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*

Abstract

We introduce and develop a general approach to dynamic sparsity modelling in multivariate time series analysis. Time-varying parameters are linked to latent processes that are thresholded to induce zero values adaptively, providing natural mechanisms of dynamic variable inclusion/selection. We discuss Bayesian analysis and prediction in dynamic regressions, time-varying vector autoregressions and multivariate volatility models using latent thresholding. Application in analyses of topical macroeconomic time series problems shows the utility of this 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.

1 Introduction

For analysis of increasingly high-dimensional time series in many areas of business and economics, dynamic modelling strategies are pressed by the needs to appropriately constrain and reduce parameter dimensions. Approaches to introducing parsimonious structure into increasingly highly parametrized models for multivariate time series are in particular needed to reflect common contexts in which relationships between variables and latent processes may change over time, being practically relevant in some time periods yet irrelevant or redundant in others.

We address these general questions with a new approach based on *latent threshold models* (LTMs), that defines a strategy of fairly broad utility. A first example is dynamic regression modelling, itself an already broadly used applied time series framework as well as an illustration of the LTM strategy for the richer class of dynamic linear models (DLMs – West and Harrison 1997; Prado and West 2010). From this basis, we develop the LTM ideas and methodology in time-varying vector autoregressive (TV-VAR) models including time-varying multivariate volatility matrix analysis (Aguilar and West 2000). In applications of such modelling approaches, the latent, time-varying parameter processes they involve grow quickly in dimension with the number of time series response variables; this often leads to increased uncertainty in estimation that degrades predictive performance. Modelling strategies that induce 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, can reduce estimation uncertainties substantially and lead to improved predictive ability and model interpretation. The LTM approach is explicitly designed to encourage such dynamic sparsity. Examples in macroeconomic time series analysis illustrate this, as well as highlighting connections with other approaches, open questions and potential extensions.

Much progress has been made in recent years in the general area of Bayesian sparsity modelling: developing model structures via hierarchical priors that are able to induce shrinkage to zero of subsets of parameters in multivariate models. Among developments of most relevance here are the nowadays standard use of sparsity priors for regression model uncertainty and variable selection (George and McCulloch 1993, 1997; Clyde and George 2004) in areas including sparse factor analysis (West 2003; Carvalho et al. 2008; Lopes et al. 2010; Yoshida and West 2010) and inherently sparse graphical modelling (Jones et al. 2005); they have also been applied to traditional time series models including constant coefficient VAR models (e.g. George et al. 2008; Chen et al. 2010). In both dynamic regression and Bayesian approaches in multivariate volatility models, these general strategies have been usefully applied to induce full shrinkage to zero of effects *globally* – that is, zeroing out regression coefficients in a time series model for all time (e.g. Carvalho and

West 2007; George et al. 2008; Korobilis 2010; Wang 2010). We build on these earlier works but go much further to address the much more general question of time-varying inclusion of effects, i.e. dynamic sparsity modelling; the LTM strategy provide flexible, local and adaptive (in time) data-dependent variable selection for dynamic regressions and autoregressions, and for volatility matrices as components of more elaborate multivariate dynamic models.

There are, of course, at least superficial connections with previous work on time series using threshold ideas for as mechanisms in regime switching and tobit/probit models (e.g. West and Mortera 1987; Polasek and Krause 1994; Wei 1999; Galvao and Marcellino 2010), and with time series 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 parameters governed by latent indicator variables. The LTM approach differs fundamentally from these approaches in that threshold mechanisms operate continuously in the parameter space and over time based on the values of underlying latent parameter processes. This temporal determination structure leads to a new paradigm and practical strategies for threshold modelling in time series analysis, with broad potential utility as our examples indicate.

Some notation: We use f, g, p for density functions, and the distributional notation $\boldsymbol{y} \sim N(\boldsymbol{a}, \boldsymbol{A})$, $d \sim U(a,b)$, $p \sim B(a,b)$, $v \sim G(a,b)$, for the normal, uniform, beta, and gamma distributions, respectively. We also use, for example, $N(\boldsymbol{y}|\boldsymbol{a}, \boldsymbol{A})$ to denote the actual density function $f(\boldsymbol{y})$ when $\boldsymbol{y} \sim N(\boldsymbol{a}, \boldsymbol{A})$, in cases where the specificity is needed. 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 $s < t$, for succinct subscripting such as in use of $\boldsymbol{y}_{1:T}$ to denote $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$.

2 Latent threshold modelling

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 $\boldsymbol{s}_t \equiv (s_{1t}, \dots, s_{kt})'$ 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 idea is simply that the i^{th} variable x_{it} has time-varying coefficient whose value is shrunk to zero when it falls below a thresholded, neatly embodying sparsity/shrinkage and parameter reduction when relevant in the dynamic regression context; see Figure 1. The shrinkage region $(-d_i, d_i)$ defines what we can refer to as 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*.

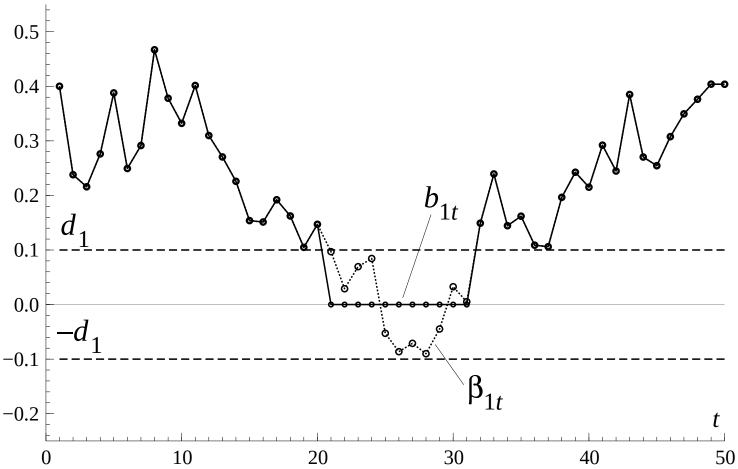


Figure 1: Illustration of LTM concept: The dynamic regression coefficient process β_{1t} arises as a thresholded version of an underlying dynamic coefficient time series.

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 taken here for example. That is,

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

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

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

Throughout we will denote the hyper-parameters of these univariate AR models by $\boldsymbol{\theta}_i = (\mu_i, \phi_i, \sigma_i)$ with $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\}$. We refer to the model of eqns. (1-3) as a *Latent Threshold Model* (LTM). Note that if $d_i \equiv 0$, the LTM reduces to a standard special case of the DLM.

2.2 Threshold parameters and sparsity

The LTM structure applies in multiple classes of dynamic models as developed and illustrated below. First, we discuss the role of threshold parameters and specification of priors for them. We continue in the dynamic regression model example in the special case of $k = 1$ for definiteness; this loses no generality as the discussion applies to thresholding of time-varying parameters in all the other classes of models. This simplest 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– better named as "regression model uncertainty analysis"– to the time series context. Standard Bayesian model selection methods assign a non-zero prior probability to zero values of regression parameters, and continuous priors centered at zero otherwise. The extension to time-varying parameters requires a non-zero *marginal* prior probability $\Pr(b_{1t} = 0)$ coupled with a continuous prior on non-zero values, but that also respects the time series context and 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 in the parameters, while the thresholding mechanism induces persistence over periods of time in the occurrences of zero/non-zero values in the effective coefficients b_{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 in regression (George and McCulloch 1993; Clyde and George 2004; Carvalho et al. 2008).

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 $p(y_t|d_1) = \pi_1 g(y_t) + (1 - \pi_1)N(y_t|0, 1)$ with $\pi_1 = \Pr(|\beta_{1t}| \geq d_1) = 2\Phi(-d_1/v_1)$ where Φ is the standard normal cdf, and with density function $g(y_t) = \pi_1^{-1}h(y_t)N(y_t|0, 1 + v_1^2)$ where $h(y_t) = 1 - \Phi((d_1 - a^2y_t)/a) + \Phi((-d_1 - a^2y_t)/a)$ with $a^2 = v_1^2/(1 + v_1^2)$. This continuous mixture density smoothly transitions from the zero-mean normal component when d_1 is large, to that from the regression model with the normal prior on β_{1t} when d_1 is small or zero. The fundamental sparsity probability is now seen to be $\Pr(b_{1t} = 0) = 1 - \pi_1 = 2\Phi(d_1/v_1) - 1$. This also indicates the importance of the scale v_1 in considering relevant values of the threshold d_1 and in specifying priors over thresholds generally; 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 $d_1 = 0.063 v_1$; 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 will assign priors over the threshold and use this line of reasoning about the fundamental role of d_1 to do so. A neutral prior will admit and 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.

