are interested in MCMC for simulation of the full joint posterior $p(\boldsymbol{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \boldsymbol{d}|\boldsymbol{y}_{1:T})$. We outline components of the MCMC computations here, and provide additional details in Appendix A, available as on-line Supplementary Material.

First, sampling the TV-VAR model parameters $\boldsymbol{\Theta}$ conditional on $(\boldsymbol{\beta}_{1:T}, \boldsymbol{d}, \sigma, \boldsymbol{y}_{1:T})$ is standard, reducing to the set of conditionally independent posteriors $p(\boldsymbol{\theta}_i|\beta_{i,1:T}, d_i)$. Traditional priors for $\boldsymbol{\theta}_i$ can be used, as can standard Metropolis-Hastings (MH) methods of sampling parameters in univariate AR models.

The second MCMC component is new, addressing the key issue of sampling the latent state process $\boldsymbol{\beta}_{1:T}$. Conditional on $(\boldsymbol{\Theta}, \sigma, \boldsymbol{d})$, we adopt a direct Metropolis-within-Gibbs sampling strategy for simulation of $\boldsymbol{\beta}_{1:T}$. This sequences through each t, using a MH sampler for each $\boldsymbol{\beta}_t$ given $\boldsymbol{\beta}_{-t} = \boldsymbol{\beta}_{1:T} \backslash \boldsymbol{\beta}_t$. Note that the usual dynamic model without thresholds formally arises by fixing $\boldsymbol{s}_t = \boldsymbol{1}$; in this context, the resulting conditional posterior at time t is $N(\boldsymbol{\beta}_t|\boldsymbol{m}_t, \boldsymbol{M}_t)$, where

$$\begin{aligned} \boldsymbol{M}_t^{-1} &= \sigma^{-2}\boldsymbol{x}_t\boldsymbol{x}_t' + \boldsymbol{\Sigma}_\eta^{-1}(\boldsymbol{I} + \boldsymbol{\Phi}'\boldsymbol{\Phi}), \\ \boldsymbol{m}_t &= \boldsymbol{M}_t\left[\sigma^{-2}\boldsymbol{x}_t y_t + \boldsymbol{\Sigma}_\eta^{-1}\left\{\boldsymbol{\Phi}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}) + (\boldsymbol{I} - 2\boldsymbol{\Phi} + \boldsymbol{\Phi}'\boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right]. \end{aligned}$$

For $t = 1$ and $t = T$ a slight modification is required, with details in Appendix A. The MH algorithm uses this as proposal distribution to generate a candidate $\boldsymbol{\beta}_t^*$ for accept/reject assessment. 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 $\boldsymbol{\beta}_t$ are large, and smaller elements in candidate draws will in any case 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

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

where $\boldsymbol{b}_t = \boldsymbol{\beta}_t \circ \boldsymbol{s}_t$ is the current LTM state at t and $\boldsymbol{b}_t^* = \boldsymbol{\beta}_t^* \circ \boldsymbol{s}_t^*$ the candidate.

It is possible to develop a block sampling extension of this MH strategy by using a forward-filtering, backward sampling (FFBS) algorithm (e.g. Frühwirth-Schnatter 1994; de Jong and Shephard 1995; Prado and West 2010) on the non-thresholded model to generate proposals for the full sequence $\boldsymbol{\beta}_{1:T}$. This follows related approaches using this idea of a global FFBS-based proposal (Prado and West 2010; Niemi and West 2010). The main drawback is that the resulting acceptance rates decrease exponentially with T and our experiences 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 LTM posterior. We have experimented with a modified multi-block approach in which FFBS is applied separately within a block conditional on state vectors in all other blocks, but with limited success in improving acceptance rates. Although the simpler single-move strategy has the potential to mix less slowly, it is computationally efficient and so can be run far longer than blocking approaches to balance mixing issues; we have experienced practically satisfactory rates of acceptance and mixing in multiple examples, and hence adopt it as standard.

The TV-VAR coefficients can, at any time point, define VAR structures with explosive autoregressive roots, depending on their values. In our macroeconomic examples, this is unrealistic and undesirable (e.g., Cogley and Sargent 2001) and can be avoided by restricting to the stationary region. We apply this restriction in the single-move sampler for $\boldsymbol{\beta}_t$, adding a rejection sampling step in generating the candidates, noting that this is trivial compared to adapting a multi-move sampler like the FFBS to address this problem.

The final MCMC component required is the generation of thresholds $\boldsymbol{d}$. We adopt a direct MH approach with candidates drawn from the prior. The simulated example of the next section illustrates this. Our experiences with the direct MCMC summarized here are 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, only additional computational burden.

3 Simulation example

A sample of size $T = 500$ was drawn from a dynamic regression LTM with $k = 3$ 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 $\sigma = 0.15$, while for $i = 1, 2$ we take parameters $(\mu_i, \phi_i, \sigma_{i\eta}, d_i) = (0.5, 0.99, 0.1, 0.4)$; for $i = 3$, $\beta_{3t} = 0$ for all t. Figure 2 graphs the true values of the time-varying coefficients, indicating the within-threshold sparsity periods by shading. The following prior distributions are used: $\mu_i \sim N(0, 1)$, $(\phi_i + 1)/2 \sim B(20, 1.5)$, $\sigma_{i\eta}^{-2} \sim G(3, 0.03)$, 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 $\sigma_{i\eta}^2$ and σ^2 are $(0.015, 0.015)$. The conditional prior for thresholds is $U(d_i|0, |\mu_i| + Kv_i)$ with $K = 3$. MCMC used $J = 50{,}000$ iterates after burn-in of 5,000. Computations use Ox (Doornik 2006) and code is available to interested readers.

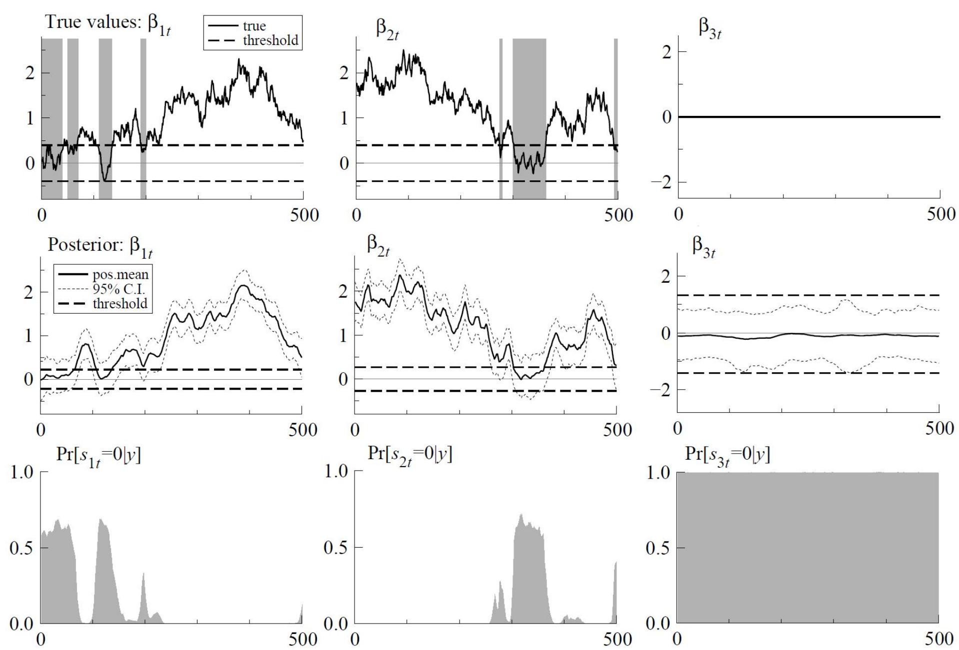


Figure 2: Simulation example: Trajectories of dynamic regression parameters. True values (top), posterior means, 95% credible intervals (middle), and posterior probabilities of $s_{it} = 0$ (bottom). The shadows in the top panels refer to the periods when $s_{it} = 0$. The thresholds in the middle panels refer to their posterior means.

Figure 2 shows some summary results. The posterior means of the time-varying coefficients trace their true values, and the posterior probabilities of $s_{it} = 0$ successfully detect the temporal

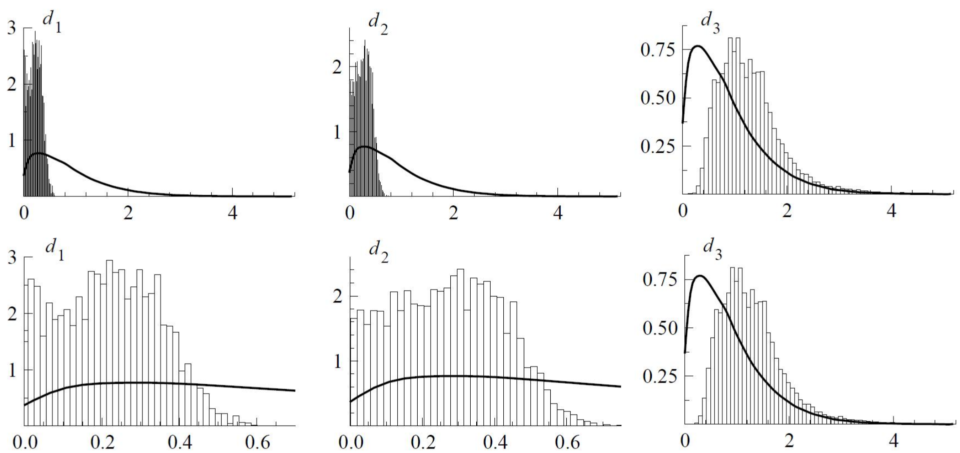


Figure 3: Simulation example: Priors (solid lines) and posteriors (histograms) for thresholds. In the lower row, the graphs are simply truncated on the d_i axes for visual clarity.

	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
$\sigma_{1,\eta}$	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

Table 1: Simulation example: Posterior estimates for selected parameters with credible intervals based on 2.5%, 97.5%- quantiles of posterior MCMC draws.

sparsity for β_{1t} and β_{2t} as well as the whole-sequence variable selection for β_{3t}. Table 1 reports the posterior estimates for selected parameters; posterior means are close enough to the true values that the corresponding 95% credible intervals include them. Figure 3 displays priors, from eqn. (5), and resulting posteriors for the thresholds. Repeat analyses with larger and moderately smaller values of K yield substantially similar inferences on the dynamic state vectors and their sparsity patterns over time, as well as for the underlying AR model parameters. Even taking lower values such as $K = 1$ has limited effect on our main findings. As discussed in detail above, there is no inherent interest in inferences on thresholds themselves; the interest is in their roles as defining the ability to shrink parameters when the data support sparsity. Hence the expected differences in priors and posteriors for d_i as we change K are of limited interest so long as the posteriors for regression states and AR parameters remain stable. Taking K smaller than 1 or so does begin to more substantially impact on primary inferences, inducing less sparse models in general. In some applications, this may be of positive relevance, but for general application we adopt $K = 3$ as a global default. Further computational details, including convergence checks and performance of the MCMC sampler, appear in Appendix B of the on-line Supplementary Material.

