# Locally Time-Varying Parameter Regression

Zhongfang He*

First Version: June 11, 2021

**Abstract**

I discuss a framework to allow dynamic sparsity in time-varying parameter regression models. The conditional variances of the innovations of time-varying parameters are time varying and equal to zero adaptively via thresholding. The resulting model allows the dynamics of the time-varying parameters to mix over different frequencies of parameter changes in a data driven way and permits great flexibility while achieving model parsimony. A convenient strategy is discussed to infer if each coefficient is static or dynamic and, if dynamic, how frequent the parameter change is. An efficiency MCMC scheme is described for model estimation. Performance of the proposed approach is illustrated in studies of both simulated and real economic data.

*Keywords:* TVP, Bayesian Shrinkage, MCMC, Economic Time Series
*JEL Codes:* C11, C22, C25, E37

*Email: hezhongfang2004@yahoo.com. Royal Bank of Canada, 155 Wellington St W, Toronto, ON, Canada, M5V 3H6. The views in this paper are solely the author's responsibility and are not related to the company the author works in.

# 1 Introduction

Regression analyses of economic time series have often found that model parameters may be unstable over time (e.g. Stock and Watson (1996), Cogley and Sargent (2005), Primiceri (2005), Dangl and Halling (2012), Belmonte et al. (2014)). Properly accounting for parameter shifts is important for improved model inference and forecast (Clements and Hendry (1999)). However there are several fundamental issues that face the researchers when tackling the varying parameter problem. In a regression model with multiple regressors, it is possible that only some of the regression parameters are time varying while the others are actually constant. In this case, how can one separate these two types of parameters? For parameters that vary over time, what kind of time variation dynamics should be specified? Continuous movements or discrete shifts? An ideal model framework should be flexible to impose minimal *a priori* restrictions on the time variation patterns of model parameters while keeping parsimonious to avoid the risk of over-fitting.

Recent Bayesian studies have applied shrinkage methods to automatically differentiate constant and time varying parameters but often require the researcher to pre-specify if the time variation in parameters is continuous and smooth (time-varying parameter or TVP approach) or infrequent and step-wise (change-point approach). Examples include Fruhwirth-Schnatter and Wagner (2010), Bitto and Fruhwirth-Schnatter (2019) and Dufays and Rombouts (2020) etc. This paper contributes to the literature by proposing an approach that applies dynamic shrinkage and is flexible to accommodate both continuous and infrequent movements in regression coefficients. The resulting model is able to mix over continuous and infrequent parameter shifts in a data-driven way through posterior exploration.

The main thrust of the proposed model is introducing a thresholding mechanism to the conditional variances of the time-varying parameter innovations in the TVP framework.

The conditional variances are allowed to be time varying and are set to zero locally if they are below fixed thresholds at each time point $t$. Thus parameter innovations can be switched on and off at each time point $t$ to control local parameter shift. As a result, both continuous and infrequent parameter shifts as well as their mixtures can be easily accommodated. In particular, the parameters can be locally constant to shut off unnecessary "small" movements that are often caused by fitting random noises in the data. The thresholds act to induce a spike-and-slab-like distribution for the parameter innovations. With higher thresholds, the frequency of zero conditional variances becomes higher, which causes the dynamics of the time-varying parameters to move from continuous smooth movements towards infrequent step-wise ones. By estimating the thresholds through their posterior distributions, the dynamics of the time-varying parameters can mix over different frequencies of parameter shifts, offering rich information to infer the type of parameter dynamics and relieving the researchers of the difficult task to pre-specify it.

The proposed model is close to the mixture innovation approach of Giordani and Kohn (2008) where the TVP framework is adopted but the innovation of each time-varying parameter follows a spike-and-slab distribution. The time-varying mixture indicators in the spike-and-slab distributions control the activation of local parameter shifts and play the same role as the thresholding function in the proposed model. The main difference is computational. As will be explained in Section 2.2, the mixture innovation approach requires $\mathcal{O}(2^K n)$ operations to sample the mixture indicators for a regression model with $K$ regressors and $n$ observations and thus the computation cost grows exponentially with $K$. In contrast, the proposed approach uses latent continuous variables to activate the thresholding function and requires only $\mathcal{O}(n)$ computation cost. As a result, the proposed approach can be applied to estimate larger and more practical TVP regression models.

The proposed model is illustrated in a few examples. In a simulation study, I show that the model is able to recover the true regression coefficients with various time variation

patterns. Moreover, by analysing the posterior distribution of some static model parameters, one can correctly identify if a regression coefficient is constant, infrequently moving or constantly changing. Two empirical applications to macroeconomic and financial data show that the proposed model offers rich information about the dynamics of regression coefficients as well as being competitive in forecasts against some popular existing TVP models.

In the remainder of the paper, Section 2 describes the proposed model in detail. The estimation algorithm is provided in Section 3. The simulation study and the empirical applications are provided in Section 4. Section 5 concludes. Additional details are provided in the appendices.

# 2 The Model

Consider the TVP model with random-walk evolving coefficients and spike-and-slab process variances:

$$\begin{aligned} y_t &= \sum_{j=1}^{K} x_{j,t}\beta_{j,t} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ \Delta\beta_{j,t} &\equiv \beta_{j,t} - \beta_{j,t-1} \sim N\left(0, w_j \mathrm{I}\{z_{j,t} > d_j\}\right) \end{aligned} \tag{1}$$

where $y_t$ is the scalar dependent variable that, conditional on the regressors $x_{1,t}$, ..., $x_{K,t}$ and time-varying coefficients $\beta_{1,t}$, ..., $\beta_{K,t}$, follows a Gaussian distribution with time-varying variance $\sigma_t^2$ over $t = 1$, ..., $n$. Each coefficient $\beta_{j,t}$ follows a random-walk process with a starting value $\beta_{j,0}$. The symbol $\mathrm{I}\{\cdot\}$ denotes an indicator function. A unique feature of Equation (1) is that the process variance of each $\beta_{j,t}$ falls to zero if the latent variable $z_{j,t}$ is below a threshold $d_j$. Thus $\beta_{j,t}$ can stay locally constant to avoid unnecessary movements due to random noises in the data and alleviate the risk of over-fitting.

The thresholding mechanism in Equation (1) allows very flexible dynamics for each

coefficient $\beta_{j,t}$. If $z_{j,t}$ crosses its threshold only occasionally, $\beta_{j,t}$ will behave like in a change-point model with infrequent jumps. On the other hand, when $z_{j,t}$ constantly exceeds its threshold, the resulting dynamics of $\beta_{j,t}$ becomes similar to the conventional TVP model and allows continuous and smooth movements. Between these two types of scenarios lies a continuum of their mixtures by letting $z_{j,t} \leq d_i$ adaptively over $t = 1, ..., n$.