Reverting to the more general model multiple regression parameter processes β_{it} and mean parameters μ_i, this prior specification extends to each of 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

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

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 $v_i = 1$, 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 in the posterior for primary time-varying parameters and the resulting predictive distributions. Correspondingly, there is no interest in the underlying values of the latent processes β_{it} when they are below threshold; all that matters is that we then know $b_{it} = 0$. Depending on the model and data context, any one of the threshold parameters may be well-estimated, or identified, in the sense of the posterior being rather precise relative to the prior; in contrast, the data may be basically uninformative about other thresholds, so that their posteriors are minor modifications of their priors. In the dynamic regression context, as the running 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 very high levels of thresholding on β_{it}. Then posterior will suggest larger values of d_i while providing no information about the β_{it} process; we have no interest in the β_{it} process values below threshold in this context, while the posterior provides full inferences 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. There we have 3 regressors; for 2 of the regressors, the latent coefficient processes exceed thresholds for reasonable periods of time, and the posterior shows clear and strong evidence for low threshold values. The third regression coefficient process stays below threshold, and the posterior on the threshold d_3 is appropriately very close to the prior, favoring only very slightly larger values, 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 based on this detection and estimation of time-varying shrinkage, are marginal with respect to thresholds; the posteriors for primary model parameters integrate over the posterior uncertainties about thresholds whatever the nature of prior-to-posterior updates for the threshold parameters themselves. Again, the latter are simply vehicles to inducing relevant time-varying sparsity mechanisms, and whether or not they are well-estimated from the data is of no intrinsic importance or interest.

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, note that sampling of 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 $\boldsymbol{s}_t = \mathbf{1}$; 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. de Jong and Shephard 1995; Durbin and Koopman 2002) 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 posterior under the LTM. We have also 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 very efficient and so can be run far longer than blocking approaches to balance mixing issues; further, we have experienced practically acceptable acceptance rates and mixing in multiple examples illustrated here, and hence adopt it as standard.

It is worth noting that TV-VAR coefficients can, at any time point, define VAR structures with explosive autoregressive roots, depending on their values. In studies such as our macroeconomic examples, this is unrealistic and undesirable (e.g., Cogley and Sargent 2001). This can be avoided by modifying the structure to assign zero prior probability to explosive roots, corresponding to constraints on the conditional posteriors for the $\boldsymbol{\beta}_t$. In the analysis of the LT-VAR models, we can then apply this restriction in the single-move sampler for $\boldsymbol{\beta}_t$, adding a rejection sampling step in generating the candidates. We note 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 candidate 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 compute burden.

3 Simulation example

For illustration, a sample of size $T = 500$ comes from the LTM with $k = 3$ predictors and where only the first two predictors are relevant. The x_{it}'s are generated from i.i.d. $U(-0.5, 0.5)$ and $\sigma = 0.15$, 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 discarding a burn-in period of 5,000 samples. Computations are obtained using Ox (Doornik 2006); code is available to interested readers.

	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.

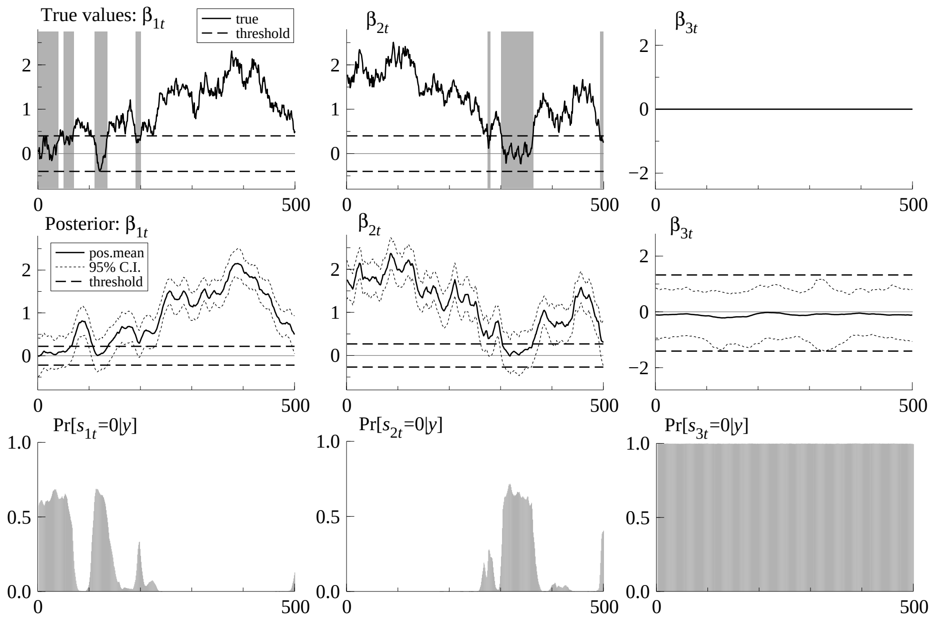


Figure 2: Simulation example: Trajectories of dynamic regression parameters. True values (top), posterior means, 95% credible intervals (middle), and posterior probabilities of $s_{it} = 0$ (bottom). The shadows in the top panels refer to the periods when $s_{it} = 0$. The thresholds in the middle panels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

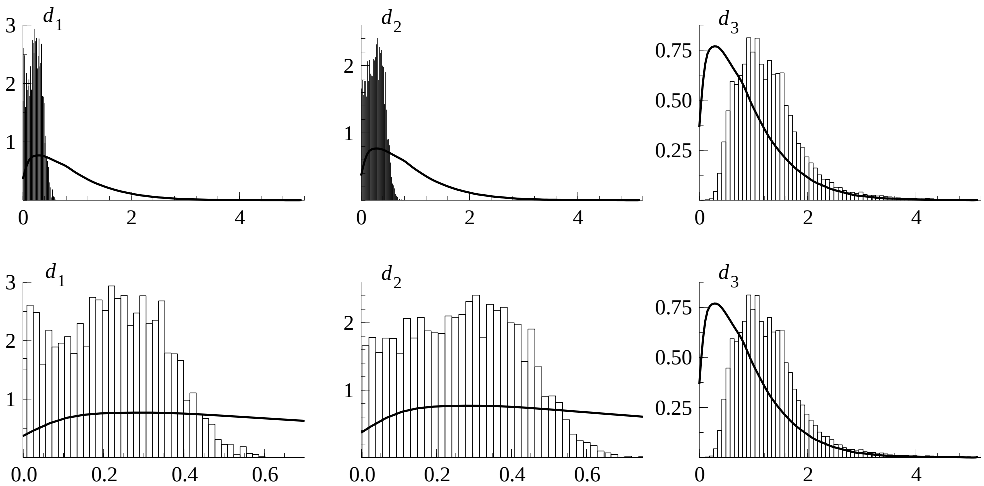


Figure 3: Simulation example: Priors (solid lines) and posteriors (histograms) for thresholds. In the lower row, the graphs are simply truncated on the d_i axes for visual clarity.

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 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 impact on these primary inferences. 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 benefit and relevance, while for general application we adopt $K = 3$ as a global, neutral 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 modelling are becoming increasingly prevalent in econometrics (e.g. Cogley and Sargent 2005; Primiceri 2005) as in other fields. With increasingly high-dimensional response series, the number of coefficients in VAR model autoregressive coefficient matrices escalates as does the need for parameter constraints. Recent Bayesian VAR analysis, address this using shrinkage and sparsity-inducing priors of various forms (Fox et al. 2008; George et al. 2008; Wang 2010) for traditional constant coefficient VAR models, but the induction of zeros into increasingly sparse *time-varying* coefficient matrices, with allowance for time-variation in the occurrence of non-zero values as well as local changes in coefficients when they are non-zero, is a challenging and open problem. The LTM ideas provide an approach.

4.1 Model structure

For the time series of $m \times 1$ vector responses $\boldsymbol{y}_t$, $(t = 1, 2, \ldots,)$, take the basic TV-VAR model with autoregressive order p as

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

where $\boldsymbol{c}_t$ is the $m \times 1$ vector of time-varying intercept, $\boldsymbol{B}_{jt}$ 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 covariance matrix that is also often time-varying. For each time 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 $\boldsymbol{X}_t = \boldsymbol{I} \otimes (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})$. Then the model can be written as a multivariate dynamic regression, viz.