4 Latent threshold time-varying VAR models

We now consider the latent threshold strategy in multivariate time series analysis using time-varying parameter vector autoregressive (TV-VAR) models. Traditional, constant coefficient VAR models are of course central to applied time series analysis (e.g. Prado and West 2010, and references therein), and various approaches to TV-VAR modeling are becoming increasingly prevalent in econometrics (e.g. Cogley and Sargent 2005; Primiceri 2005) as in other fields. Recent studies have introduced sparsity-inducing priors of various forms (Fox et al. 2008; George et al. 2008; Wang 2010) into traditional constant coefficient VAR models. 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 an open and challenging problem. The LTM ideas provide an approach.

4.1 Model structure

For the $m \times 1$-vector time series $\boldsymbol{y}_t$, $(t = 1, 2, \ldots)$, the TV-VAR(p) model is

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

where $\boldsymbol{c}_t$ is the $m \times 1$ vector of time-varying intercepts, $\boldsymbol{B}_{jt}$ is the $m \times m$ matrix of time-varying coefficients at lag j, $(j = 1, \ldots, p)$, and $\boldsymbol{\Sigma}_t$ is the $m \times m$ innovations variance matrix that is also often time-varying. For each t, define the $m(1 + pm) \times 1$ vector $\boldsymbol{b}_t$ by stacking the set of $\boldsymbol{c}_t$ and $\boldsymbol{B}_{jt}$ by rows and by order $j = 1, \ldots, p$; define the corresponding $m \times m(1 + pm)$ matrix $\boldsymbol{X}_t = \boldsymbol{I} \otimes (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})$. The model can then be recast as a multivariate dynamic regression

$$\boldsymbol{y}_t \;=\; \boldsymbol{X}_t\boldsymbol{b}_t + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim N(\boldsymbol{u}_t|\mathbf{0}, \boldsymbol{\Sigma}_t). \tag{6}$$

The time-varying coefficient vector $\boldsymbol{b}_t$ is often assumed to follow a VAR(1) process, the simplest and often most useful model. We begin with this and then generalize to the LTM framework by overlaying thresholds as in Section 2.1, eqns. (2)-(3). We refer to this specification 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.

Posterior simulation of the full sequence $\boldsymbol{\beta}_{1:T}$ and LTM model hyper-parameters $\boldsymbol{\Theta}$ conditional on the variance matrices $\boldsymbol{\Sigma}_{1:T}$ is performed via a direct extension to the multivariate dynamic regression of the ideas of Section 2.3.

4.2 Time-varying variance matrices

Modeling time-varying variance matrices, both residual/error matrices in observation equations of dynamic models and innovations/evolution variance matrices such as $\boldsymbol{\Sigma}_t$ in eqn. (6), is key to many analyses of financial and macroeconomic data. We build on prior Bayesian modeling approaches here for the TV-VAR innovations volatility matrices.

Consider the triangular reduction $\boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}_t' = \boldsymbol{\Lambda}_t^2$ where $\boldsymbol{A}_t$ is the lower triangular matrix of covariance components with unit diagonal elements and $\boldsymbol{\Lambda}_t$ is diagonal with positive elements. That is, $\boldsymbol{\Lambda}_t(\boldsymbol{A}_t')^{-1}$ is the Cholesky component of $\boldsymbol{\Sigma}_t$, viz.

$$\boldsymbol{\Sigma}_t = \boldsymbol{A}_t^{-1}\boldsymbol{\Lambda}_t^2(\boldsymbol{A}_t')^{-1},\ \boldsymbol{A}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{m1,t} & \cdots & a_{m,m-1,t} & 1 \end{pmatrix},\ \boldsymbol{\Lambda}_t = \begin{pmatrix} \sigma_{1t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{mt} \end{pmatrix},$$

and $\boldsymbol{u}_t = \boldsymbol{A}_t^{-1}\boldsymbol{\Lambda}_t\boldsymbol{e}_t$, where $\boldsymbol{e}_t \sim N(\boldsymbol{e}_t|\boldsymbol{0}, \boldsymbol{I})$.

Linked in part to developments in sparse factor modeling (e.g. West 2003; Carvalho et al. 2008), some recent innovations in sparsity modeling for variance matrices have adopted new priors on elements of triangular square-roots of variance matrices such as $\boldsymbol{A}_t$. George et al. (2008) do this with priors allowing zero elements in Cholesky components of constant variance matrices, building on previous, non-sparse approaches (e.g. Pinheiro and Bates 1996; Pourahmadi 1999; Smith and Kohn 2002). The construction has also appeared in time-varying variance matrix modeling in VAR contexts (Cogley and Sargent 2005; Primiceri 2005; Lopes et al. 2010) which is one point of departure for us; in the following section, we use the above Cholesky structure and embed it in a novel LTM framework, combining models for stochastic time-variation with dynamic sparsity to shrink subsets of the lower-triangle of $\boldsymbol{A}_t$ to zero adaptively and dynamically.

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 the time-varying VAR coefficients $\boldsymbol{\beta}_t$ as

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_{\boldsymbol{\beta}} + \boldsymbol{\Phi}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\beta}}) + \boldsymbol{\eta}_{\boldsymbol{\beta}t}, \tag{7}$$

$$\boldsymbol{a}_t = \boldsymbol{\mu}_{\boldsymbol{a}} + \boldsymbol{\Phi}_{\boldsymbol{a}}(\boldsymbol{a}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{a}}) + \boldsymbol{\eta}_{\boldsymbol{a}t}, \tag{8}$$

$$\boldsymbol{h}_t = \boldsymbol{\mu}_{\boldsymbol{h}} + \boldsymbol{\Phi}_{\boldsymbol{h}}(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{h}}) + \boldsymbol{\eta}_{\boldsymbol{h}t}, \tag{9}$$

where $(\boldsymbol{e}_t', \boldsymbol{\eta}_{\boldsymbol{\beta}}', \boldsymbol{\eta}_{\boldsymbol{a}}', \boldsymbol{\eta}_{\boldsymbol{h}}')' \sim N[\boldsymbol{0}, \mathrm{diag}(\boldsymbol{I}, \boldsymbol{V}_{\boldsymbol{\beta}}, \boldsymbol{V}_{\boldsymbol{a}}, \boldsymbol{V}_{\boldsymbol{h}})]$ and with each of the matrices ($\boldsymbol{\Phi}_{\boldsymbol{\beta}}$, $\boldsymbol{\Phi}_{\boldsymbol{a}}$, $\boldsymbol{\Phi}_{\boldsymbol{h}}$, $\boldsymbol{V}_{\boldsymbol{\beta}}$, $\boldsymbol{V}_{\boldsymbol{a}}$, $\boldsymbol{V}_{\boldsymbol{h}}$) diagonal. Thus all univariate time-varying parameters follow stationary AR(1) models, in parallel to the latent VAR model for dynamic regression parameters of Section 2. Note that the specific cases of the log variances h_{it} define traditional univariate stochastic volatility 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 variance matrices is that we can translate the resulting dynamic model for $\boldsymbol{\Sigma}_t$ into a conditional DLM with $\boldsymbol{a}_t$ as the latent state vector. Define $\tilde{\boldsymbol{y}}_t = (\tilde{y}_{1t}, \ldots, \tilde{y}_{mt})' = \boldsymbol{y}_t - \boldsymbol{X}_t \boldsymbol{\beta}_t$ and the $m \times m(m-1)/2$ matrix

$$\tilde{\boldsymbol{X}}_t = \begin{pmatrix} 0 & \cdots & & & & & & 0 \\ -\tilde{y}_{1t} & 0 & 0 & \cdots & & & & \vdots \\ 0 & -\tilde{y}_{1t} & -\tilde{y}_{2t} & 0 & \cdots & & & \\ 0 & 0 & 0 & -\tilde{y}_{1t} & \cdots & & & \\ \vdots & & & & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & & & 0 & -\tilde{y}_{1t} & \cdots & -\tilde{y}_{m-1,t} \end{pmatrix}.$$

From the model identity $\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + \boldsymbol{A}_t^{-1} \boldsymbol{\Lambda}_t \boldsymbol{e}_t$ and using the lower triangular form of $\boldsymbol{A}_t$ we deduce $\tilde{\boldsymbol{y}}_t = \tilde{\boldsymbol{X}}_t \boldsymbol{a}_t + \boldsymbol{\Lambda}_t \boldsymbol{e}_t$ for all t. This couples with the state evolution of eqn. (8) to define a conditional DLM; the MCMC analysis will then extend to include a component to resample the $\boldsymbol{a}_{1:T}$ sequence at each iteration, using the efficient FFBS strategy for conditionally normal DLMs.

4.3 Latent threshold time-varying variance matrices

Our LTM structure for time-varying variance matrices directly adapts the LTM strategy from Section 2 to apply to the state vector $\boldsymbol{a}_t$ in the reformulated model above. That is, introduce a latent VAR process $\boldsymbol{\alpha}_t$ to substitute for $\boldsymbol{a}_t$ in the conditional model of the preceding section. With $\boldsymbol{\alpha}_t$ having elements $\alpha_{ij,t}$ stacked as are the elements of $\boldsymbol{a}_t$, define $\boldsymbol{a}_t = \boldsymbol{\alpha}_t \circ \boldsymbol{s}_{\boldsymbol{a}t}$ with indicator vector $\boldsymbol{s}_{\boldsymbol{a}t}$ of the form discussed in Section 2. That is, for each of the strictly lower triangular elements i, j of $\boldsymbol{A}_t$, we have

$$a_{ij,t} = \alpha_{ij,t} s_{aij,t}, \quad s_{aij,t} = I(|\alpha_{ij,t}| \geq d_{aij}), \quad i = 1, \ldots, m, \ j = 1, \ldots, i-1.$$

The LTM extension of Section 4.2 is then

$$\tilde{\boldsymbol{y}}_t = \tilde{\boldsymbol{X}}_t \boldsymbol{a}_t + \boldsymbol{\Lambda}_t \boldsymbol{e}_t, \tag{10}$$

$$\boldsymbol{a}_t = \boldsymbol{\alpha}_t \circ \boldsymbol{s}_{\boldsymbol{a}t}, \tag{11}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}} (\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \boldsymbol{\eta}_{\boldsymbol{\alpha}t}, \quad \boldsymbol{\eta}_{\boldsymbol{\alpha}t} \sim N(\boldsymbol{\eta}_{\boldsymbol{\alpha}t} | \boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\alpha}}), \tag{12}$$

where $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ and $\boldsymbol{V}_{\boldsymbol{\alpha}}$ are diagonal matrices, eqn. (8) is now deleted and replaced by $\boldsymbol{a}_t = \boldsymbol{\alpha}_t \circ \boldsymbol{s}_{\boldsymbol{a}t}$, while all other elements of the model in eqns. (7) and (9) are unchanged. The MCMC estimation procedure developed in Section 2, extended as described above, can now be straightforwardly applied to this time-varying, sparsity-inducing LTM extension of the AR(1) Cholesky based volatility model. Computational details for the LT-VAR model with the LT time-varying variance matrix are explained in Appendix A, available as on-line Supplementary Material.