A convenient approach to model the latent variable $z_{j,t}$ is an AR process of order 1:

$$z_{j,t} = \rho_j z_{j,t-1} + \xi_{j,t}, \quad \xi_{j,t} \sim N(0,1) \tag{2}$$

where $|\rho_j| < 1$ and the starting value is $z_{j,1} \sim N(0, 1/(1-\rho_j^2))$. Note that the mean of $z_{j,t}$ is set to 0 while the conditional variance of $z_{j,t}$ is set to 1 in order to identify the threshold $d_j$. The resulting TVP model consisting of Equations (1) and (2) allows regression coefficients to change at a subset of time points and hence is labeled as a locally time-varying parameter or LTVP model.

To facilitate model inference, the threshold $d_j$ in the LTVP model can be reparametrized as $d_j = \Phi^{-1}(q_j)/\sqrt{1-\rho_j^2}$, where $q_j \equiv P(\mathrm{I}\{z_{j,t} > d_j\} = 0)$ is the stationary probability of zero process variance and the function $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal distribution. The relation between $d_j$ and $q_j$ can be derived by observing that the stationary distribution of the latent variable $z_{j,t}$ is $N(0, 1/(1-\rho_j^2))$. It follows that $q_j = P(z_{j,t} \leq d_j) = \Phi\left(\sqrt{1-\rho_j^2}\, d_j\right)$ where $\Phi(\cdot)$ is the CDF of a standard normal distribution. Being the stationary probability of $\mathrm{I}\{z_{j,t} > d_j\} = 0$, the parameter $q_j$ determines the degree of mixing between infrequent jumps ($q_j \to 1$) and continuous movements ($q_j \to 0$) for the dynamics of the corresponding coefficient $\beta_{j,t}$.

Such a reparameterization suggests an interesting strategy to infer the type of the dynamics of $\beta_{j,t}$ by examining the posterior distribution of static model parameters. The slab variance parameter $w_j$ serves as a global factor for $\beta_{j,t}$. As $w_j \to 0$, the conditional variance of $\beta_{j,t}$ becomes zero globally and leads to a constant $\beta_{j,t}$. By examining if $w_j = 0$,

one can thus infer if each $\beta_{j,t}$ is constant or time varying. If $\beta_{j,t}$ is found to be time varying (i.e. $w_j > 0$), one can further examine the posterior distribution of the stationary probability $q_j$ to infer if infrequent jumps or continuous movements of $\beta_{j,t}$ are supported by the data. Such analyses are not substitute for more rigorous tests of model specification but reveal useful and important information to the researcher in a convenient way.

In my experiments, allowing individualized AR coefficient $\rho_j$ for each latent variable $z_{j,t}$ appears to be over-parametrized and tends to produce numerically unstable estimates of the stationary probability $q_j$. Imposing the commonality constraint $\rho_j = \rho$ for $j = 1, ..., K$ stabilizes the estimate of $q_j$.

As for the conditional variance of the dependent variable, the stochastic volatility (SV) process is specified to capture volatility clustering typically observed in economic time series:

$$s_t \equiv \log(\sigma_t^2) = (1-\rho_s)\mu_s + \rho_s s_{t-1} + \epsilon_{s,t}, \;\; \epsilon_{s,t} \sim N(0, \sigma_s^2), \tag{3}$$

where $s_1 \sim N(\mu_s, \sigma_s^2/(1-\rho_s^2))$.

## 2.1 Prior Specification

For the slab variance $w_j$, it is of most interest to examine if it is zero, which is unfortunately a value at the boundary of its support. In this paper, I adopt the strategy in Fruhwirth-Schnatter and Wagner (2010) to instead model the signed standard deviation (SD) $v_j = \pm\sqrt{w_j}$ through the following reparameterization of the LTVP model:

$$\begin{aligned} y_t &= \sum_{j=1}^{K} x_{j,t}\beta_{j,0} + \sum_{j=1}^{K} x_{j,t}\beta_{j,t}^* v_j + \epsilon_t, \;\; \epsilon_t \sim N(0, \sigma_t^2), \\ \Delta\beta_{j,t}^* &\sim N\left(0, \omega_{j,t}\right) \end{aligned} \tag{4}$$

where $\omega_{j,t} \equiv \mathrm{I}\left\{z_{j,t} > \frac{\Phi^{-1}(q_j)}{\sqrt{1-\rho^2}}\right\}$, the starting value $\beta_{j,0}^* = 0$ and the normalized coefficient $\beta_{j,t}^* = \frac{\beta_{j,t}-\beta_{j,0}}{v_j}$. Since the signed SD $v_j$ is the coefficient of a linear regression conditional on

$\beta^*_{j,t}$, a zero-mean Gaussian prior is placed as $v_j \sim N(0, \tau_v)$, where the prior variance $\tau_v$ is a fixed hyper-parameter. By examining if the posterior of $v_j$ is uni-modal around zero or bimodal at non-zero values, one can infer if $w_j = 0$ is supported by the data.

The prior for the stationary probability $q_j$ is a standard uniform distribution $U(0, 1)$ to allow the posterior to mix freely over different frequencies of parameter changes. As for the AR coefficient $\rho$, the prior is $\frac{1+\rho}{2} \sim Beta(\tau_{\rho,1}, \tau_{\rho,2})$, where $\tau_{\rho,1}$ and $\tau_{\rho,2}$ are fixed hyper-parameters.

The initial regression coefficient $\beta_{j,0}$ plays the role of a static regression coefficient as seen in Equation (4). To encourage model parsimony, the horseshoe prior of Carvalho et al. (2010) is applied: $\beta_{j,0} \sim N(0, \tau_0\tau_j)$ with $\sqrt{\tau_0} \sim C^+(0, 1/n)$ and $\sqrt{\tau_j} \sim C^+(0, 1)$, where $C^+$ denotes the half Cauchy distribution over positive values. The global component $\tau_0$ is scaled following Piironen and Vehtari (2017).

In the applications, the hyper-parameters are set as $\tau_v = 0.01$, $\tau_{\rho,1} = 20$, $\tau_{\rho,1} = 1.5$. A relatively tight prior for $v_j$ is found to benefit producing sensible results and reasonable mixing behavior of posterior draws, while a tight prior close to one for the AR coefficient $\rho$ favors strong persistence in the process variances.