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

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 covariance matrix

Modelling time-varying covariance 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 financial and macroeconomic data. We build on prior Bayesian modelling approaches here for the TV-VAR innovations volatility matrices.

Consider a triangular reduction of $\boldsymbol{\Sigma}_t$, defined by $\boldsymbol{A}_t\boldsymbol{\Sigma}_t\boldsymbol{A}'_t = \boldsymbol{\Lambda}_t^2$, where $\boldsymbol{A}_t$ is the lower triangular matrix of covariance components with all diagonal elements equal to one 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 modelling (e.g. West 2003; Carvalho et al. 2008), some recent innovations in sparsity modelling for variance matrices have adopted new priors on elements of triangular square-roots of variance matrices such as $\boldsymbol{A}_t$ defined here. George et al. (2008) do this with priors allowing zero elements in Cholesky components of constant covariance matrices in a constant parameter VAR model context, building on previous, non-sparse approaches utilizing such constructions (e.g. Pinheiro and Bates 1996; Pourahmadi 1999; Smith and Kohn 2002). The construction has also appeared in time-varying variance matrix modelling in VAR contexts (Cogley and Sargent 2005; Primiceri 2005; Lopes et al. 2010) which is one point of departure for us; we use the above Cholesky structure and embed it in a novel LTM framework in the following section, combining models for stochastic time-variation in variance matrices with the natural threshold-based sparsity inducing mechanism to shrink subsets of the lower-triangle of $\boldsymbol{A}_t$ to zero adaptively and dynamically.

The basic time-varying model for the Cholesky parameters is as 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 $\boldsymbol{h}_t = (h_{1t}, \dots, h_{kt})'$ where $h_{jt} = \log \sigma_{jt}^2$, for $j = 1, \dots, k$. The dynamics of the covariances and variances are specified jointly with the time-varying VAR coefficients $\boldsymbol{\beta}_t$ as

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

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

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

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, chap. 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 $\tilde{\boldsymbol{y}}_t = (\tilde{y}_{1t}, \ldots, \tilde{y}_{kt})' = \boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}_t$ and the $m \times m(m-1)/2$ matrix

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

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 matrix

We now proceed to introduce the LTM structure for time-varying variance matrices, to enable shrinkage to zero of subsets of the elements of $\boldsymbol{a}_t$ over periods of time consistent with sparse structure. This simply directly adapts the LTM strategy from Section 2 into this context, applying the latent thresholding ideas now 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 $\alpha_{ij,t}$'s 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 now 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 $\mathbf{\Phi}_{\boldsymbol{\alpha}}$ 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 $\mathbf{\Omega}_t = \boldsymbol{A}_t' \mathbf{\Lambda}_t^{-2} \boldsymbol{A}_t$; the more zeros there are in the lower triangle of $\boldsymbol{A}_t$, the more zeros there will be in the precision matrix. Hence the LTM defines an approach to time-varying sparsity modelling for precision matrices, and hence time-varying 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 now clearly defines a new class of models for time-variation in the *structure* of the graphical model underlying $\mathbf{\Sigma}_t$, since the appearance of zeros in its inverse $\mathbf{\Omega}_t$ 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.

Note finally that, from a full MCMC analysis of the model, we will recover 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 an assessment of the posterior probabilities of edge inclusion at each time t as well.

5 Application to US macroeconomic data

5.1 Introduction and literature

The use of Bayesian analyses of TV-VAR models is becoming increasingly common in analyses of macroeconomic data. Recent works such as Primiceri (2005), Benati (2008), Benati and Surico (2008), Koop et al. (2009) and Nakajima et al. (2011), for example, 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 C, 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 analyses yield intuitive and interpretable results, particularly with respect to inferred impulse response functions; and *(iii)* again relative to the standard, non-thresholded models, the LTM analyses yield practically significant improvements in multi-step, out-of-sample predictions, these being particularly relevant to policy uses of such models. The consonance of results across the US and Japanese studies can also be regarded as additionally indicative of the relevance of the LTM concepts in these kinds of applications.

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 (2010) developed Bayesian variable selection for TV-VAR coefficients as well as structural breaks. Chan et al. (2011) 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 and priors

We analyze the $m = 3$ time series giving the quarterly inflation rate, unemployment rate and short-term nominal interest rate in the US economy during 1963/Q1–2001/Q3; this is a time series that is of some topical interest and has been previously studied by a number of authors (Cogley and Sargent 2005; Primiceri 2005; Koop et al. 2009). The 3 variables are ordered this way to define $\boldsymbol{y}_t$ for $t = 1, \ldots, T = 155$ quarters, 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 Treasure bills; see Figure 4.

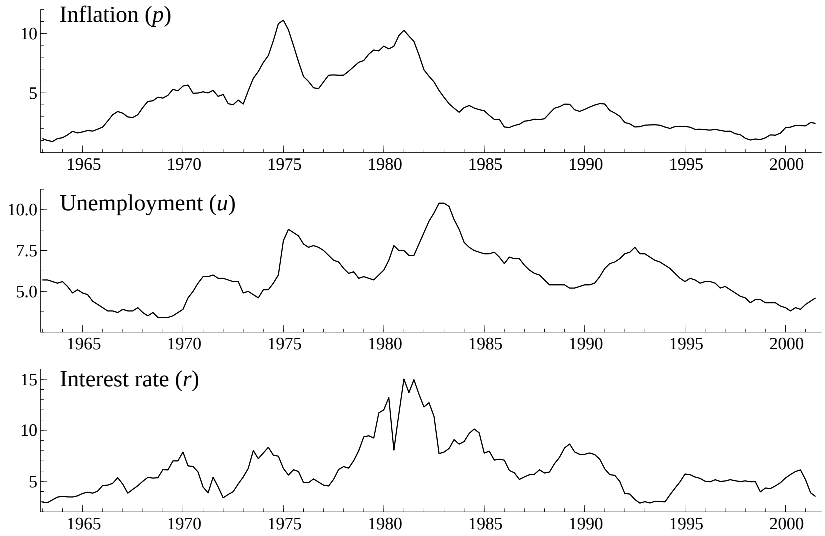


Figure 4: US macroeconomic time series (indices $\times 100$ for % basis).

For all model analyses reported we take the following prior components. With $v^2_{\boldsymbol{\beta}i}$, $v^2_{\boldsymbol{\alpha}i}$ and $v^2_{\boldsymbol{h}i}$ denoting the i^{th} diagonal elements of $\boldsymbol{V}_{\boldsymbol{\beta}}$, $\boldsymbol{V}_{\boldsymbol{\alpha}}$ and $\boldsymbol{V}_{\boldsymbol{h}}$, respectively, we use $v^{-2}_{\boldsymbol{\beta}i} \sim G(20, 0.01)$, $v^{-2}_{\boldsymbol{\alpha}i} \sim G(2, 0.01)$ and $v^{-2}_{\boldsymbol{h}i} \sim G(2, 0.01)$. 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 Forecasting performance and comparisons

We fit and compare predictions using both the new LT-VAR and the standard, non-threshold (NT) TV-VAR models; in each case, we use the LT/NT time-varying variance matrices of Section 4.3. Specifically, we consider four TV-VAR models with the following settings: (1) NT for both $\boldsymbol{b}_t$ and $\boldsymbol{a}_t$, (2) NT for $\boldsymbol{b}_t$ and LT for $\boldsymbol{a}_t$, (3) LT for $\boldsymbol{b}_t$ and NT for $\boldsymbol{a}_t$, and (4) LT for both $\boldsymbol{b}_t$ and $\boldsymbol{a}_t$; all the models have stochastic volatility $\{\boldsymbol{h}_t\}$. Model (1) is almost equivalent to a commonly used TV-VAR of Primiceri (2005); the only difference is that Primiceri (2005) assumes random walk process for the time-varying parameters rather than the stationary AR processes we adopt.

The MCMC analyses were applied to each of the resulting models, and analysis was repeated across models with different maximum lags $p = 1, 2, 3$ or 4 in the VAR for the series. We evaluated each of the resulting suite of models by fitting each model to the data from 1963/Q1–2000/Q3 and then forecasting (by simulation from the posterior predictive distributions implied) over the final $1, 2, 3$ and 4 quarters. Based on evaluation of root mean squared forecast errors (RMSFE) across these out-of-sample forecasts, we found that all the VAR models perform best when $p = 3$ is assumed, which is taken in the following analysis

Model			Horizon (quarters)			
	$\boldsymbol{b}_t$	$\boldsymbol{a}_t$	1	2	3	4
RMSFE						
(1)	NT	NT	0.234	0.336	0.491	0.687
RMSFE relative to Model (1)						
(2)	NT	LT	0.995	0.995	0.996	0.993
(3)	LT	NT	0.957	0.979	0.984	0.916
(4)	LT	LT	0.946	0.912	0.911	0.864

Table 2: Forecasting performance for US macroeconomic data: RMSFE for one- to four- quarter ahead prediction. NT and LT refer to the non-threshold and latent threshold models, respectively.