One point of interest is that a sparse $\boldsymbol{A}_t$ matrix can translate into a sparse precision matrix $\boldsymbol{\Omega}_t = \boldsymbol{A}_t' \boldsymbol{\Lambda}_t^{-2} \boldsymbol{A}_t$; the more zeros there are in the lower triangle of $\boldsymbol{A}_t$, the more zeros there can be in the precision matrix. Hence the LTM defines an approach to time-varying sparsity modeling for precision matrices, and hence *dynamic graphical models* as a result. 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 (Jones et al. 2005). The LTM approach 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.

From a full MCMC analysis we can obtain posterior inferences on sparsity structure. For each pair of elements i, j and each time t, the posterior simulation outputs provide realizations of the indicators $s_{aij,t} = 0$ so that we have direct Monte Carlo estimates of the posterior probability of $s_{aij,t} = 0$. This translates also to the precision matrix elements and the implied graphical model at each time t, providing inferences on edge inclusion at each time t.

5 Application: A study of US macroeconomic time series

5.1 Introduction 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. Here we use the LTM strategy for TV-VAR models and volatility matrices with stochastic volatility as described above in analysis of a topical time series of US data. A parallel study of Japanese macroeconomic data with similar goals, but some additional features related to Japanese monetary policy, is detailed in the Appendix D, available as on-line Supplementary Material. In terms of broad summaries, we note that in each of the two applications we find that: *(i)* there is strongly significant evidence for dynamic thresholding when compared to the models with no thresholding; *(ii)* the LTM yields intuitive and interpretable results, particularly with respect to inferred impulse response functions; and *(iii)* again relative to the standard, non-thresholded models, the LTM can yield practically significant improvements in multi-step, out-of-sample predictions, these being particularly relevant to policy uses of such models; these predictive improvements are robust and sustained over time.

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 (2011) 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. 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. Comparisons are of interest with, in particular, Markov switching structures as in some of the above references, and popular in econometric studies of regime changes in particular. Though such models share similar technical aspects with the LTM approach, they are inherently focused on very 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. We focus here 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.

5.2 Data, models and analysis goals

We analyze 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. The variables are ordered this way in $\boldsymbol{y}_t$. Prior studies and especially that of Primiceri (2005) focused on the data over the period 1963/Q1–2001/Q3, which we extend here to the full data set up to more recent times, 1963/Q1–2011/Q4; see Figure 4.

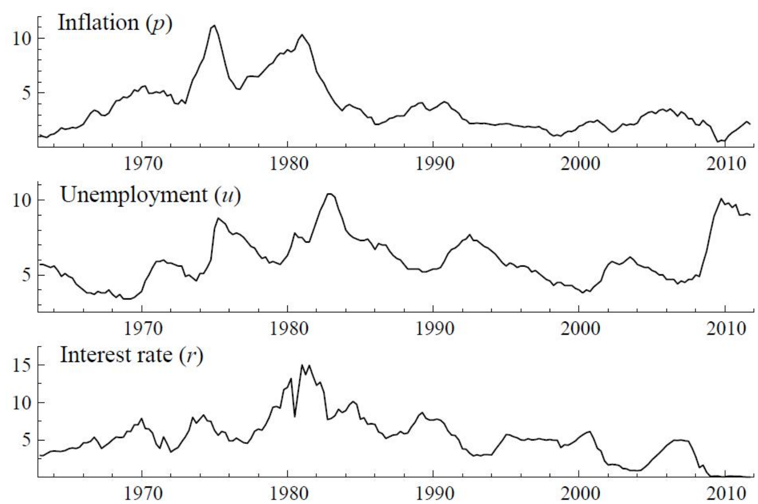


Figure 4: US macroeconomic time series (indices $\times 100$ for % basis).

We compare analyses based on the LT-VAR model with time-varying variance matrices of Section 4.3, and the standard, non-threshold (NT-VAR) model. The latter is almost equivalent to a commonly used TV-VAR of Primiceri (2005); the difference is that Primiceri (2005) uses random walks for the time-varying parameters rather than the stationary AR processes we adopt.

We use VAR models of order $p = 3$. Analysis methodology could be extended to include formal Bayesian inference on model order; the LTMs have the opportunity to advance the methodology for order selection since they can naturally "cut-back" to models of lower order if the data so suggest. This is, however, beyond our goals and scope of the current paper. In any case, we stress a key practical issue related to this and that partly underlies the method we used to specify $p = 3$. Based on the full MCMC analysis of the series over the period 1963/Q1–2000/Q3, we generated out-of-sample forecasts 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 our extensive forecasting study of the next section, 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. 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 macro-economic studies such as the example here.

For all model analyses reported we take the following prior components. With $v_{\boldsymbol{\beta}i}^2$, $v_{\boldsymbol{\alpha}i}^2$ and $v_{\boldsymbol{h}i}^2$ denoting the i^{th} diagonal elements of $\boldsymbol{V_\beta}$, $\boldsymbol{V_\alpha}$ and $\boldsymbol{V_h}$, respectively, we use $v_{\boldsymbol{\beta}i}^{-2} \sim G(20, 0.01)$, $v_{\boldsymbol{\alpha}i}^{-2} \sim G(2, 0.01)$ and $v_{\boldsymbol{h}i}^{-2} \sim G(2, 0.01)$. For $\boldsymbol{\mu}_h$ we assume $\exp(-\mu_{hi}) \sim G(3, 0.03)$. The other prior specifications for the LTM models and the simulation size are as in Section 3.

5.3 Out-of-sample forecasting performance and comparisons

5.3.1 Point forecasting accuracy over multiple horizons

We summarize out-of-sample forecast performance to compare the LT- and NT-VAR models in predicting 1,2,3 and 4 quarters ahead. We do this in the traditional recursive forecasting format that mirrors the reality facing forecasters: given data from the start of 1963 up to any quarter t, we refit the full MCMC analysis of each model to define the posterior based on $\boldsymbol{y}_{1:t}$. We then simulate the resulting out-of-sample predictive distributions over the following 4 quarters, times $t+1 : t+4$, to generate realized out-of-sample forecast errors. We then move ahead one quarter to observe the next observation y_{t+1}, rerun the MCMC based on the updated data series $\boldsymbol{y}_{1:t+1}$, and then forecast 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 we find.

Results are summarized in terms of root mean squared forecast errors (RMSFE) for each variable and each horizon, averaged over the 16 years of quarterly forecasting; see Table 2. With the exception of the $1-$step forecasts of inflation (p), 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. As might be expected, 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 (p), which we found is due to relatively less accurate point forecasts for p under the LTM in just two quarters: 2008/Q4 and 2010/Q2. If we remove just these two times 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. We plan to explore such extensions in future.

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, as we have noted above, further investigation shows that they are robust and sustained over time. Some further summaries in Appendix C (Supplementary Material) show posteriors of RMSFEs that more formally substantiate the 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 overall and at longer horizons in the less predictable years of 2007-2011.

	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

Table 2: Forecasting performance for US macroeconomic data over the 16 years 1996/Q4–2011/Q4: RMSFE for each horizon and each variable, comparing LT-VAR and NT-VAR models.

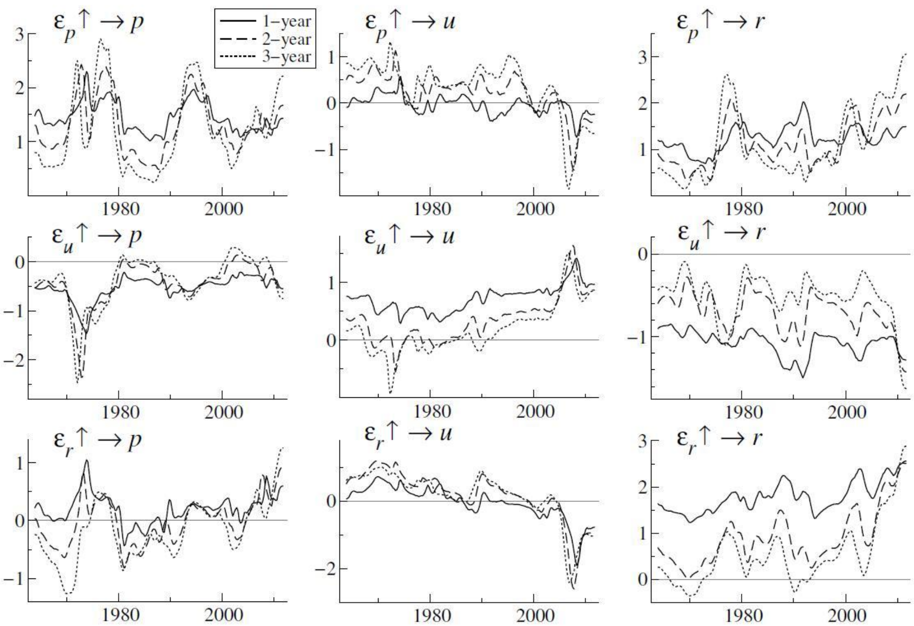
We note that we have also explored analyses in which the LTM structure is applied to $\boldsymbol{b}_t$ but not $\boldsymbol{a}_t$, and vice-versa. Results from those analyses (not detailed here) confirm that: (i) allowing dynamic sparsity in either $\boldsymbol{b}_t$ or $\boldsymbol{a}_t$ improves out of sample predictions relative to the standard NT-VAR model; (ii) the LTM for $\boldsymbol{b}_t$ alone leads to substantial time-varying shrinkage of the coefficients and contributes much of the predictive improvement; while (iii) the full LT-VAR model with sparsity in both $\boldsymbol{b}_t$ and $\boldsymbol{a}_t$ substantially improves predictions relative to those models with sparsity in only one or the other components.

5.3.2 Impulse response analysis

We discuss further aspects of out-of-sample prediction of key practical interest, namely impulse response analysis. From the recursive forecasting analysis of the above section, at each quarter we can also compute impulse response functions projecting ahead multiple quarters. Here we 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. We show some summary results and implications from both the the LT- and NT-VAR models, for further insights and comparisons of predictive implications of the latent thresholding structure.