For the conditional variance $\sigma^2_t$ of the dependent variable (Equation (3)), the priors are described in Appendix A along with the description of estimating the SV model.

## 2.2 Related Literature

Using the notations in this paper, a typical form of the mixture innovation model of Giordani and Kohn (2008) would specify the dynamics of the time-varying regression coefficients as $\Delta\beta_{j,t} \sim N(0, w_j c_{j,t})$ where $c_{j,t} \in \{0, 1\}$ is the binary mixture indicator and $w_j$ is the variance of the slab part for $j = 1, ..., K$ and $t = 1, ..., n$. The proposed LTVP model resembles the mixture innovation model but uses indicator functions of latent continuous variables to control local parameter shifts as opposed to directly using Bernoulli variables

in the mixture innovation model. As such, large computational gain can be achieved in the LTVP model. A $K$-dimensional mixture indicator $c_t = [c_{1,t} \ ... \ c_{K,t}]'$ has $2^K$ combinatorial scenarios. To sample the elements in $c_t$ jointly, the model likelihood needs to be evaluated over all $2^K$ scenarios which would incur impractically high computation cost when $K$ grows. A single-move Gibbs sampler to cycle through individual elements in $c_t$ reduces computation cost but suffers from high inefficiency due to the interdependence between elements within $c_t$. In contrast, sampling the latent continuous variables $z_t = [z_{1,t} \ ... \ z_{K,t}]'$ in the indicator functions of the LTVP model (Equation (1)) needs to evaluate the model likelihood only twice in Metropolis-Hastings (MH) steps and greatly reduces the computation burden. The details of sampling the indicator function are described in Section 3.

The proposed model fits into the recent literature on Bayesian dynamic shrinkage priors for TVP regressions that, being dynamic, allow adaptive shrinkage at different time points for regression coefficients. This is in contrast to the static shrinkage approaches for TVP regressions where the entire path of each time varying coefficient is regularized towards a constant value and does not allow local behaviors such as infrequent shifts or temporary zero values.[1] Besides the mixture innovation model, other examples in the dynamic TVP shrinkage literature include Kalli and Griffin (2014), Kowal et al. (2019), Uribe and Lopes (2020) and Rockova and McAlinn (2021) etc.[2] A common feature of these approaches is allowing unnecessary parameter changes to be approximately zero locally but is unable to enforce exact local constancy in parameters as in the mixture innovation model and the proposed LTVP model. As such, one can not infer the type of the dynamics of the time-varying coefficient by these dynamic TVP shrinkage models. Within this strand of

[1]Examples include Fruhwirth-Schnatter and Wagner (2010), Belmonte et al. (2014), Bitto and Fruhwirth-Schnatter (2019), Cadonna et al. (2020), Hauzenberger et al. (2020) etc.

[2]Another example is Huber et al. (2019) that applies a thresholding approximation in MCMC draws of the mixture innovation model to speed up computation but suffers from convergence issues (Dufays et al. (2021)).

literature, the dynamic horseshoe model of Kowal et al. (2019) is particularly influential due to its intuitive design and good performance.[3] In the empirical sections, I will compare the forecast performance of the proposed approach with the dynamic horseshoe model.

The use of the indicator function as a truncation device to model time-varying regression parameters has been explored in many previous studies (e.g. Tong (1990), Chang et al. (2017)) but typically works to truncate the level of regression coefficients. In particular, Nakajima and West (2013) applies the indicator function to truncate time-varying regression coefficients to zero locally when the absolute values of these coefficients are below certain thresholds, which enables temporary zero values of the coefficients but does not allow non-zero ones. In contrast, the indicator function in the proposed model operates on the innovation of regression coefficients and thus is flexible to accommodate a much wider range of possible time variation patterns in the coefficients.

# 3 MCMC Estimation

For notational ease, let the dependent variable $y = \{y_t\}_{t=1}^n$, the normalized regression coefficients $\beta^* = \{\beta_t^*\}_{t=1}^n$ with $\beta_t^* = [\beta_{1,t}^* \ ... \ \beta_{K,t}^*]'$, the latent variable $z = \{z_t\}_{t=1}^n$ with $z_t = [z_{1,t} \ ... \ z_{K,t}]'$, $v = \{v_j\}_{j=1}^K$, $q = \{q_j\}_{j=1}^K$ and $\beta_0 = \{\beta_{j,0}\}_{j=1}^K$. Denote $\theta_v = \{v, \beta_0\}$, $\theta_q = \{q, \rho\}$, $\theta_0 = \{\tau_0, \tau_1, ..., \tau_K\}$ collecting the hyper-parameters for $\beta_0$, $\theta_s$ collecting $\{\sigma_t^2\}_{t=1}^n$ and the other parameters in the SV model, and finally $\Theta = \{\theta_v, \theta_q, \theta_0, \theta_s\}$ collecting all static model parameters. I will use $S_{-j}$ to denote the elements of a generic set $S$ excluding its subset $j$. For simplicity, the regressors $\{x_{j,t}\}_{j=1,...,K,t=1,...,n}$ will be dropped from the conditioning set of posterior distributions below.

The target is to sample from the posterior distribution $p(\beta^*, z, \Theta|y)$ based on the repre-

[3]See Kowal et al. (2019) for documented superior performance of the dynamic horseshoe model in studies comparing with other dynamic shrinkage models such as the dynamic normal-gamma approach of Kalli and Griffin (2014).

sentation of Equation (4). A Gibbs sampler is applied that divides the model parameters into 4 blocks $\theta_s$, $\theta_0$, $\theta_v$ and $\{\beta, z^*, \theta_q\}$ and is outlined below.

Standard approaches are available to sample the blocks $\theta_s$, $\theta_0$ and $\theta_v$. For parameters $\theta_s$ in the SV model, the sampler of Kastner and Fruhwirth-Schnatter (2014) is applied with the details in Appendix A. Conditional on $\beta_0$, the posterior of the hyper-parameters $\theta_0$ is inverse gamma distributions based on the hierarchical representation approach of Makalic and Schmidt (2016). The details of sampling $\theta_0$ are provided in Appendix B. Conditional on $\{\beta^*, \theta_s, \theta_0\}$, $\theta_v$ is the parameters of a linear regression whose posterior is a normal distribution and is described in Appendix C.