Extensive predictive evaluation and model comparisons are summarized in Table 2. We evaluated the competing models based on their forecasting performance for twenty different selections of subsets of data to hold-out for prediction. We begin with the sample period from 1963/Q1–1995/Q4, fit the model and then forecast one- to four-quarters ahead over 1996/Q1–Q4. We next use the sample period 1963/Q2–1996/Q1 and forecast the following four quarters. We repeat this rolling estimation, moving ahead one quarter at time to obtain twenty sample periods and sets of forecasts over the following year for each. This tests the predictive ability of the various models in different time periods to generate a detailed set of comparisons under different economic conditions and regimes. The table reports the resulting average RMSFE measures for each forecast horizon. The LT-$\boldsymbol{b}_t$ models (Models 3 and 4) uniformly outperform the NT-$\boldsymbol{b}_t$ models (Models 1 and 2), indicating that the time-varying shrinkage structure on coefficients evidently contributes to the forecasting performance of time-varying VARs. The LT-($\boldsymbol{b}_t$,$\boldsymbol{a}_t$) model (Model 4) performs the best, with as much as 14% improvement from the standard TV-VAR model (Model 1) of Primiceri (2005) at four-quarter horizon. Some additional insights into contributions to the model fit and forecasting edge generated by the LTM strategies, for both predictive model parameters and volatility matrices, can be seen in the posterior summaries that follow.

5.4 Some summaries of posterior inferences

We report results from the LT-VAR model where both $\boldsymbol{b}_t$ and $\boldsymbol{a}_t$ follow the LT structure with stochastic volatility process $\boldsymbol{h}_t$ (Model 4). Figure 5 displays the posterior probabilities of $s_{it} = 0$ for coefficients $\boldsymbol{b}_t$. This comes from a full MCMC analysis of the entire time series. In the figure, $b_{ij,\ell,t}$ denotes the (i, j) element in the coefficient matrix $B_{\ell t}$. Time-varying sparsity is observed for several coefficients, with estimated values shrinking to zero for some time periods but not others, whereas other coefficients are automatically selected out over the entire time-frame.

Figure 6 plots the posterior means, the 95% credible intervals of the time-varying Cholesky covariance elements $a_{ij,t}$, as well as the posterior probabilities of $s_{aij,t} = 0$. The $a_{21,t}$ is entirely shrunk to zero, whereas the other two covariances, associated with interest rates, are roughly 80-90% distinct from zero over most of the time frame with stable time variation in their values based on the posterior means. Considerable sparsity for only one covariance element leads to relatively smaller improvements in predictive performance than that for $\boldsymbol{b}_t$; the notable time-varying shrinkage across the latter coefficients drives more significant predictive improvements.

In the LT-VAR model, setting $s_{it} = 1$ for all (i,t) and $s_{aij,t} = 1$ for all (i,j,t) reduces 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 $s_{*,t}$ in the figures, 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 7 graphs the posterior means of the stochastic volatility, h_{it} and $\exp(h_{it}/2)$, together with their 95% credible intervals. Several volatile periods are observed for three series reflecting non-systematic shocks hitting the economy. After volatile periods in the 1970s, the stochastic volatility clearly decays, implying a possible source of the Great Moderation as discussed in the literature (Cogley and Sargent 2005; Primiceri 2005).

A final summary of practical interest relates to impulse response analysis. We evaluate this to compare the LT-VAR and NT-VAR models. For models with time-varying coefficients, we need to compute impulse responses repeatedly for each quarter as we move the time window across a sample period. There are several ways to explore impulse responses; 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. The comparative NT-VAR analysis has both $\boldsymbol{b}_t$ and $\boldsymbol{a}_t$ following non-thresholded models and with stochastic volatility $\{\boldsymbol{h}_t\}$ (Model 1).

Figure 8 displays posterior means of the impulse response for one-, two- and three-year ahead horizons. The NT-VAR provides similar responses to 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 trajectory of the response from the LT-VAR model is smoother than that of the NT-VAR. Effective shrinkage in the TV-VAR coefficients and

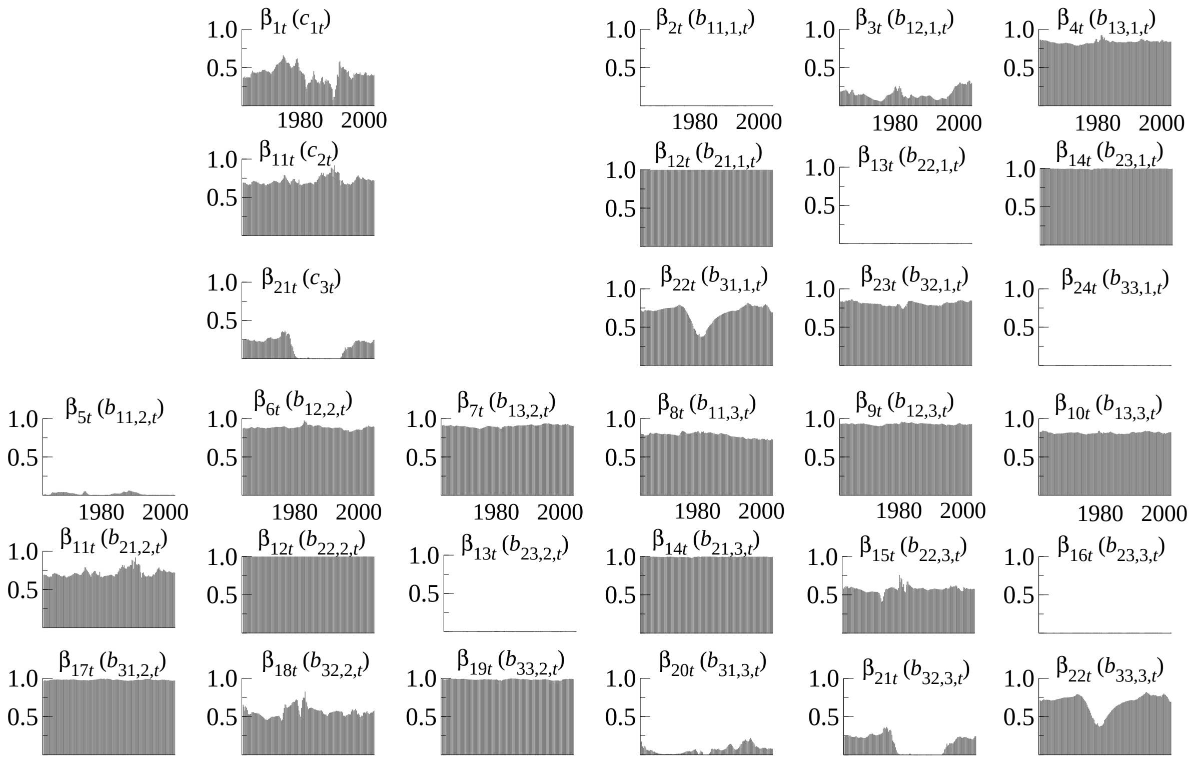


Figure 5: Posterior probabilities of $s_{it} = 0$ for US macroeconomic data. The corresponding indices of $\boldsymbol{c}_t$ or $\boldsymbol{B}_{\ell t}$ are in parentheses.