Figure 5 displays posterior means of the impulse response for $1-, 2-$ and $3-$year ahead hori-

(i) NT-VAR model



(ii) LT-VAR model

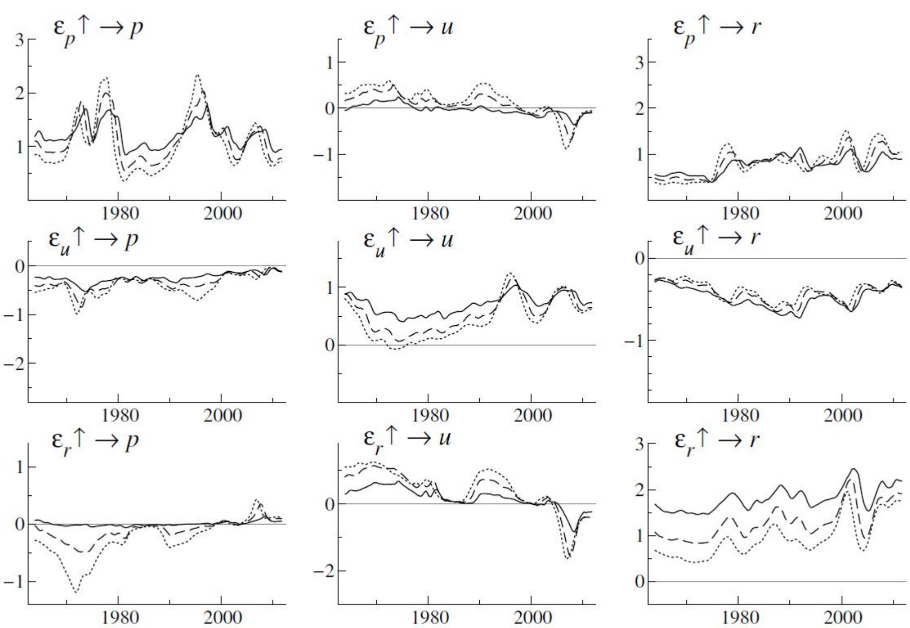


Figure 5: Impulse response trajectories for $1-, 2-$ and $3-$year ahead horizons from NT-VAR (upper) and LT-VAR models (lower) for US macroeconomic data. The symbols $\varepsilon_u\uparrow \to w$ refer to the response of the variable w to a shock to the innovation of variable u. The shock size is set at the average of the stochastic volatility across time for each series. Compare with results in Primiceri (2005).

zons from both models. Up to the period to the end of 2001 as earlier analyzed by previous authors, we find similar results under our NT-VAR model, as expected. Over those years, 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. Our focus here is in comparing the NT-VAR and LT-VAR models. First, the trajectories of the response from the LT-VAR model are smoother than under 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. We note that these features are replicated in the analysis of the Japanese data from the parallel econometric context, discussed in detail in Appendix D in the on-line Supplementary Material.

5.4 Additional predictive assessment and model comparisons

Additional formal model assessments are provided by out-of-sample predictive densities. In comparing LT-VAR with NT-VAR, the *log predictive density ratio* for forecasting h quarters ahead from quarter t is $\mathrm{LPDR}_t(h) = \log\{p_{\text{LT-VAR}}(\boldsymbol{y}_{t+h}|\boldsymbol{y}_{1:t})/p_{\text{NT-VAR}}(\boldsymbol{y}_{t+h}|\boldsymbol{y}_{1:t})\}$ where $p_{\mathrm{M}}(\boldsymbol{y}_{t+h}|\boldsymbol{y}_{1:t})$ 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 $\mathrm{LPDR}_t(1)$ 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 $\mathrm{LPDR}_t(h)$ for $h > 1$ provide further insights into relative forecasting ability at longer horizons.

Figure 6 shows that LT-VAR dominates NT-VAR at all time points. 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. In connection with our earlier discussion of differing time periods, it is notable that the the density ratios are lower in the later recessionary years 2007–2011; nevertheless, the dominance of the LTM is maintained uniformly. With $h = 1$, cumulating values over the full time period gives the log Bayes' factors (log model likelihood ratios) comparing LT-VAR to NT-VAR of about 415.17; this implies very substantial formal evidence in favor of the LT-VAR. Figure 6 also reflects the increased ability of LT-VAR to improve forecasts with increasing horizon, as the LPDR values are increasing with h, essentially uniformly. A part of this is that the LT-VAR model generally produces more precise predictive distributions as well as more accurate point forecasts; the increased spread of the predictive distributions under the less parsimonious NT-TVAR model at longer horizons is detrimental, relative to the LTM.

5.5 Summaries of posterior inferences from analysis of full series

We now discuss and highlight some aspects of posterior inference under the LT-VAR model based on a final MCMC analysis to fit the model to the full series, 1963/Q1–2011/Q4.

Figures 7–9 show some selected summaries that highlight the adaptive, dynamic sparsity in TV-VAR and volatility matrix parameters. Time-varying sparsity is observed for several of the $b_{ij,\ell,t}$

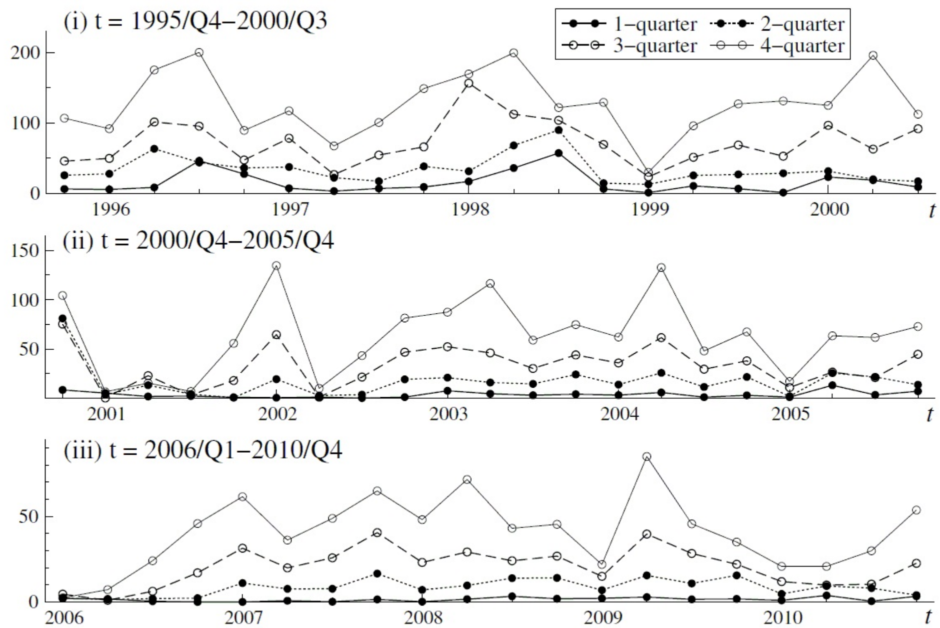


Figure 6: Log predictive density ratios LPDR$_t(h)$ for LT-VAR over NT-VAR model in analysis of the US time series: the per quarter LPDR values are plotted at the times t forecasts are made and for each of the 4 forecast horizons h.

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 8 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 $\boldsymbol{b}_t$.

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 9 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 Sargent 2005; Primiceri 2005); 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.

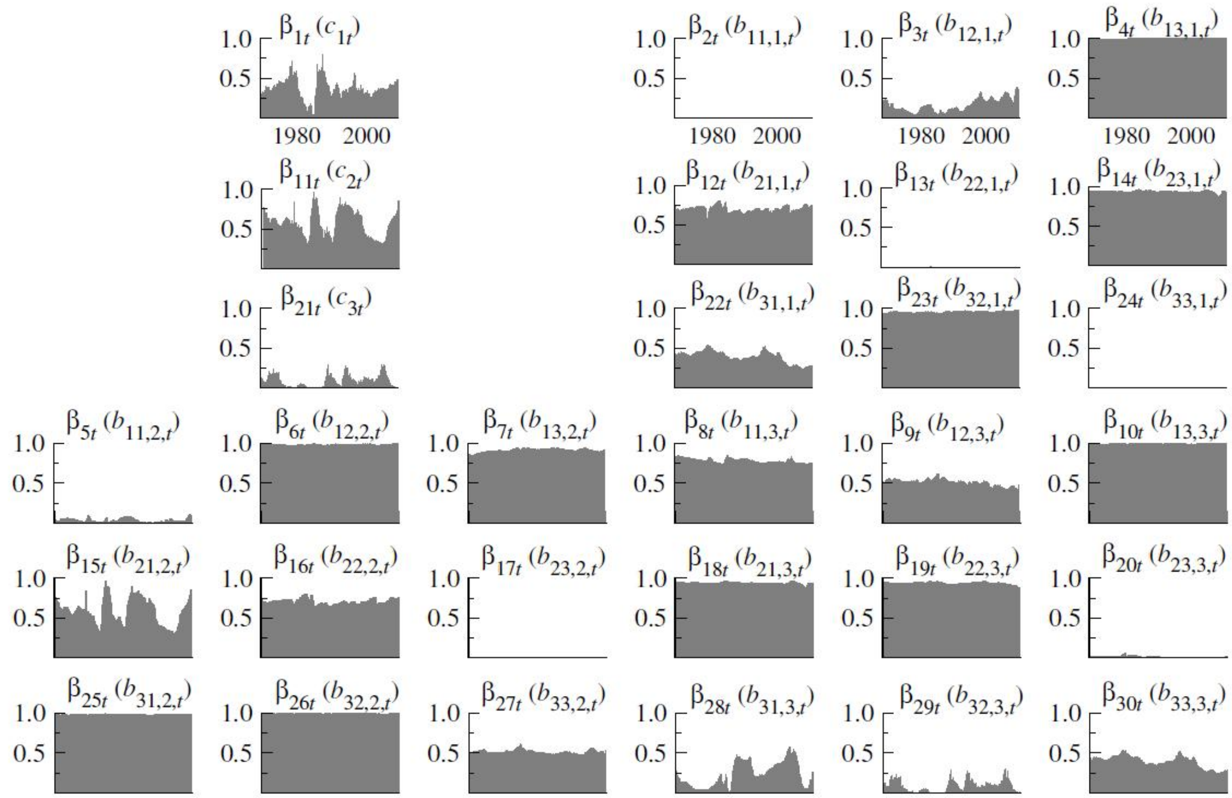


Figure 7: Posterior probabilities of $s_{it} = 0$ for US macroeconomic data. The corresponding indices of $\boldsymbol{c}_t$ or $\boldsymbol{B}_{\ell t}$ are in parentheses.

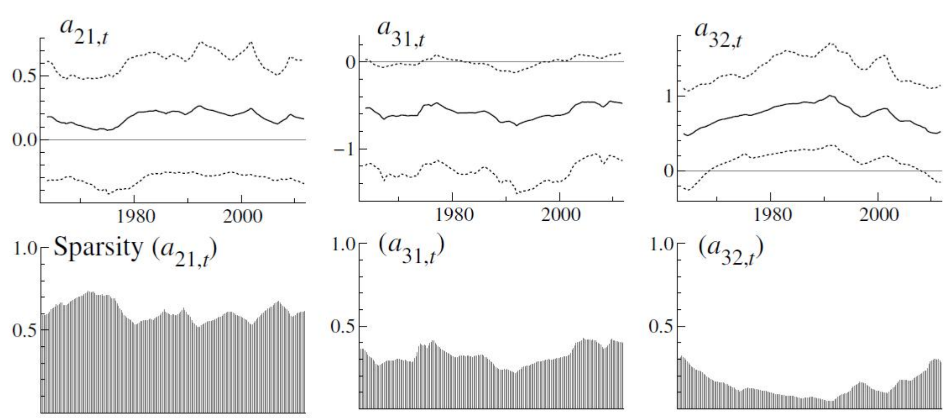


Figure 8: Posterior trajectories of $a_{ij,t}$ for US macroeconomic data: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

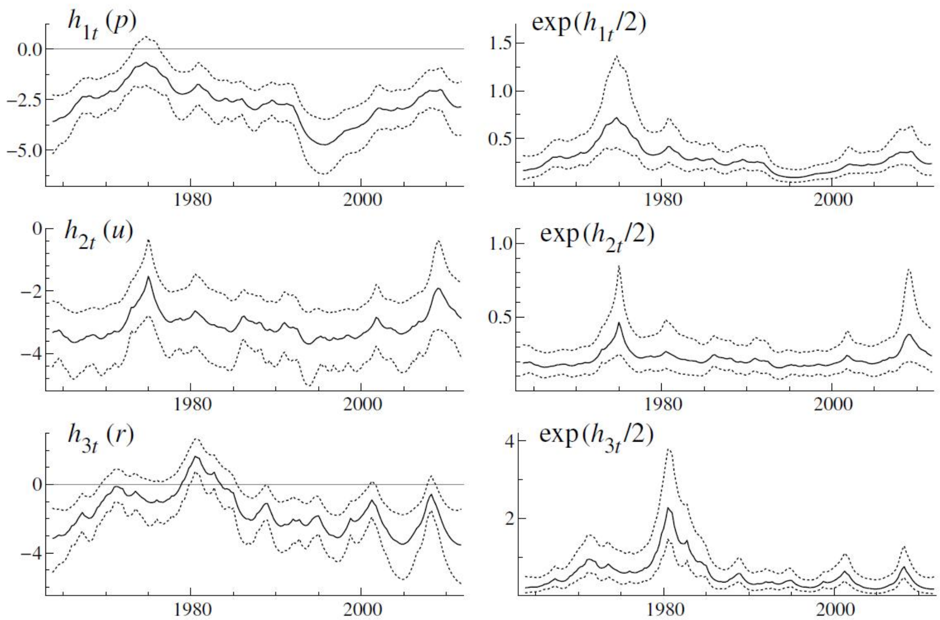


Figure 9: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$ for US macroeconomic data: posterior means (solid) and 95% credible intervals (dotted).