For the remaining block $\{\beta^*, z, \theta_q\}$, a straightforward approach would divide it into sub-blocks of $\beta^*$, $z$ and $\theta_q$ and apply a nested Gibbs sampler. The problem of this approach, however, is that sampling $z$ conditional on $\beta^*$ is highly inefficient and can completely break down due to the high correlation between $z$ and $\beta^*$, similar to the difficulty of sampling the mixture indicators in the mixture innovation model that has long been observed in the literature (McCulloch and Tsay (1993), Gerlach et al. (2000)). Similarly $\theta_q$ is also highly dependent on $\beta^*$. For improved sampling efficiency, $\beta^*$ is integrated out when sampling $z$ and $\theta_q$. The posterior is decomposed as:

$$p(\beta^*, z, \theta_q | y, \Theta_{-\theta_q}) = p(z, \theta_q | y, \Theta_{-\theta_q}) p(\beta^* | y, z, \Theta)$$

The part $p(\beta^*|y, z, \Theta)$ can be treated as sampling the latent states in a linear Gaussian state space system and is sampled by the simulation smoother of Durbin and Koopman (2002).[4] On the other hand, sampling the part $p(z, \theta_q|y, \Theta_{-\theta_q})$ is unconventional in that the normalized coefficients $\beta^*$ are integrated out. To sample from this posterior, a nested Gibbs sampler is applied to iterate over the sub-blocks $p(z|y, \Theta)$ and $p(\theta_q|y, z, \Theta_{-\theta_q})$. Of note, sampling $\{z, \theta_q\}$ should precede sampling $\beta^*$ given that the former integrates out $\beta^*$.

[4] Alternative approaches to simulate the latent states from a linear Gaussian state space system include Fruhwirth-Schnatter (1994), Rue (2001) and McCausland et al. (2011) etc.

The sub-block $p(z|y,\Theta)$ is sampled by a single-move MH-within-Gibbs step:

$$p(z_t|y,\Theta,z_{-t}) \propto p(z_t|\Theta,z_{-t})p(y|z,\Theta)$$

over $t = 1, ..., n$. It may appear that one can directly apply the Kalman filter to compute the likelihood part $p(y|z,\Theta)$ by integrating out $\beta^*$. But this approach would require $\mathcal{O}(n^2)$ operations to sample $z$ in each MCMC sweep and is too costly for typical data sample length $n$. Following Gerlach et al. (2000) (*GCK* hereafter), this paper applies the factorization $p(y|z,\Theta) \propto p(y_t, ..., y_n|y_1, ..., y_{t-1}, z, \Theta)$ for sampling $z_t$, where the right-hand side of the factorization can be computed efficiently in one step by the GCK algorithm. The overall cost of sampling $z$ is in the order of $\mathcal{O}(n)$ operations. The details of sampling $z$ are provided in Appendix D.

It is worth noting here that, to sample each $z_t$ and hence computing the conditional variances of $\beta_t^*$ in the LTVP model, the likelihood $p(y|z,\Theta)$ only needs to be evaluated twice via the GCK algorithm in an MH step. This is in contrast to sample the discrete mixture indicator in the conditional variances of $\beta_t$ from the mixture innovation model of Giordani and Kohn (2008) where the model likelihood needs to be evaluated by the GCK algorithm over all $2^K$ combinatorial scenarios of the mixture indicator and is not generally practical.

For the other sub-block $p(\theta_q|y,z,\Theta_{-\theta_q})$, a nested MH-within-Gibbs sampler is applied to draw $q$ and $\rho$ separately. Details are described in Appendix E. The critical part of the posterior distributions of $q$ and $\rho$ is computing the marginalized likelihood $p(y|z,\Theta)$ that can be derived by applying the Kalman filter to integrate out $\beta^*$.

MH steps are heavily used in the sampler. To avoid manual tuning, the adaptive optimal scaling method of Garthwaite et al. (2016) is adopted to automatically tune the random walk proposals of MH steps towards the acceptance rate of 25% for each $z_t$ and $q^*$ and 44% for the univariate $\rho^*$. The details are provided in Appendix F.

In my experiments, posterior draws from the above Gibbs sampler could mix unsatisfactorily in certain situations. For example, when a regression coefficient contains infrequent change points as in the simulation example, draws of the corresponding parameter $v$ and $q$ could mix poorly. To robustify the sampler's performance, I adopt the *ancillarity-sufficiency interweaving strategy* (ASIS) of Yu and Meng (2011) that adds two extra steps in each MCMC sweep to re-sample the parameters $\{v, \beta_0\}$ and $q$. The re-sampling in the ASIS steps is from closed-form distributions and hence requires minimum extra computation cost (a few seconds per 1,000 draws in the examples of this paper) while noticeably improving the mixing of posterior draws. The details of the ASIS boosting can be found in Appendix G.

# 4 Illustrations

## 4.1 Synthetic Data

The proposed model is illustrated in a simulated data sample with $n = 300$ and $K = 6$ where the true regression coefficients exhibit different time variation patterns. The first 3 coefficients are as follows:

1. Random walk: $\beta_{1,t} = \sum_{j=1}^{t} u_j$ with $u_j \sim N(0, 0.01)$.

2. Two change points: $\beta_{2,t} = \mathrm{I}_{\{t_1 < t \leq t_2\}} - 0.5\, \mathrm{I}_{\{t > t_2\}}$.

3. One change point: $\beta_{3,t} = \mathrm{I}_{\{t > t_3\}}$.

where $t_1 = 100$, $t_2 = 200$ and $t_3 = 150$. Coefficient 4 is a constant of one and the other coefficients are all zeros. The regressors are drawn from standard normal distributions. The dependent variable is generated by adding a noise from $N(0, \sigma_t^2)$ where $\log(\sigma_t^2) = -0.1 + 0.9 \log(\sigma_{t-1}^2) + N(0, 0.01)$ and $\log(\sigma_1^2) = -1$.

Posterior analysis is based on 15,000 draws by discarding 5,000 burn-ins.[5] The posterior draws mix reasonably well. For example, the point-wise inefficiency factor (*IF* hereafter) of $\beta_t$ is all below 100.[6] The acceptance rates of the MH steps are between 23% and 26% for $z$ and $q$ and 46% for $\rho$ and are close to their target values.

Figure 1 plots the point-wise posterior median and 90% credible set of the regression coefficient $\beta_t$ along with the true coefficient value. It can be seen that the true regression coefficients are well covered by the posterior estimates despite their heterogeneous shapes. As a further illustration, Figure 2 shows two posterior draws of the 1-change-point coefficient. Both draws track the shape of the true coefficient. However one draw shows frequent yet small movements while the other draw shows mostly sparse parameter changes. This example provides a glimpse of how the posteriors of the coefficients mix over different types of dynamics.