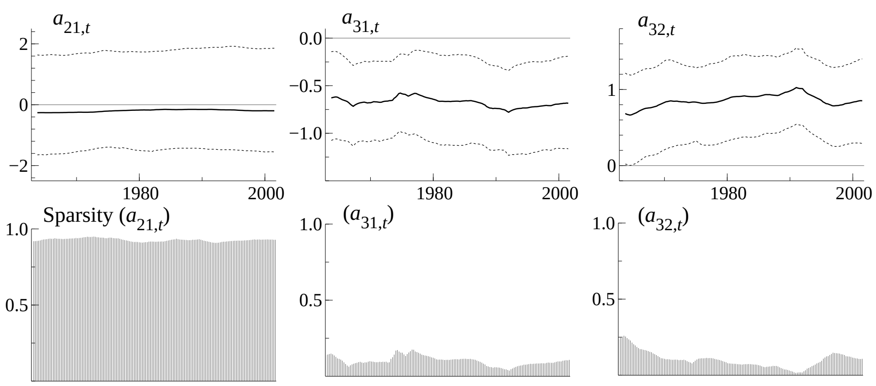


Figure 6: Posterior trajectories of $a_{ij,t}$ for US macroeconomic data: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

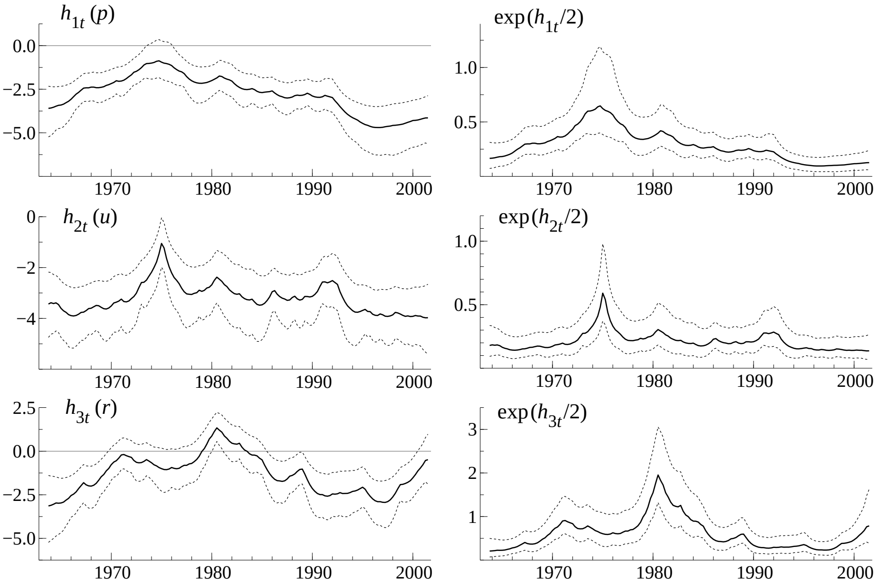


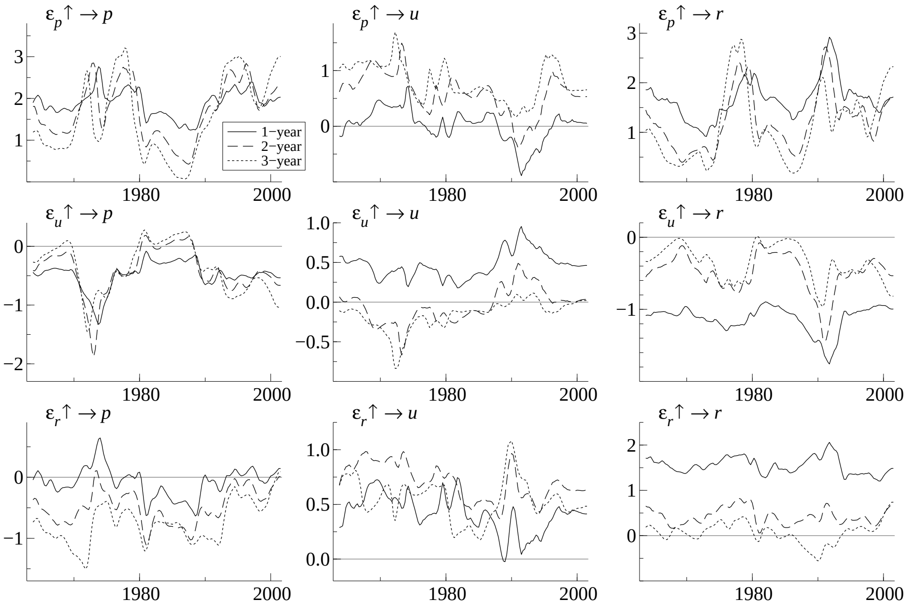
Figure 7: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$ for US macroeconomic data: posterior means (solid) and 95% credible intervals (dotted).

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 that from the NT-VAR analysis, being clearly shrunk towards zero due to the LTM structure. Based on the significant improvements in step-ahead 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 the Appendix in the on-line Supplementary Material.

6 Summary comments

By introducing latent threshold process modelling as a general strategy, we have defined a approach to time-varying parameter time series that overlays several existing model classes. The LTM approach provides a general framework for time-varying sparsity modelling, 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 dynamics 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 in-

(i) NT-VAR model



(ii) LT-VAR model

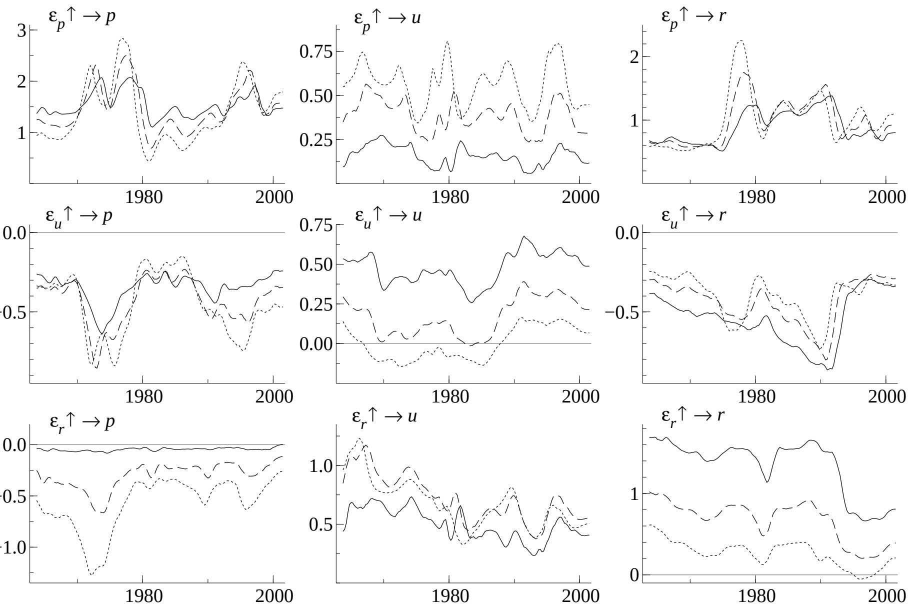


Figure 8: Impulse response trajectories for one-, two- and three-year ahead horizons from the VAR model (upper) and LT-VAR model (lower) for US macroeconomic data. The symbols $\varepsilon_a \uparrow \rightarrow b$ refer to the response of the variable b to a shock to the innovation of variable a. The shock size is set equal to the average of the stochastic volatility across time for each series.

cluding, 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 substantive example in analysis of a US macroeconomic time series, and a companion example with related Japanese data, illustrate the practical use and impact of the LTM structure. Data induced dynamic sparsity feeds through to substantial improvements in forecasting performance and contextually reasonable shrinkage of inferred impulse response function. The Japanese example investigates the dynamic relationship among three macroeconomic variables including the zero interest rate periods. The LTM structure naturally adapts to these zero-value data periods by eliminating unnecessary fluctuations of time-varying parameters that arise in standard time-varying parameter 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 these kinds of models in short-term forecasting and portfolio analyses for financial time series data are critical, and further application to detailed portfolio decision problems represents a key future direction (Nakajima and West 2011).

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.

Finally, we note that software implementing all analyses discussed in the paper is freely available from the authors to interested readers.

Acknowledgements

The authors thank the editor, the associate editor, an anonymous referee, Francesco Bianchi and Herman van Dijk, for helpful comments and suggestions.

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. (2011), "Time varying dimension models," CAMA Working Paper 28/2011, The Australian National University.

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.

Durbin, J. and Koopman, S. J. (2002), "Simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–616.

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*.

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," Manuscript.

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. (2010), "VAR forecasting using Bayesian variable selection," Working Paper, The Rimini Centre for Economic Analysis (RCEA), Universite Catholique de Louvain.

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. (2011), "Dynamic factor volatility modeling: A Bayesian latent threshold approach," Working Paper 11-16, Department of Statistical Science, Duke University.

Niemi, J. B. and West, M. (2010), "Adaptive mixture modelling Metropolis methods for Bayesian analysis of non-linear state-space models," *Journal of Computational and Graphical Statistics*, 19, 260–280.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic volatility with leverage: Fast likelihood inference," *Journal of Econometrics*, 140, 425–449.

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.

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.