6 Summary comments

By introducing latent threshold process modeling as a general strategy, we have defined an approach to time-varying parameter time series that overlays several existing model classes. The LTM approach provides a general framework for time-varying sparsity modeling, in which time-varying parameters can be shrunk to zero for some periods of time while varying stochastically at non-zero values otherwise. In this sense, the approach defines automatic parsimony in dynamic models, with the ability to dynamically select in/out potential predictor variables in dynamic regression, lagged predictors in dynamic VAR models, edges in dynamic graphical models induced in novel multivariate volatility frameworks, and by extension other contexts not explicitly discussed here including, for example, dynamic simultaneous equations models and dynamic factor models. Global relevance, or irrelevance, of some variables and parameters is a special case, so the model also allows for global model comparison and variable selection.

The US macroeconomic studies, and a companion Japanese data study (Appendix D, Supplementary Material), illustrate the practical use and impact of the LTM structure. Data induced dynamic sparsity feeds through to improvements in forecasting performance and contextually reasonable shrinkage of inferred impulse response functions. Our extensive model evaluations 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 overlaying time-varying VAR and volatility models. Importantly for applications, the LTMs are empirically seen to be increasingly preferred by measures of forecasting accuracy at longer horizons, which also supports the view that impulse response analyses, as demonstrated in our two studies, will be more robust and reliable under latent thresholding.

The Japanese study in a very different economy 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. In a different context, the roles of LTMs in short-term forecasting and portfolio analyses for financial time series data are critical; application to detailed portfolio decision problems represents a key current direction (Nakajima and West 2012).

In addition to such broader applications and extension of the LTM concepts to other models, including to topical contexts such as Bayesian dynamic factor models in economic and financial applications (e.g. Aguilar and West 2000; Carvalho et al. 2011; Wang and West 2009), there are a number of methodological and computational areas for further investigation. Among these, we note the potential of more elaborate state process models, asymmetric and/or time-varying thresholds, as well as refined MCMC methods including potential reversible jump approaches.

Acknowledgements

The authors thank the editor and the associate editor for detailed comments and suggestions that led to substantial improvements in the paper, and to Francesco Bianchi and Herman van Dijk for helpful comments and suggestions. The research reported here was partly supported by a grant from the National Science Foundation [DMS-1106516 to M.W.]. Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

References

Aguilar, O. and West, M. (2000), "Bayesian dynamic factor models and portfolio allocation," *Journal of Business and Economic Statistics*, 18, 338–357.

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.

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., 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., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., 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*, in press.

Chen, C. W. S., Liu, F., and Gerlach, R. (2010), "Bayesian subset selection for threshold autoregressive moving-average models," *Computational Statistics*, forthcoming.

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. (2001), "Evolving post World War II U.S. inflation dynamics," *NBER Macroeconomics Annual*, 16, 331–373.

— (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.

Doornik, J. (2006), *Ox: Object Oriented Matrix Programming*, London: Timberlake Consultants Press.

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 - Vancouver, Canada, December*.

Frühwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, 15, 183–202.

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.

Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., New York: Oxford University Press, vol. 4, pp. 169–188.

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.

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. (2011), "VAR forecasting using Bayesian variable selection," *Journal of Applied Econometrics*.

Lopes, H. F., McCulloch, R. E., and Tsay, R. (2010), "Cholesky stochastic volatility," Tech. rep., University of Chicago, Booth Business School.

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. (2012), "Dynamic factor volatility modeling: A Bayesian latent threshold approach," *Journal of Financial Econometrics*, -, –, -.

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.

Pinheiro, J. C. and Bates, D. M. (1996), "Unconstrained parametrizations for variance-covariance matrices," *Statistics and Computing*, 6, 289–296.

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. and West, M. (2010), *Time Series Modeling, Computation, and Inference*, New York: Chapman & Hall/CRC.

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

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.

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.

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.

— (2012), "Bayesian dynamic modelling," in *Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian F. M. Smith*, eds. Damien, P., Dellaportes, P., Polson, N. G., and Stephens, D. A., Clarendon: Oxford University Press, chap. 8, pp. 145–166.

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.

Xia, Y. and Tong, H. (2011), "Feature matching in time series modeling," *Statistical Science*, 26, 21–46.

Yoshida, R. and West, M. (2010), "Bayesian learning in sparse graphical factor models via annealed entropy," *Journal of Machine Learning Research*, 11, 1771–1798.

Bayesian Analysis of Latent Threshold Dynamic Models

Jouchi Nakajima & Mike West

Supplementary Material: Appendices

A Posterior computation and MCMC algorithm

A.1 LT regression model

In the LT regression model defined by eqns. (1)-(3), we describe a MCMC algorithm for simulation of the full joint posterior $p(\mathbf{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \boldsymbol{d}|\boldsymbol{y}_{1:T})$. We assume prior forms of the following: $\mu_i \sim N(\mu_{i0}, w_{i0}^2)$, $(\phi_i + 1)/2 \sim \pi(\phi_i)$, $\sigma_{i\eta}^{-2} \sim G(v_{0i}/2, V_{0i}/2)$, $\sigma^{-2} \sim G(n_0/2, S_0/2)$, $\beta_{i1}|\Theta \sim N(\mu_i, v_i^2)$, and $d_i \sim U(0, |\mu_i| + K_i v_i)$.

A.1.1 Sampling $\mathbf{\Theta}$ and σ

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

$$p(\mu_i|\phi_i, \sigma_{i\eta}, \beta_{i,1:T}, d_i) \propto TN_{D_i}(\mu_i|\hat{\mu}_i, \hat{w}_i^2)(|\mu_i| + Kv_i)^{-1},$$

where TN_{D_i} denotes the density of a truncated normal for μ_i on $D_i = \{\mu_i : d_i < |\mu_i| + K_i v_i\}$, and

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

A Metropolis-Hastings step draws a candidate μ_i^* from this truncated normal, accepting the draw with probability

$$\min\left\{1,\ \frac{|\mu_i| + Kv_i}{|\mu_i^*| + Kv_i}\right\}.$$

Second, the conditional posterior density of ϕ_i is

$$p(\phi_i|\mu_i, \sigma_{i\eta}, \beta_{i,1:T}, d_i) \propto \pi(\phi_i)(1-\phi_i^2)^{1/2} TN_{(-1,1)\times E_i}(\hat{\phi}_i, \sigma_{\phi_i}^2)\{|\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^2)^{1/2}\}^{-1},$$

where $\hat{\phi}_i = \sum_{t=1}^{T-1} \bar{\beta}_{i,t+1}\bar{\beta}_{it} / \sum_{t=2}^{T-1} \bar{\beta}_{it}^2$, $\sigma_{\phi_i}^2 = \sigma_{i\eta}^2 / \sum_{t=2}^{T-1} \bar{\beta}_{it}^2$ with $\bar{\beta}_{it} = \beta_{it} - \mu_i$, and E_i is the truncation region $E_i = \{\phi_i : d_i < |\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^2)^{1/2}\}$. A Metropolis-Hastings step draws a candidate ϕ_i^* from this truncated normal, accepting the draw with probability

$$\min\left\{1,\ \frac{\pi(\phi_i^*)(1-\phi_i^{*2})^{1/2}\{|\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^2)^{1/2}\}}{\pi(\phi_i)(1-\phi_i^2)^{1/2}\{|\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^{*2})^{1/2}\}}\right\}.$$

Third, the conditional posterior density of $\sigma_{i\eta}^{-2}$ is

$$p(\sigma_{i\eta}^{-2}|\mu_i, \phi_i, \beta_{i,1:T}, d_i) \propto TG_{F_i}(\sigma_{i\eta}^{-2}|\hat{v}_i/2, \hat{V}_i/2)\{|\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^2)^{1/2}\}^{-1},$$

where TG_{F_i} is the density of the implied gamma distribution truncated to $F_i = \{\sigma_{i\eta}^{-2} : d_i < |\mu_i| + K_i\sigma_{i\eta}/(1-\phi_i^2)^{1/2}\}$, and

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

A Metropolis-Hastings step draws a candidate $1/\sigma_{i\eta}^{*2}$ from this truncated gamma, accepting the draw with probability

$$\min\left\{1, \frac{|\mu_i| + K\sigma_{i\eta}/(1-\phi_i^2)^{1/2}}{|\mu_i| + K\sigma_{i\eta}^*/(1-\phi_i^2)^{1/2}}\right\}.$$

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

A.1.2 Sampling $\boldsymbol{\beta}_{1:T}$

Conditional on $(\boldsymbol{\Theta}, \sigma, \boldsymbol{d}, \boldsymbol{y}_{1:T})$, we sample the conditional posterior at time t, $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{-t})$, sequentially for $t = 1 : T$ using a Metropolis-Hastings sampler. The MH proposals come from a non-thresholded version of the model specific to each time t, as follows. Fixing $\boldsymbol{s}_t = \boldsymbol{1}$, take proposal distribution $N(\boldsymbol{\beta}_t|\boldsymbol{m}_t, \boldsymbol{M}_t)$ where

$$\begin{aligned}
\boldsymbol{M}_t^{-1} &= \sigma^{-2}\boldsymbol{x}_t\boldsymbol{x}_t' + \boldsymbol{\Sigma}_\eta^{-1}(\boldsymbol{I} + \boldsymbol{\Phi}'\boldsymbol{\Phi}), \\
\boldsymbol{m}_t &= \boldsymbol{M}_t\left[\sigma^{-2}\boldsymbol{x}_t y_t + \boldsymbol{\Sigma}_\eta^{-1}\left\{\boldsymbol{\Phi}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}) + (\boldsymbol{I} - 2\boldsymbol{\Phi} + \boldsymbol{\Phi}'\boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right],
\end{aligned}$$

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

$$\begin{aligned}
\boldsymbol{M}_1^{-1} &= \sigma^{-2}\boldsymbol{x}_1\boldsymbol{x}_1' + \boldsymbol{\Sigma}_{\eta 0}^{-1} + \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi}, \\
\boldsymbol{m}_1 &= \boldsymbol{M}_1\left[\sigma^{-2}\boldsymbol{x}_1 y_1 + \boldsymbol{\Sigma}_{\eta 0}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\left\{\boldsymbol{\beta}_2 - (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right], \\
\boldsymbol{M}_T^{-1} &= \sigma^{-2}\boldsymbol{x}_T\boldsymbol{x}_T' + \boldsymbol{\Sigma}_\eta^{-1}, \\
\boldsymbol{m}_T &= \boldsymbol{M}_T\left[\sigma^{-2}\boldsymbol{x}_T y_T + \boldsymbol{\Sigma}_\eta^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{\beta}_{T-1} + (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right],
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\eta 0} = \mathrm{diag}(v_1^2, \ldots, v_k^2)$. The candidate is accepted with probability

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

where $\boldsymbol{b}_t = \boldsymbol{\beta}_t \circ \boldsymbol{s}_t$ is the current LTM state at t and $\boldsymbol{b}_t^* = \boldsymbol{\beta}_t^* \circ \boldsymbol{s}_t^*$ the candidate.