The posterior of each $v_j$ directly from the MCMC sampler does not necessarily have a symmetric and clear-cut bimodal shape when it supports non-zero slab variance $w_j$ for $j = 1, ..., K$. To see the pattern more easily, the shape of $v^* = \{v_j^*\}_{j=1}^K$ with $v_j^* = \pm\sqrt{v_j^2}$ is examined where the sign of $\sqrt{v_j^2}$ is randomly drawn. Figure 3 shows the shape of $v^*$. It is clear from the figure that posterior draws of $v^*$ for time-varying coefficients are all of bimodal shape, while those for constant coefficients are unimodal and tightly concentrate around 0. Focusing on the time-varying coefficients, the left panels in Figure 4 show the posterior distribution of the stationary probabilities $q$ for these coefficients. Even though the prior of $q$ is uniform between 0 and 1, the posterior of $q$ for the random-walk coefficient is mostly located in regions away from 1 which correctly supports frequent movements in the coefficient. On the other hand, the posterior distributions of $q$ for the

[5]Producing 1,000 posterior draws takes about 110 seconds on a standard desktop computer with a 3.0 GHz Intel Core i5 CPU running in MATLAB R2020b.

[6]The IF is computed by the initial monotone sequence method of Geyer (1992) A smaller IF value implies less correlated and hence better mixed posterior draws.

change point coefficients all concentrate in regions very close to 1, consistent with infrequent movements in these two coefficients. Therefore by examining the posterior distribution of the parameters $v^*$ and $q$, one could infer whether a coefficient is constant or time varying and, if time varying, whether parameter changes are frequent or infrequent.

In principle, one could use the point-wise posterior mean of $\omega \equiv \{\omega_{j,t}\}_{j=1,\ldots,K,t=1,\ldots,n}$ in Equation (4) to infer the locations of parameter changes. As observed in the right panels of Figure 4, the point-wise posterior mean of $\omega$ is all above 0.5 for the random-walk coefficient and correctly shows spikes around the true break locations for the change-point coefficients. However frequent minor spikes in the posterior mean of $\omega$ due to noises in the data are also evident in the figure and could result in ambiguous signals about the locations of parameter changes in real world applications which would certainly be more complicated than the simulation example.

## 4.2 Industrial Output and the Yield Curve

This exercise is predicting the growth rate of industrial output by the yield curve of interest rates. The literature on this topic is voluminous. See Haubrich (2020) for a comprehensive review on predicting economic output by the yield curve.

The dependent variable in the study is the quarterly log growth rate of the U.S. industrial production index. There are 7 regressors including a constant, AR 1 to 4 lags of the log growth rate, 1-quarter lag of the quarterly change of the short rate (3-month treasury bill rate) and 1-quarter lag of the term spread (difference between the 10-year treasury constant maturity rate and the 3-month treasury bill rate). Quarterly values of interest rates are calculated as the average monthly values within each quarter. The data sample has a long span from Q3 1953 to Q1 2021 with 271 quarterly observations. All data are from the FRED database of the U.S. Federal Reserve Bank of St. Louis.[7]

[7]The series names are *INDPRO*, *TB3MS* and *GS10*.

In estimation, all non-constant regressors are normalized by subtracting their sample means and dividing by their sample standard deviations. A total of 15,000 posterior draws is collected after a burn-in of 5,000. Figure 5 shows the point-wise posterior median and 90% credible set of the regression coefficient $\beta_t$. Inspecting Figure 5, it is clear that determining if a coefficient is constant or time varying is not obvious by eyeballing the estimate of $\beta_t$, let alone inferring the type of time variation dynamics. In most cases, both constants of different values and time variation patterns of different types can fit into the estimated credible set. This is why the posterior of some key static model parameters becomes useful for more informed model identification.

To identify the time varying coefficients, Figure 6 plots the posterior densities of $v^*$ where bi-modality is evident for the intercept while a tiny amount of bi-modality can be found for the AR 4 lag and the term spread, suggesting that time variation in the coefficients of these three regressors is supported to various degrees by the data. Examining the estimate of $\beta_t$ in Figure 5, the intercept steadily declines throughout the data sample from 1953 to 2021. The AR 4 coefficient shifts to a higher level around mid 1960s and stays around that level until mid 2000s before moving down. The coefficient of the term spread is of most empirical interest and is found to steadily decline till early 1980s and, since then, is flat and mostly insignificant by consistently containing the value of zero in its point-wise 90% credible set, suggesting that the predictive power of the term spread deteriorates since 1980s.[8]

To further infer the type of dynamics for the time varying coefficients, Figure 7 plots the posterior densities of $q$ for the 3 regressors including the intercept, the AR 4 lag and the term spread. It is found that the posterior densities of $q$ for all these 3 regressors are mostly located in regions close to one and hence favor infrequent changes in the corresponding

[8]Such a finding is consistent with previous studies such as Estrella et al. (2003).

coefficients as opposed to continuous TVP-type dynamics.[9] Among these 3 regressors, the posterior density of $q$ for the intercept is more dispersed than the others, suggesting more uncertainty in the frequency of shifts in the intercept.

## 4.3 Equity Premium

The second empirical exercise predicts the U.S. equity premium one quarter ahead with the predictors examined in Welch and Goyal (2008). The dependent variable is the value-weighted quarterly return of the S&P500 index minus the corresponding risk free rate. A total of 12 economic predictors are selected based on Welch and Goyal (2008) that include stock characteristics, interest rates and other macroeconomic indicators and are listed in Table 1. Detailed descriptions of the predictors can be found in the original paper of Welch and Goyal (2008). In estimation, I also include an intercept and an AR 1 lag of the equity premium. The sample runs from Q1 1947 to Q4 2020 with a total of 296 quarterly observations. The data is publicly available from Amit Goyal's website.[10] Similar to the example of industrial production, all non-constant regressors are normalized in estimation. Analysis is based on 15,000 posterior draws after 5,000 burn-ins.

As the number of regressors grows, visual inspection of the shapes of the posterior densities of $v^*$ becomes increasingly inconvenient. Moreover the degree of bi-modality in the posterior densities of some bimodal elements of $v^*$ could be very small. Recall from Equation 4, the latent variable $z_t$ and parameters such as the stationary probabilities $q$ are not identifiable if the corresponding parameters $v$ equal zero (i.e. constant coefficients). Therefore, when an element of $v^*$ is only weakly bimodal, the estimate of the corresponding $z_t$ and $q$ could be numerically unstable. As such I focus on presenting the results for coefficients that are "strongly" time varying based on the magnitude of the mode of $|v|$. Table 2

[9] Additional results of the point-wise posterior mean of $\omega$ are provided in Appendix H.