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

Appendix
for
Bayesian Analysis of Latent Threshold Dynamic Models

Jouchi Nakajima & Mike West

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 $p(\boldsymbol{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \boldsymbol{d}|\boldsymbol{y}_{1:T})$. 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 $\boldsymbol{\Theta}$ and σ

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

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

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

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

Second, the conditional posterior density of ϕ_i is

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

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

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

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

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

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 $\boldsymbol{s}_t = \mathbf{1}$, 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:

$$\begin{aligned}
\boldsymbol{M}_1^{-1} &= \sigma^{-2}\boldsymbol{x}_1\boldsymbol{x}_1' + \boldsymbol{\Sigma}_{\eta0}^{-1} + \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi},\\
\boldsymbol{m}_1 &= \boldsymbol{M}_1\left[\sigma^{-2}\boldsymbol{x}_1 y_1 + \boldsymbol{\Sigma}_{\eta0}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Phi}\left\{\boldsymbol{\beta}_2 - (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right],\\
\boldsymbol{M}_T^{-1} &= \sigma^{-2}\boldsymbol{x}_T\boldsymbol{x}_T' + \boldsymbol{\Sigma}_\eta^{-1},\\
\boldsymbol{m}_T &= \boldsymbol{M}_T\left[\sigma^{-2}\boldsymbol{x}_T y_T + \boldsymbol{\Sigma}_\eta^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{\beta}_{T-1} + (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{\mu}\right\}\right],
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\eta0} = \mathrm{diag}(v_1^2, \ldots, v_k^2)$. 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 d

We adopt a direct MH algorithm to sample the conditional posterior distribution of d_i, conditional on $(\mathbf{\Theta}, \sigma, \boldsymbol{\beta}_{1:T}, \boldsymbol{d}_{-i}, \boldsymbol{y}_{1:T})$ 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

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

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 $\mathbf{\Theta}_{\boldsymbol{\gamma}} = (\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \mathbf{\Phi}_{\boldsymbol{\gamma}}, \boldsymbol{V}_{\boldsymbol{\gamma}})$ 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 $(\mathbf{\Theta}_{\boldsymbol{\beta}}, \boldsymbol{d}, \boldsymbol{\alpha}_{1:T}, \boldsymbol{h}_{1:T}, \boldsymbol{y}_{1:T})$, $\boldsymbol{\beta}_t$ 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

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

   and the MH acceptance probability is

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

2. Sampling $\boldsymbol{\alpha}_{1:T}$
   Conditional on $(\mathbf{\Theta}_{\boldsymbol{\alpha}}, \boldsymbol{d}_a, \boldsymbol{\beta}_{1:T}, \boldsymbol{h}_{1:T}, \boldsymbol{y}_{1:T})$ 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 $(\mathbf{\Theta}_{\boldsymbol{h}}, \boldsymbol{\beta}_{1:T}, \boldsymbol{\alpha}_{1:T}, \boldsymbol{y}_{1:T})$, 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:

$$\begin{aligned} y_{it}^* &= \exp(h_{it}/2)e_{it}, \\ h_{it} &= \mu_{hi} + \phi_{hi}(h_{i,t-1} - \mu_{hi}) + \eta_{hit}, \\ (e_{it}, \eta_{hit})' &\sim N\left(\mathbf{0}, \mathrm{diag}(1, v_{hi}^2)\right), \end{aligned}$$

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 9 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.

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.

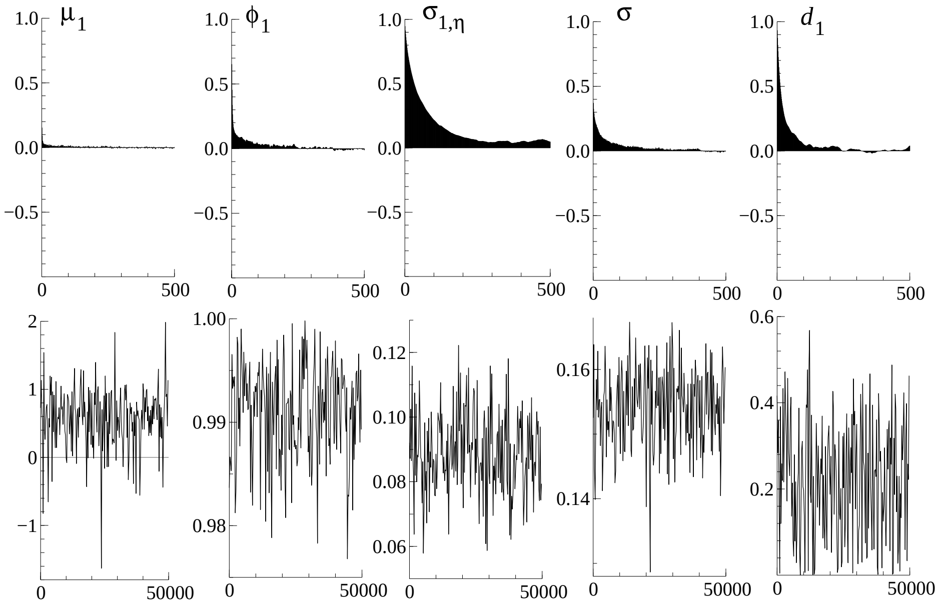


Figure 9: Performance of the MCMC: Autocorrelations (top) and sample paths (bottom) of MCMC draws for selected parameters in simulation example.

	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 Application to Japanese macroeconomic data

C.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 10. 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.

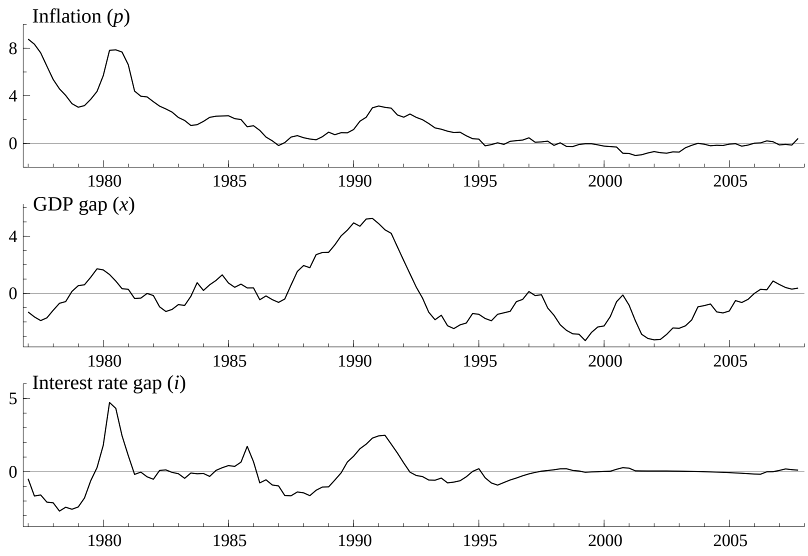


Figure 10: Japanese macroeconomic time series (indices $\times 100$ for % basis).

C.2 Forecasting performance and comparisons

We fit and compare predictions using the same four models as the previous application. Based on evaluation of RMSFE across the out-of-sample forecasts for the final four 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.

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

Table 4: Forecasting performance for Japanese macroeconomic data: RMSFE for one- to four-quarter ahead prediction. NT and LT refer to the non-threshold and latent threshold models, respectively.

The model comparisons are summarized in Table 4. We computed RMSFE for ten different selections of subsets of data, beginning with the sample period from 1977/Q1–2004/Q3, fit the model and then forecast one- to four-quarters ahead over 2004/Q4–2005/Q3 and repeating the rolling estimation as in the previous application. Interestingly, Model 2 (NT-$\boldsymbol{b}_t$, LT-$\boldsymbol{a}_t$) marks slightly larger RMSFEs than Model 1, which indicates little evidence of forecasting improvement by incorporating sparsity only in the covariance component of time-varying volatility matrix. This difference of RMSFE is negligible, although this arises partly because shrinking covariance components may leads to a bias in the time-varying coefficients $\boldsymbol{b}_t$ that follows NT structure in this model. Again, the LT-($\boldsymbol{b}_t$,$\boldsymbol{a}_t$) model (Model 4) dominates all others, confirming that the time-varying shrinkage structure contributes to the forecasting performance of time-varying VARs. Improvement from the standard NT model (Model 1) is much more distinctive than that of the US macroeconomic data, providing almost half of RMSFE at four-quarter horizon. The Japanese data include zero interest rate periods, therefore a benefit from time-varying shrinkage is considered to be larger than the US data.