A.1.3 Sampling $\boldsymbol{d}$

We adopt a direct MH algorithm to sample the conditional posterior distribution of d_i, conditional on $(\boldsymbol{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \boldsymbol{d}_{-i}, \boldsymbol{y}_{1:T})$ where $\boldsymbol{d}_{-i} = d_{1:k}\backslash d_i$. A candidate is drawn from the current conditional prior, $d_i^* \sim U(0, |\mu_i| + K_i v_i)$, and accepted with probability

$$\alpha(d_i, d_i^*) = \min\left\{1, \prod_{t=1}^{T}\frac{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t^*, \sigma^2)}{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t, \sigma^2)}\right\},$$

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

A.2 LT-VAR model

We detail sampling steps for posterior computations in the LT-VAR model where both the VAR coefficients and covariance components of Cholesky-decomposed variance matrices follow LT-AR(1) processes; see eqns. (6)-(7), and (9)-(12). Let $\boldsymbol{\Theta}_{\boldsymbol{\gamma}} = (\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Phi}_{\boldsymbol{\gamma}}, \boldsymbol{V}_{\boldsymbol{\gamma}})$ where $\gamma \in \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{h}\}$. Standard MCMC algorithms for TV-VAR models are well documented; see, for example, Primiceri (2005), Koop and Korobilis (2010), and Nakajima (2011). These form a basis for the new MCMC sampler in our latent thresholded model extensions.

1. Sampling $\boldsymbol{\beta}_{1:T}$
   Conditional on $(\boldsymbol{\Theta}_{\boldsymbol{\beta}}, \boldsymbol{d}, \boldsymbol{\alpha}_{1:T}, \boldsymbol{h}_{1:T}, \boldsymbol{y}_{1:T})$, $\boldsymbol{\beta}_t$ is generated using the MH sampler implemented in Section A.1.2. Note that the response here is multivariate; the ingredients in the proposal distribution are generalized to
$$\begin{aligned} \boldsymbol{M}_t^{-1} &= \boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t + \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{I} + \boldsymbol{\Phi}'\boldsymbol{\Phi}), \\ \boldsymbol{m}_t &= \boldsymbol{M}_t\left[\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{y}_t + \boldsymbol{V}_{\boldsymbol{\beta}}^{-1}\left\{\boldsymbol{\Phi}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}) + (\boldsymbol{I} - 2\boldsymbol{\Phi} + \boldsymbol{\Phi}'\boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right], \end{aligned}$$
   and the MH acceptance probability is
$$\alpha(\boldsymbol{\beta}_t, \boldsymbol{\beta}_t^*) = \min\left\{1, \frac{N(\boldsymbol{y}_t|\boldsymbol{X}_t\boldsymbol{b}_t^*, \boldsymbol{\Sigma}_t)N(\boldsymbol{\beta}_t|\boldsymbol{m}_t, \boldsymbol{M}_t)}{N(\boldsymbol{y}_t|\boldsymbol{X}_t\boldsymbol{b}_t, \boldsymbol{\Sigma}_t)N(\boldsymbol{\beta}_t^*|\boldsymbol{m}_t, \boldsymbol{M}_t)}\right\}.$$

2. Sampling $\boldsymbol{\alpha}_{1:T}$
   Conditional on $(\boldsymbol{\Theta}_{\boldsymbol{\alpha}}, \boldsymbol{d}_a, \boldsymbol{\beta}_{1:T}, \boldsymbol{h}_{1:T}, \boldsymbol{y}_{1:T})$ where $\boldsymbol{d}_a = \{d_{aij}\}$, sampling $\boldsymbol{\alpha}_{1:T}$ requires the same MH sampling strategy as $\boldsymbol{\beta}_{1:T}$ based on the model (10)-(12).

3. Sampling $\boldsymbol{h}_{1:T}$
   Conditional on $(\boldsymbol{\Theta}_{\boldsymbol{h}}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \boldsymbol{y}_{1:T})$, defining $\boldsymbol{y}_t^* = \boldsymbol{A}_t(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}_t)$ and $\boldsymbol{y}_t^* = (y_{1t}^*, \ldots, y_{mt}^*)'$ yields a form of univariate stochastic volatility:
$$\begin{aligned} y_{it}^* &= \exp(h_{it}/2)e_{it}, \\ h_{it} &= \mu_{hi} + \phi_{hi}(h_{i,t-1} - \mu_{hi}) + \eta_{hit}, \\ (e_{it}, \eta_{hit})' &\sim N\left(\boldsymbol{0}, \mathrm{diag}(1, v_{hi}^2)\right), \end{aligned}$$
   where μ_{hi}, ϕ_{hi} and v_{hi}^2 are the i-th (diagonal) element of $\boldsymbol{\mu}_{\boldsymbol{h}}, \boldsymbol{\Phi}_{\boldsymbol{h}}$ and $\boldsymbol{V}_{\boldsymbol{h}}$, 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.

4. Sampling $(\boldsymbol{\Theta}_{\boldsymbol{\beta}}, \boldsymbol{\Theta}_{\boldsymbol{\alpha}}, \boldsymbol{\Theta}_{\boldsymbol{h}})$

   Conditional on $(\boldsymbol{\beta}_{1:T}, \boldsymbol{d})$ and $(\boldsymbol{\alpha}_{1:T}, \boldsymbol{d}_a)$, sampling $\boldsymbol{\Theta}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\alpha}}$, respectively, is implemented as in Section A.1.1. Conditional on $\boldsymbol{h}_{1:T}$, sampling $\boldsymbol{\Theta}_{\boldsymbol{h}}$ also follows the same sampling strategy, although it does not require the rejection step associated with the thresholds.

5. Sampling $(\boldsymbol{d}, \boldsymbol{d}_a)$

   Conditional on all other parameters, we generate the latent thresholds $\boldsymbol{d}$ and $\boldsymbol{d}_a$ using the sampler described in Section A.1.3.

B Empirical evaluation of MCMC sampling

This appendix reports performance of the MCMC sampler for the LTM in the simulation example. Figure 10 plots autocorrelations and sample paths of MCMC draw for selected parameters of the simulation example (Section 3). In spite of non-linearity of the model structure, the autocorrelations decay quickly and sample paths appear to be stable, indicating the chain mixes well. In addition, MH acceptance rates are empirically high: about 80% for the generation of $\boldsymbol{\beta}_t$ and $\boldsymbol{\alpha}_t$, about 40% for $\boldsymbol{d}$ and $\boldsymbol{d}_a$, and about 95% for $(\boldsymbol{\Theta}_{\boldsymbol{\beta}}, \boldsymbol{\Theta}_{\boldsymbol{\alpha}})$ in the application to macroeconomic data.

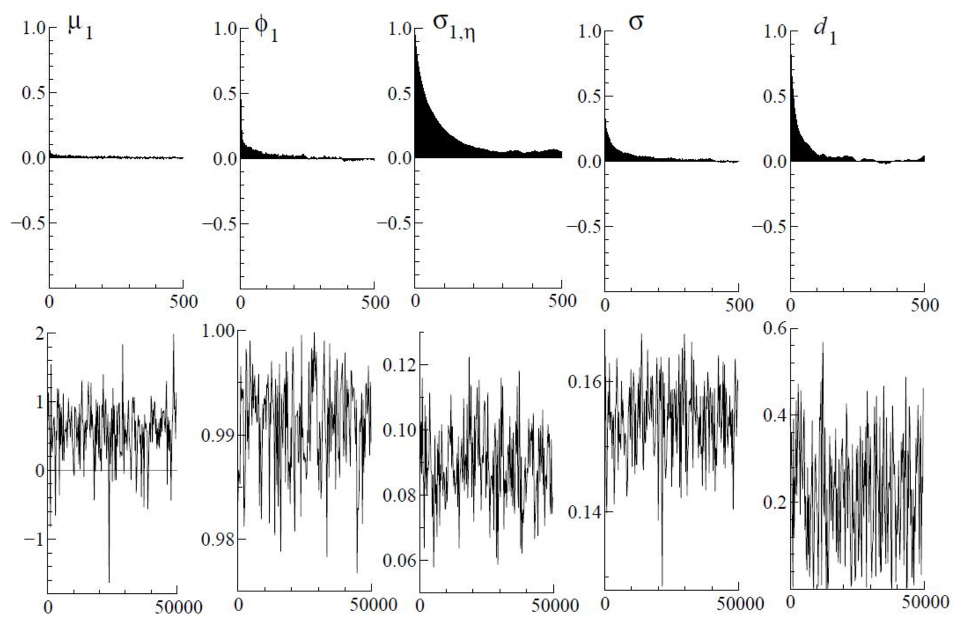


Figure 10: Performance of the MCMC: Autocorrelations (top) and sample paths (bottom) of MCMC draws for selected parameters in simulation example.

To check convergence of MCMC draws, the convergence diagnostic (CD) and relative numerical efficiency measure (a.k.a., effective sample size) of Geweke (1992) are computed. Table 3 reports the CDs (p-values for null hypothesis that the Markov chain converges) as well as inefficiency factors (IFs) for the selected parameters. The CDs indicate the convergence of the MCMC run and the effective sample size is fairy small relative to standard non-linear dynamic models.

	CD	IF
μ_1	0.326	5.0
ϕ_1	0.582	22.1
$\sigma_{1,\eta}$	0.378	107.2
σ	0.150	26.6
d_1	0.503	52.1

Table 3: MCMC diagnostics: Convergence diagnostic (CD) of Geweke (1992) (p-value) and inefficiency factor (IF) for selected parameters in simulation example.