[10] The web address is https://sites.google.com/view/agoyal145/?redirpath=/.

shows the 6 regressors with non-zero mode of $|v|$ where the mode of $|v|$ is about 0.02 for the intercept while being an order of magnitude smaller for the 5 other regressors. Thus only the result for the intercept is described in the main text.[11] Of note, the coefficients of some popular predictors of equity premium such as the dividend payout ratio are not supported to be time varying as the posterior densities of $v^*$ for these regressors are unimodal.

Figure 8 shows the estimates for the intercept. The first panel shows the point-wise posterior median and 90% credible set of $\beta_t$ for the intercept. The intercept grows till mid 1960s and then declines in the 1970s before resuming growth in the 1980s and 1990s. Since the early 2000s, the intercept largely stays flat. The posterior density of the stationary probability $q$ for the intercept is shown in the second panel of Figure 8 and, as seen in the figure, is rather dispersed, though the bulk of it is in regions away from the end of one, suggesting that frequent shifts in the intercept are supported by the data. Such a pattern of the posterior density of $q$ is consistent with the point-wise posterior mean of $\omega$ for the intercept which is shown in the third panel of Figure 8 and is above 0.5 at all time points, supporting continuous movements in the intercept.

## 4.4 Comparing with Existing TVP Models

The proposed LTVP model is compared to a number of popular existing TVP models in terms of predictive performance. The alternative TVP models include a special version of the triple gamma prior of Cadonna et al. (2020), a restricted version of the mixture innovation model of Giordani and Kohn (2008), the dynamic horseshoe prior of Kowal et al. (2019) and the latent threshold model of Nakajima and West (2013). The triple gamma prior is a static shrinkage model while the 3 other models apply dynamic shrinkage to time-varying coefficients. A brief description of the 4 alternative models is provided in Appendix J.

[11] The results for the full set of regressors are provided in Appendix I.

The predictive performance of the various TVP models is compared via the predictive likelihood that integrates out all model parameters and has been a standard Bayesian approach for comparing competing models' forecasts (Geweke and Whiteman (2006)). Similar to Bitto and Fruhwirth-Schnatter (2019), the Kalman filter is applied to improve numerical stability when computing the predictive likelihood of the TVP models except for the latent threshold one of Nakajima and West (2013).

The prediction sample is from Q2 2011 to Q1 2021 for the industrial output exercise and is from Q1 2011 to Q4 2020 for the equity premium exercise. Table 3 shows the cumulative log predictive likelihoods produced by the TVP models under study. The proposed LTVP model and the triple gamma prior are always among the best performing models in the two prediction exercises and produce comparable cumulative predictive likelihoods. Thus the greater complexity of the LTVP model does not compromise its forecasting performance relative to the more parsimonious static shrinkage model of the triple gamma prior. The restricted mixture innovation model has comparable performance as the LTVP and the triple gamma models in the industrial output exercise but is the worst performing model in the equity premium exercise, suggesting the importance of selecting the "right" scenarios for the mixture indicators which unfortunately is not obvious in practice. The dynamic horseshoe and the latent threshold models lag behind the LTVP model, showing the benefit of sparsity (relative to the dynamic horseshoe model) and greater flexibility (relative to the latent threshold model) in the LTVP model.

# 5 Conclusion

A new approach of TVP regressions is proposed that applies a thresholding mechanism to switch on and off parameter shifts locally at each time point $t$. Such a feature allows the dynamics of the time-varying coefficients to be very flexible and to mix over different

frequencies of parameter shifts through posterior simulation. I also discuss a practically convenient strategy to infer the type of the dynamics of each regression coefficient by checking the posterior distribution of static model parameters. A computationally efficient Bayesian algorithm is provided. Examples based on both simulated and real economic data show encouraging results by the proposed approach.

An unresolved issue is inferring the locations of parameter shifts for which the estimate by the proposed model appears sensitive to noises in the data. Investigating dynamic shrinkage priors in the TVP framework that can reliably identify the locations of parameter shifts would be an interesting future research topic.

Table 1: List of Predictors for Equity Premium

| Name | Description |
|---|---|
| Dividend price ratio | Log dividends minus log price |
| Dividend payout ratio | Log dividends minus log earnings |
| Stock variance | Log of sum of squared daily returns on the S&P500 index |
| Book-market ratio | Ratio of book to market value for the Dow Jones Industrial Average index |
| Equity expansion | Ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks |
| Treasury bill rate | Quarterly change of 3-month secondary market Treasury bill rate |
| Long term yield | Quarterly change of long-term government bond yield from Ibbotson's *Stocks, Bonds, Bills and Inflation Yearbook* |
| Term spread | Long term yield minus treasury bill rate |
| Default yield spread | Difference between BAA and AAA-rated corporate bond yields |
| Default return spread | Difference between long-term corporate and government bond returns |
| Inflation rate | Consumer price index (all urban consumers) |
| Investment-capital ratio | Ratio of aggregate (private non-residential fixed) investment to aggregate capital |

Note: The data is publicly available from Amit Goyal's website https://sites.google.com/view/agoyal145/?redirpath=/. Detailed descriptions of the variables can be found in Welch and Goyal (2008).

Table 2: List of Regressors with Non-Zero Mode of $|v|$: Equity Premium

| | Mode of $\|v\|$ |
|---|---|
| Intercept | 0.0186 |
| Term Spread | 0.0045 |
| Equity Expansion | 0.0036 |
| Default Yield Spread | 0.0029 |
| Default Return Spread | 0.0026 |
| Dividend Price Ratio | 0.0020 |

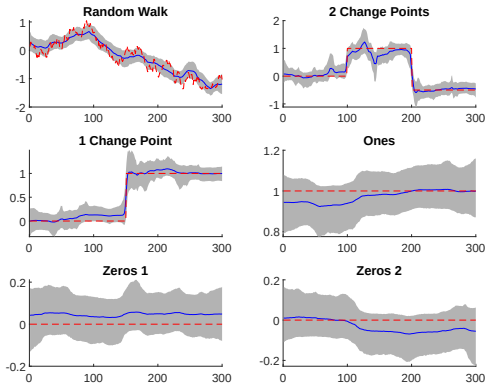
Note: The description of the regressors can be found in Table 1.