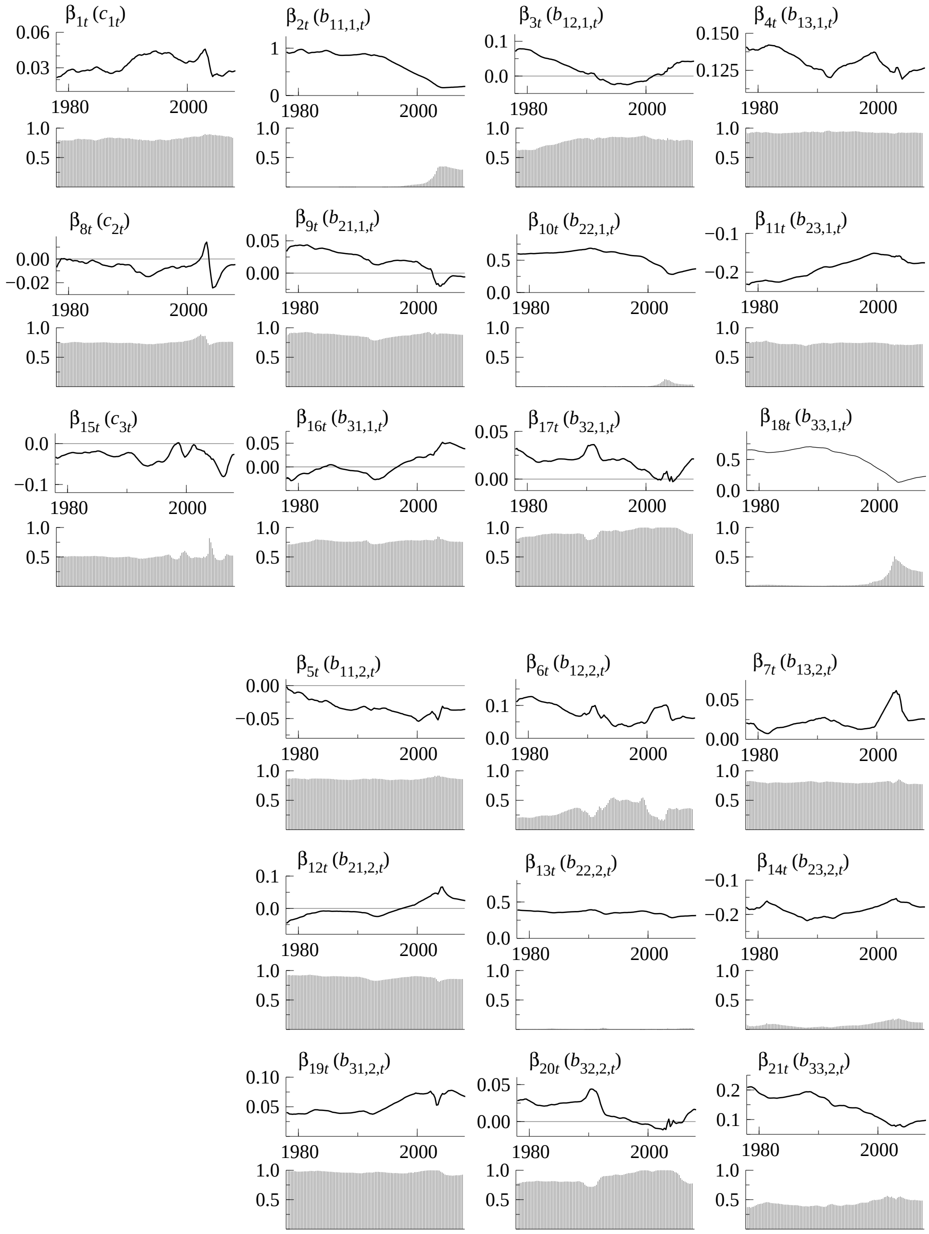


Figure 11: Posterior means of $\boldsymbol{\beta}_t$ for Japanese macroeconomic data. Posterior probabilities of $s_{it} = 0$ are plotted below each trajectory. The corresponding indices of $\boldsymbol{c}_t$ or $\boldsymbol{B}_{\ell t}$ are in parentheses.

C.3 Some summaries of posterior inferences

Figure 11 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 6 plots the posterior means, the 95% credible intervals of $a_{ij,t}$ and the posterior probabilities of $s_{aij,t} = 0$. Here $a_{21,t}$ has a relatively distinctive shrinkage pattern, with a coefficient that varies slightly and is roughly 50% distinct from zero over most of the time frame, whereas the other two elements – that link directly to the interest rate series – are shrunk to zero with posterior probability close to one across the entire period.

Figure 7 graphs the posterior means of the stochastic volatility, h_{it} and $\exp(h_{it}/2)$, together with their 95% credible intervals. Several volatile periods are observed for the inflation and interest rates series around 1980. It is quite understandable and appropriate that the volatility of the interest rate gap series is estimated close to zero during the zero interest rate periods.

Figure 8 displays posterior means of impulse response for one-, two- and three-year ahead horizons. In this comparison, we fitted both the non-threshold 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 step-ahead predictions already discussed.

In addition, Figure 15 reports impulse response with credible intervals computed from posterior draws from NT-VAR and LT-VAR models. Trajectories of posterior median and 75% credible intervals are plotted for responses of the interest rate and inflation to a GDP shock. In both responses, the credible intervals from the LT-VAR model are narrower than that from the NT-VAR 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.

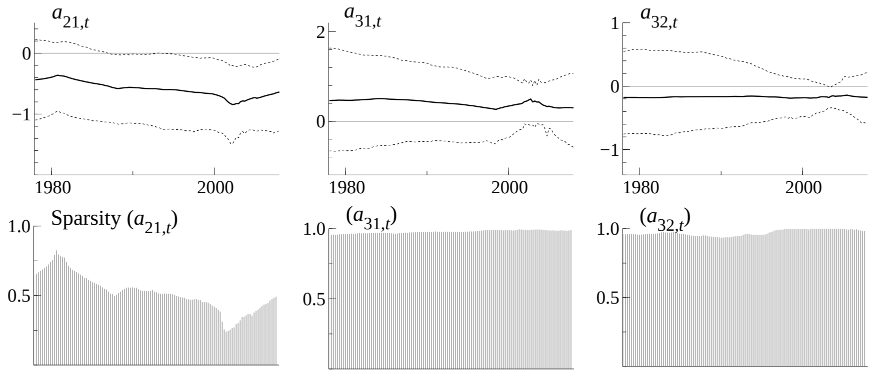


Figure 12: Posterior trajectories of $a_{ij,t}$ for Japanese macroeconomic data: posterior means (solid) and 95% credible intervals (dotted) in the top panels, with posterior probabilities of $s_{aij,t} = 0$ below.

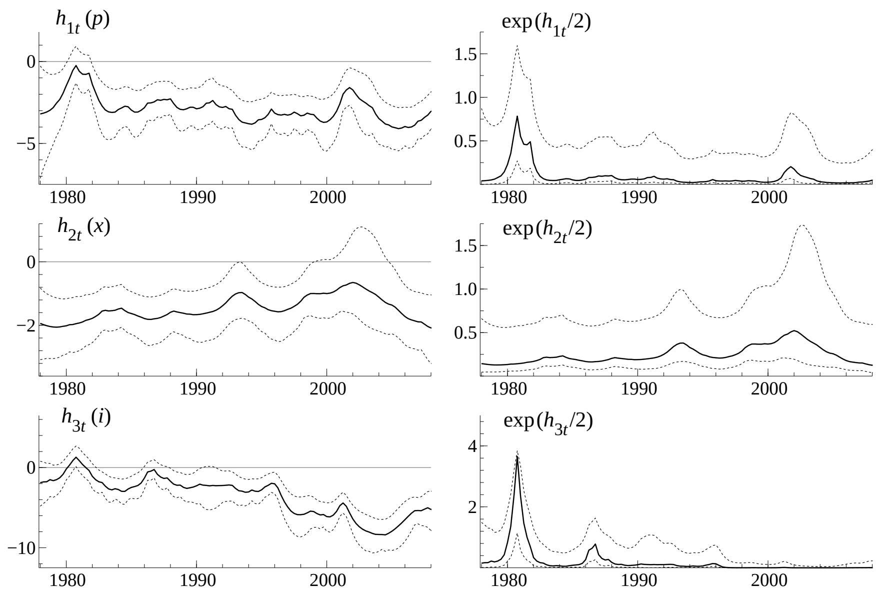


Figure 13: Posterior trajectories of h_{it} and $\exp(h_{it}/2)$ for Japanese macroeconomic data: posterior means (solid) and 95% credible intervals (dotted).