C Additional assessment summaries for US macroeconomic study

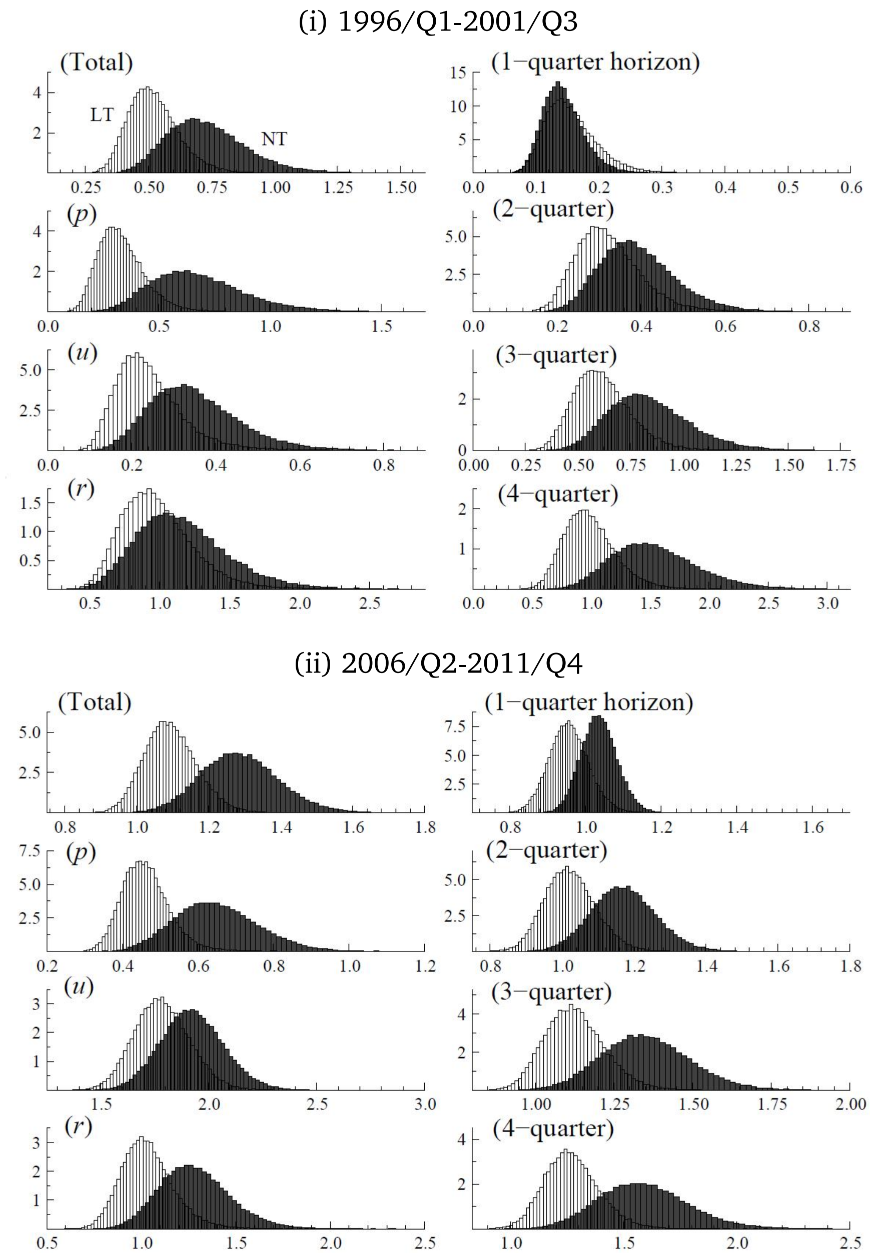


Figure 11: Posteriors of RMSFE from MCMC analysis of US macroeconomic data: (i) 1996/Q1-2001/Q3 and (ii) 2006/Q2-2011/Q4, using NT-VAR (black) and LT-VAR (light) models. Plots are by variable averaged across forecast horizons (left) and by horizon averaged across variables (right). Note the uniform improvements under the LT structure, increased improvements as forecast horizon increases, and increased ability of the LT-VAR models to maintain improved predictive ability in the more volatile second period (ii). Details of RMSFE by variable and horizon, across the full period of recursive out-of-sample forecasting from 2001–2011, are in Table 2.

D Application to Japanese macroeconomic data

D.1 Data

We analyze 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 12. 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 nominal 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. We take the same priors as previous analyses.

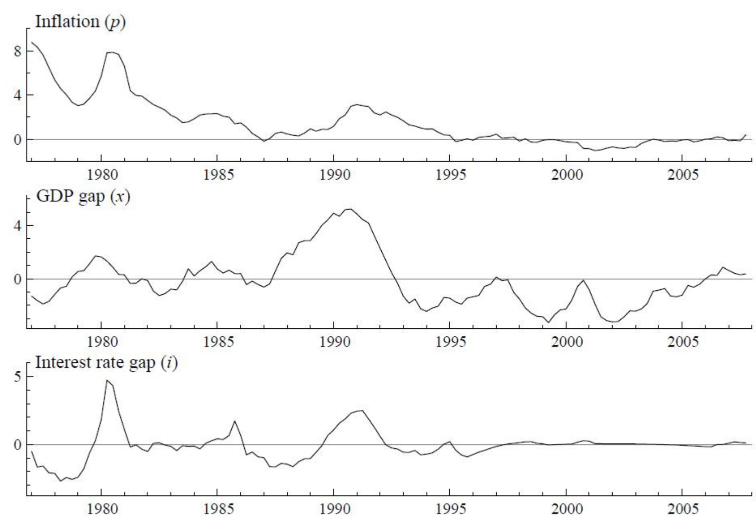


Figure 12: Japanese macroeconomic time series (indices $\times$100 for % basis).

D.2 Forecasting performance and comparisons

We fit and compare predictions from the NT-VAR and LT-VAR models, as in the study of the US time series. Based on evaluation of RMSFE across the out-of-sample forecasts for the final 4 quarters, it is clear that the LT-VAR models perform best when $p = 2$ is assumed, while the non-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.

Some summaries of out-of-sample predictive accuracy appear in Table 4. We computed RMSFE

		Horizon (quarters)			
		1	2	3	4
RMSFE	NT-VAR	0.253	0.387	0.525	0.633
RMSFE	LT-VAR	0.225	0.263	0.311	0.321
Ratio	LT-VAR/NT-VAR	0.889	0.680	0.592	0.507

Table 4: Forecasting performance for Japanese macroeconomic data: RMSFE for 1- to 4- quarter ahead predictions from NT-VAR and LT-VAR models, averaged over the 3 variables at each horizon.

for ten different selections of subsets of data: beginning with the sample period from 1977/Q1–2004/Q3, we fit the model and then forecast 1- to 4-quarters ahead over 2004/Q4–2005/Q3, and then repeat 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.

D.3 Some summaries of posterior inferences

Figure 13 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 8 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 – that link directly to the interest rate series – are shrunk to zero with posterior probability close to one across the entire period.

Figure 9 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 5 displays posterior means of impulse response for one-, two- and three-year ahead horizons. In this comparison, we 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 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 17 reports impulse response with credible intervals computed from posterior

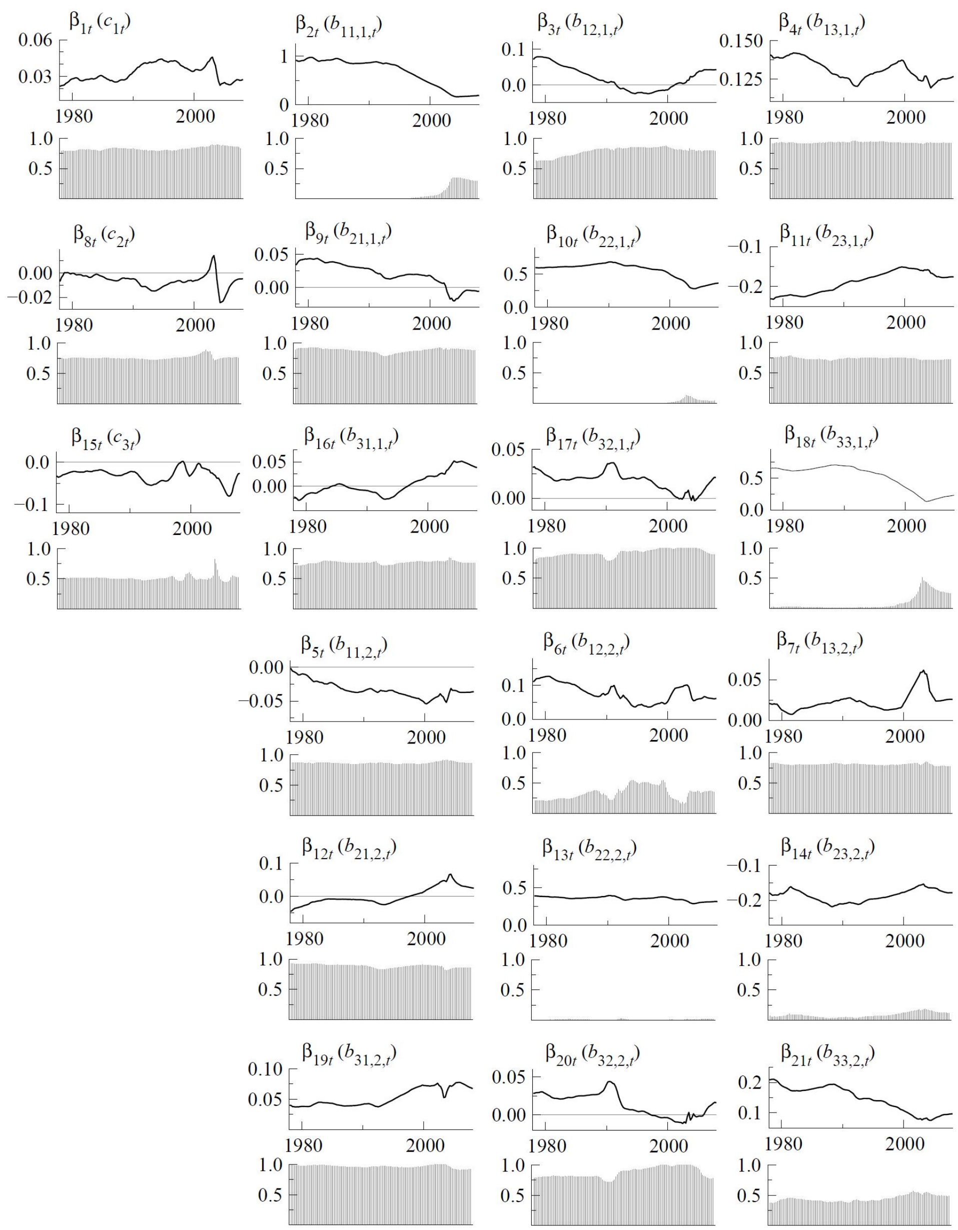


Figure 13: Posterior means of $\boldsymbol{\beta}_t$ for Japanese macroeconomic data. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory. The corresponding indices of $\boldsymbol{c}_t$ or $\boldsymbol{B}_{\ell t}$ are in parentheses.