Table 3: Comparing Cumulative Log Predictive Likelihoods

| | Industrial Production | Equity Premium |
|---|---|---|
| LTVP | -21.82 | 49.80 |
| Triple Gamma | -21.86 | 50.51 |
| Restricted Mixture Innovation | -21.49 | 29.67 |
| Dynamic Horseshoe | -23.79 | 35.78 |
| Latent Threshold | -23.87 | 44.56 |

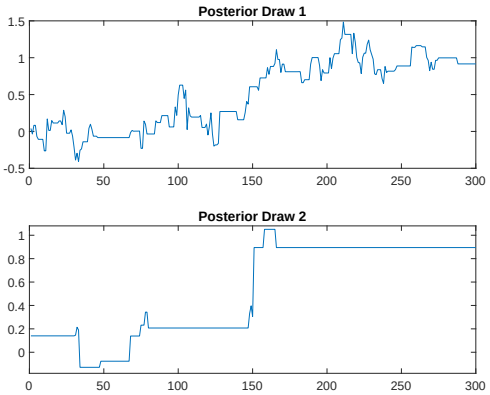
Note: The cumulative log predictive likelihood is the cumulative sum of log predictive likelihoods produced by the competing TVP models over the prediction samples for the two exercises of predicting the industrial output and the equity premium.

Figure 1: Estimate of $\beta_t$ in the LTVP Model: Simulation



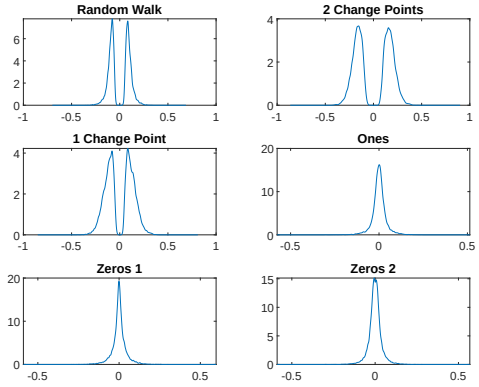

Note: This figure shows the point-wise posterior median (solid line) and 90% credible set (gray area) for the coefficients $\beta_t$ of the LTVP model in the simulation study, where the dashed line is the true coefficient.

Figure 2: Selected Posterior Paths of $\beta_t$ in the LTVP Model: Simulation



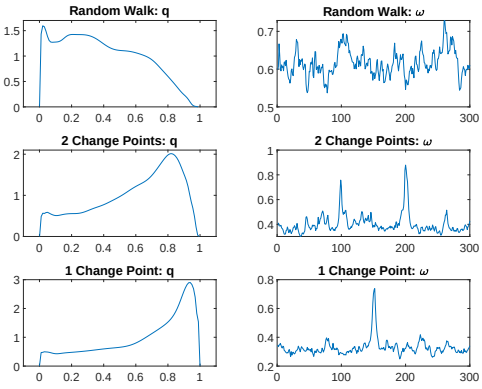

Note: This figure shows two posterior paths of the 1-change-point coefficient $\beta_{3,t}$ of the LTVP model in the simulation study. The true coefficient equals 0 up to the point 150 and equals 1 afterwards.

Figure 3: Posterior of $v^*$ in the LTVP Model: Simulation



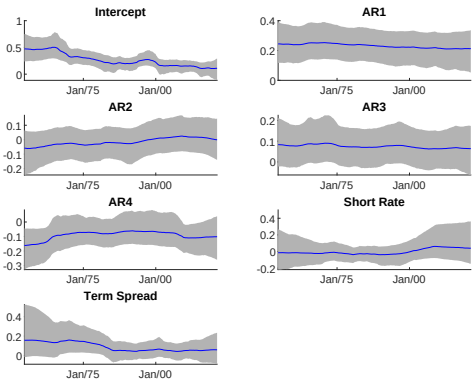

Note: This figure shows the posterior distribution of the parameters $v^*$ for the LTVP model in the simulation study.

Figure 4: Estimate of $q$ and $\omega$ for Time-Varying Coefficients in the LTVP Model: Simulation



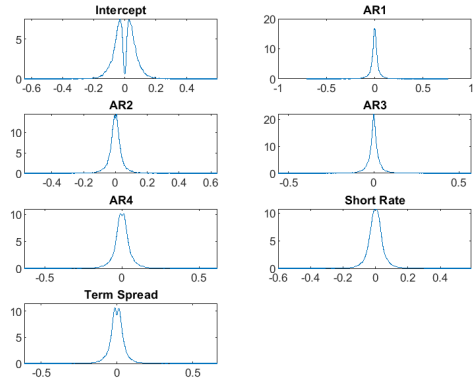

Note: This figure shows the posterior distribution of the parameters $q$ and the point-wise posterior mean of $\omega$ of the time-varying coefficients in the LTVP model for the simulation study.

Figure 5: Estimate of $\beta_t$ in the LTVP Model: Industrial Output



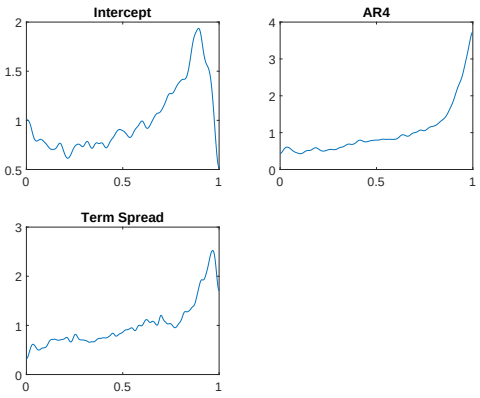

Note: This figure shows the point-wise posterior median (solid line) and 90% credible set (gray area) for the coefficients $\beta_t$ of the LTVP model for predicting the industrial output.

Figure 6: Posterior of $v^*$ in the LTVP Model: Industrial Output



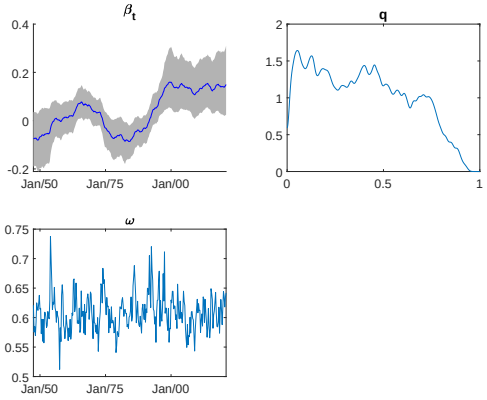

Note: This figure shows the posterior distribution of the parameters $v^*$ in the LTVP model for predicting the industrial output.

Figure 7: Posterior of $q$ for Time-Varying Coefficients in the LTVP Model: Industrial Output




Note: This figure shows the posterior distribution of the parameters $q$ of time-varying coefficients in the LTVP model for predicting the industrial output.