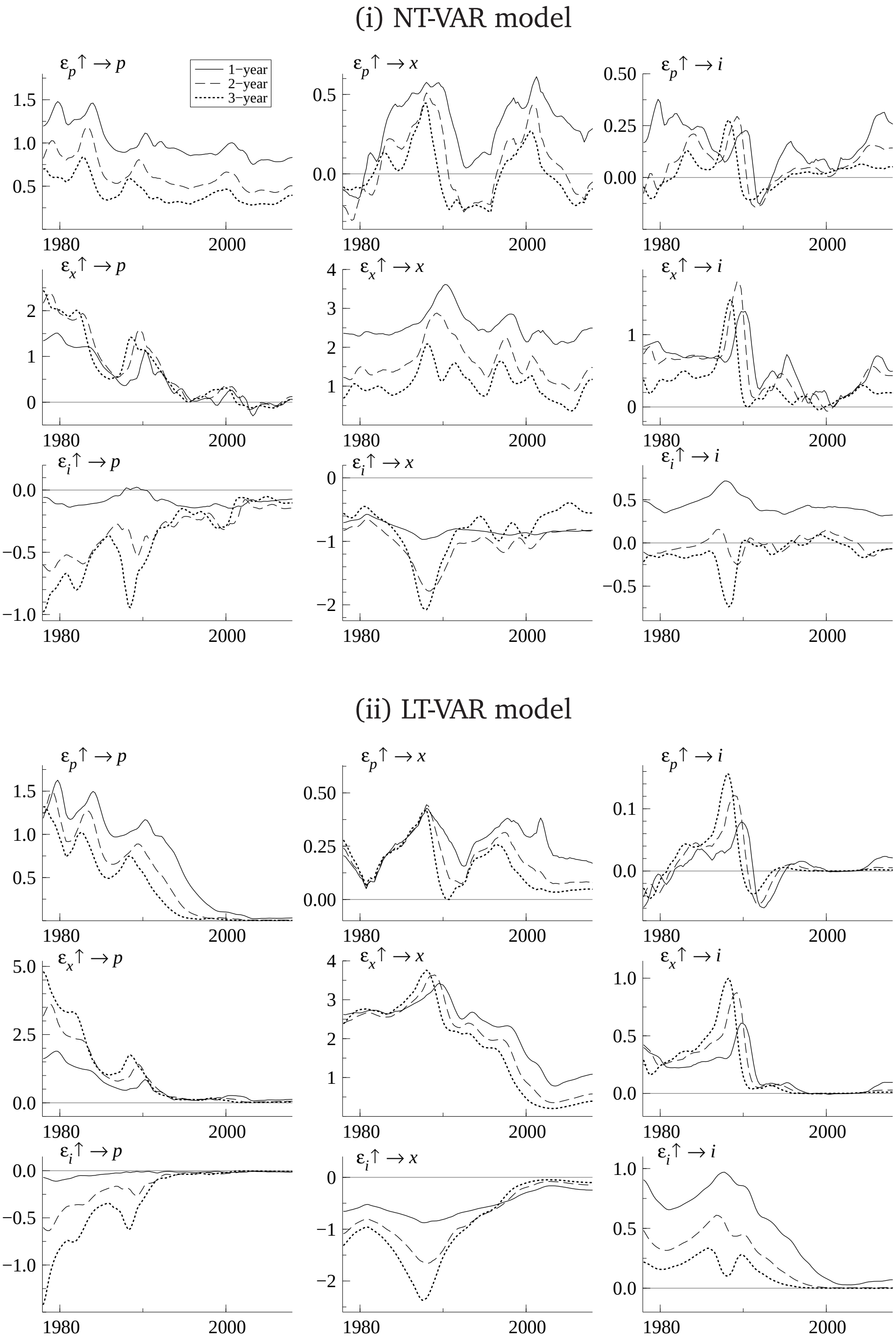


Figure 14: Impulse response trajectories for one-, two- and three-year ahead horizons from the VAR model (upper) and LT-VAR model (lower) for Japanese macroeconomic data. The symbols $\varepsilon_a\uparrow \to b$ refer to the response of the variable b to a shock to the innovation of variable a. The shock size is set equal to the average of the stochastic volatility across time for each series.

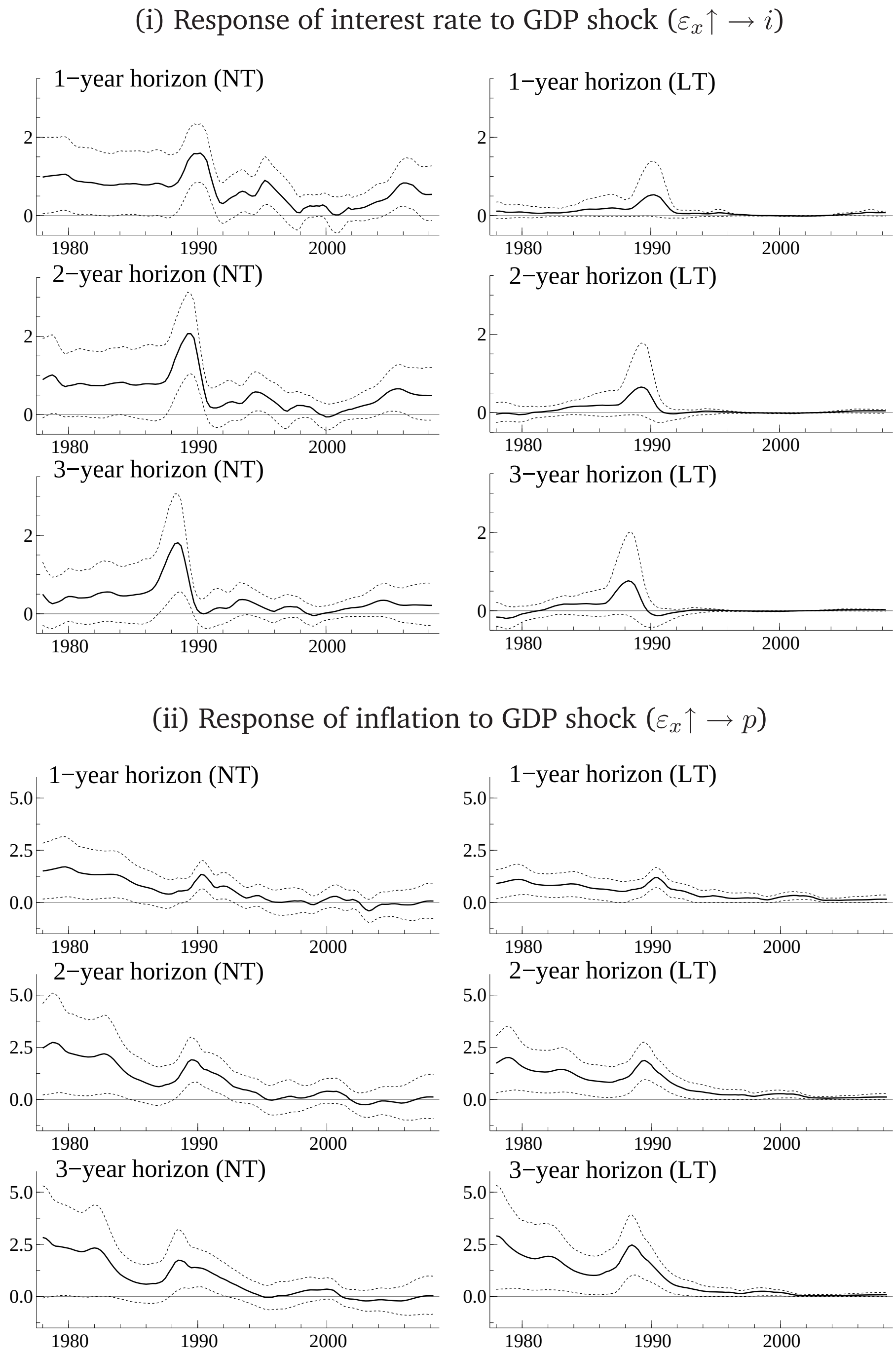


Figure 15: Impulse response trajectories with credible intervals from the NT-VAR (left) and LT-VAR (right) models for Japanese macroeconomic data. Posterior median (solid) and 75% credible intervals (dotted).