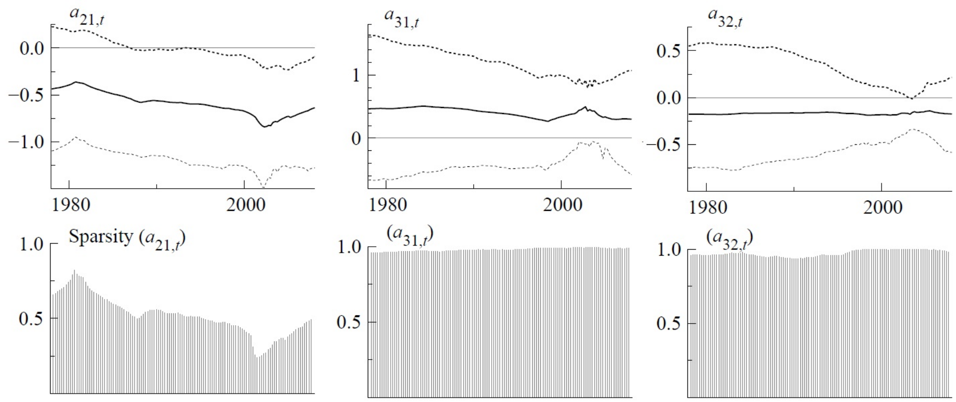


Figure 14: Posterior trajectories of $a_{ij,t}$ for Japanese macroeconomic data: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

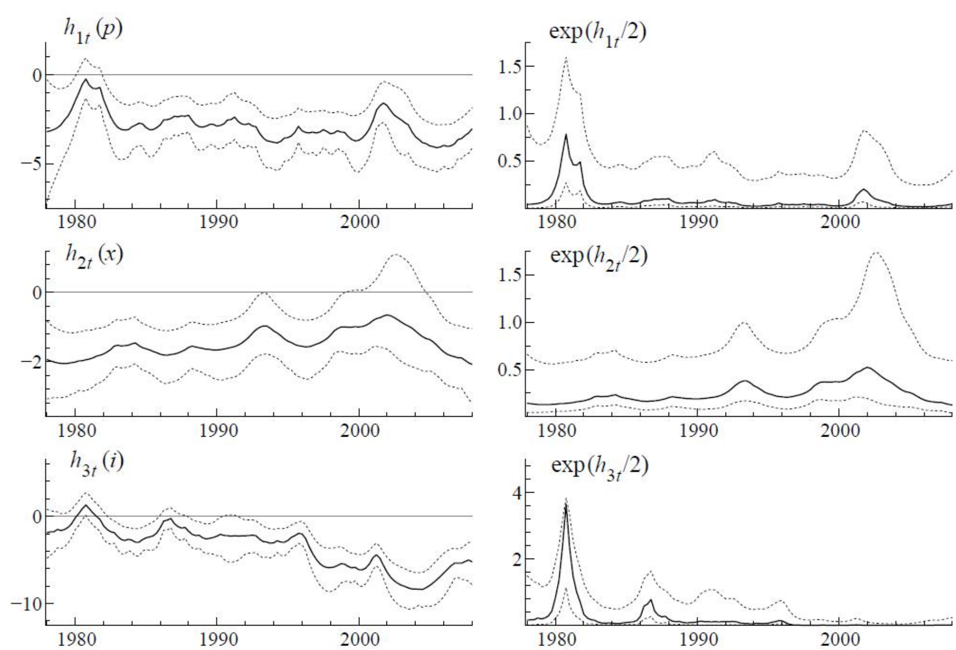


Figure 15: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$ for Japanese macroeconomic data: posterior means (solid) and 95% credible intervals (dotted).

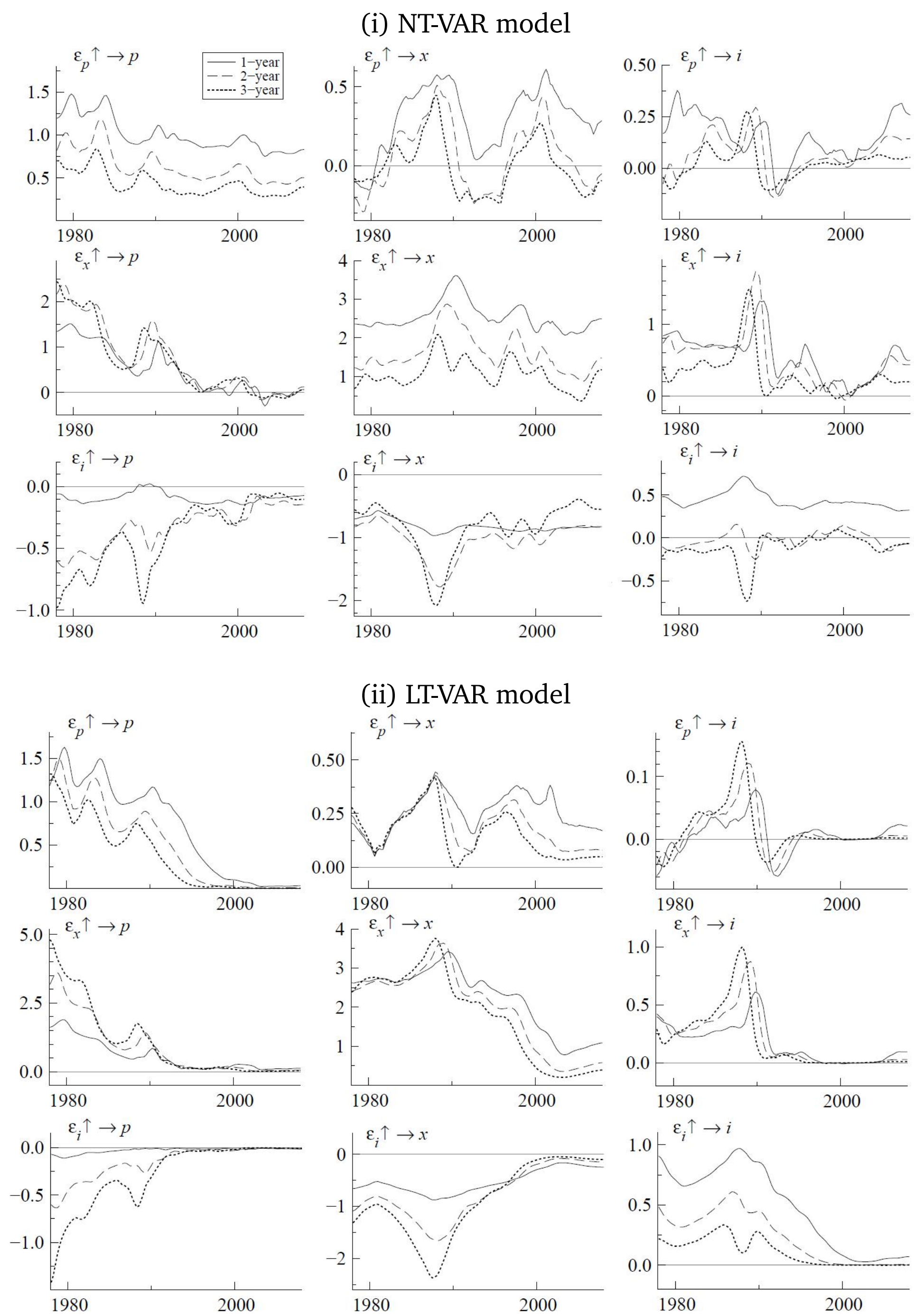


Figure 16: Impulse response trajectories for one-, two- and three-year ahead horizons from the VAR model (upper) and LT-VAR model (lower) for Japanese macroeconomic data. The symbols $\varepsilon_a\uparrow \to b$ refer to the response of the variable b to a shock to the innovation of variable a. The shock size is set equal to the average of the stochastic volatility across time for each series.

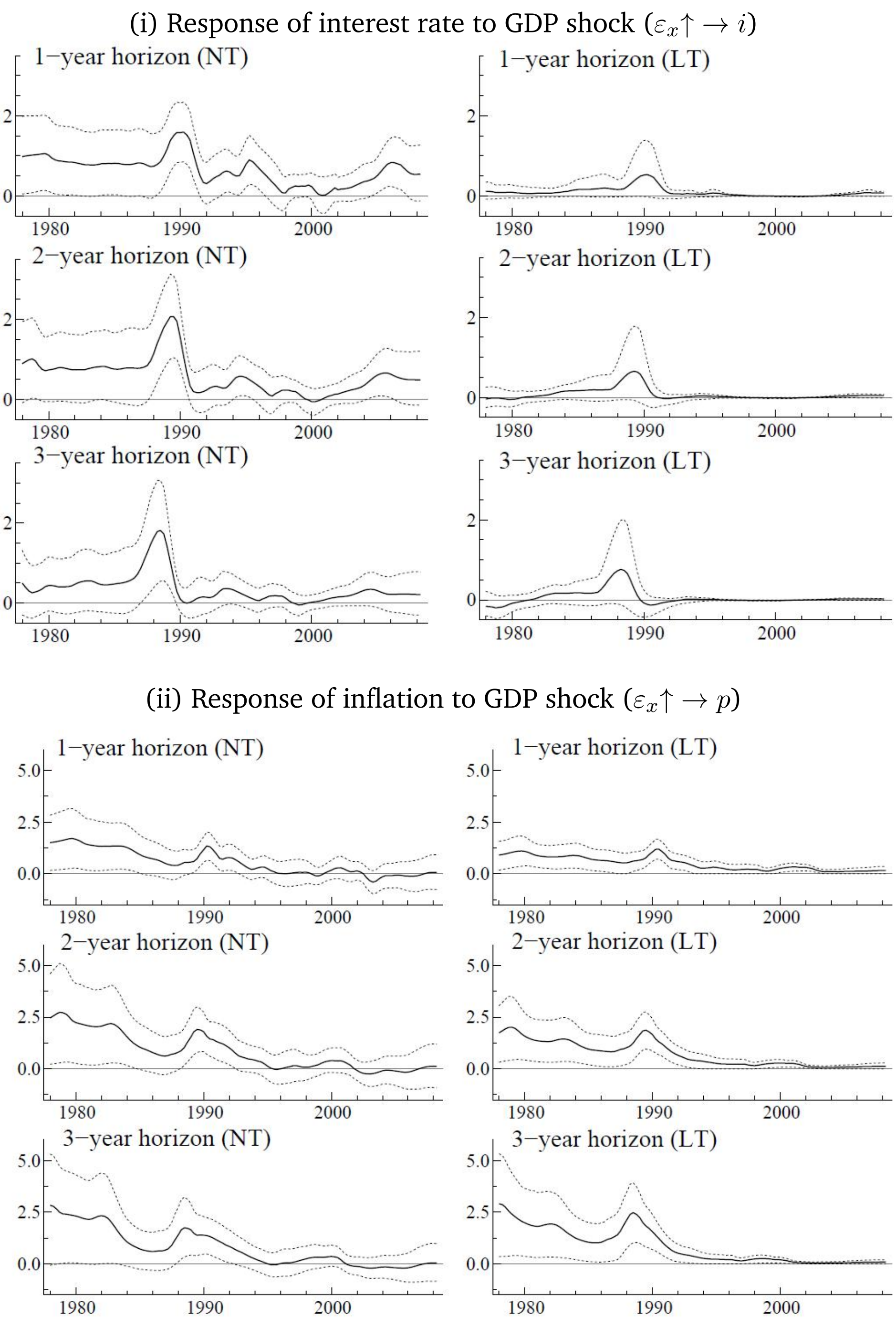


Figure 17: Impulse response trajectories with credible intervals from the NT-VAR (left) and LT-VAR (right) models for Japanese macroeconomic data. Posterior median (solid) and 75% credible intervals (dotted).

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 and theirs spread 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.