Figure 8: Estimate of $\beta_t$, $q$, $\omega$ for the Intercept in the LTVP Model: Equity Premium




Note: The first panel shows the point-wise posterior median (solid line) and 90% credible set (gray area) of the coefficient $\beta_t$ for the intercept in the LTVP model for predicting the equity premium. The second and third panels show the posterior density of the parameter $q$ and the point-wise posterior mean of $\omega$ respectively for the intercept.

# Appendix

## A Estimating the SV Model

The SV model is specified in Equation (3) of Section 2. The priors for model parameters are $\mu_s \sim N(0, 10)$, $\rho_s \sim N(0.95, 0.04)\mathrm{I}\{-1 < \rho_s < 1\}$, $\sigma_s^2|s_h \sim G(0.5, 2s_h)$ and $\sqrt{s_h} \sim C^+(0, 1)$.

Estimation of the SV model follows Kastner and Fruhwirth-Schnatter (2014) that adopts the log linearization strategy of Omori et al. (2007) to approximate the logarithm of a $\chi^2(1)$-distributed variable by a mixture of normal distributions. Kastner and Fruhwirth-Schnatter (2014) applies the interweaving strategy of Yu and Meng (2011) to boost the sampling efficiency of the long-run mean and the variance parameter of the log volatility process and has become a standard approach for Bayesian studies of the SV model (e.g. Bitto and Fruhwirth-Schnatter (2019), Huber et al. (2021), Huber and Pfarrhofer (2021)). The details of the sampler can be found in Kastner and Fruhwirth-Schnatter (2014) and are not repeated here to save space.

The main difference in this paper from Kastner and Fruhwirth-Schnatter (2014) is the prior of the variance parameter in the log volatility process. Instead of setting a fixed value for the scale parameter $s_h$ in the gamma prior $\sigma_s^2 \sim G(0.5, 2s_h)$, this paper specifies a prior $\sqrt{s_h} \sim C^+(0, 1)$ to determine $s_h$ in a data driven way. The conditional posterior of $s_h$ can be obtained by applying the hierarchical inverse gamma representation in Makalic and Schmidt (2016): $s_h|a_h, \sigma_s^2 \sim IG\left(1, \frac{1}{a_h} + \frac{\sigma_s^2}{2}\right)$ where $a_h$ is an auxiliary variable with the prior $a_h \sim IG(0.5, 1)$ and the posterior $a_h|s_h \sim IG\left(1, 1 + \frac{1}{s_h}\right)$.

## B Horseshoe Hyper-Parameters

The horseshoe prior of the initial regression coefficients is $\beta_{j,0} \sim N(0, \tau_0\tau_j)$ with $\sqrt{\tau_0} \sim C^+(0, 1/n)$ and $\sqrt{\tau_j} \sim C^+(0, 1)$ for $j = 1, ..., K$. Following Makalic and Schmidt (2016),

these distributions can be represented as hierarchical inverse gamma ones by introducing auxiliary variables::

$$\sqrt{\tau_0} \sim C^+(0, 1/n) \Longleftrightarrow \tau_0|\kappa_0 \sim IG\left(0.5, \frac{1}{\kappa_0}\right), \;\; \kappa_0 \sim IG(0.5, n^2)$$
$$\sqrt{\tau_j} \sim C^+(0, 1) \Longleftrightarrow \tau_j|\kappa_j \sim IG\left(0.5, \frac{1}{\kappa_j}\right), \;\; \kappa_j \sim IG(0.5, 1)$$

with the following posteriors:

$$\tau_0|\kappa_0, \beta_0, \tau_1, ..., \tau_K \sim IG\left(\frac{1+K}{2}, \frac{1}{\kappa_0} + \frac{1}{2}\sum_{j=1}^{K}\frac{1}{\tau_j}\beta_{j,0}^2\right),$$
$$\kappa_0|\tau_0 \sim IG\left(1, n^2 + \frac{1}{\tau_0}\right),$$
$$\tau_j|\beta_0, \tau_0, \kappa_j \sim IG\left(1, \frac{1}{\kappa_j} + \frac{1}{2\tau_0}\beta_{j,0}^2\right),$$
$$\kappa_j|\tau_j \sim IG\left(1, 1 + \frac{1}{\tau_j}\right).$$

# C Sampling the Parameters $\beta_0$ and $v$

The target is to sample from the posterior $p(v, \beta_0|y, \beta^*, \theta_s, \theta_0) \propto p(v, \beta_0|\theta_0)p(y|v, \beta_0, \beta^*, \theta_s)$. Let $\gamma = [\beta_0' \; v']'$, $B_{\gamma,0}$ be a $2K$-by-$2K$ diagonal matrix with the diagonal $[\tau_0\tau_1 \; ... \; \tau_0\tau_K \; \tau_v \; ...\tau_v]'$, $\tilde{y}_\gamma = [\frac{y_1}{\sigma_1} \; ... \; \frac{y_n}{\sigma_n}]'$ and $\tilde{x}_\gamma$ be a $n$-by-$2K$ matrix stacking $[\frac{x_t'}{\sigma_t} \;\; \frac{(x_t \odot \beta_t^*)'}{\sigma_t}]$ with $t = 1, ..., n$. It can be derived as a linear regression based on Equation (4) that the posterior of $\gamma$ is a normal distribution $N(b_\gamma, B_\gamma)$ where $B_\gamma^{-1} = B_{\gamma,0}^{-1} + \tilde{x}_\gamma'\tilde{x}_\gamma$ and $B_\gamma^{-1}b_\gamma = \tilde{x}_\gamma'\tilde{y}_\gamma$.

# D Sampling the Latent Variables $z$

The target is to sample from the posterior $p(z_t|y, \Theta, z_{-t}) \propto p(z_t|\Theta, z_{-t})p(y|z, \Theta)$. The notations are defined in the main text.

Given the AR specification of $z_t$ in Equation (2), the prior part can be written as $p(z_t|\Theta, z_{-t}) \propto p(z_t|z_{t-1}, \rho)p(z_{t+1}|z_t, \rho)$ for $t = 2, ..., n-1$. One can derive $z_t|z_{-t}, \rho \sim$

$N(b_{z,t}, B_{z,t})$ where $B_{z,t}$ is a $K$-by-$K$ diagonal matrix with the $j^{\text{th}}$ diagonal entry $\frac{1}{1+\rho^2}$ and $b_{z,t}$ is a $K$-by-1 vector with the $j^{\text{th}}$ entry $\frac{\rho}{1+\rho^2}(z_{j,t-1}+z_{j,t+1})$ for $j = 1, ..., K$ and $t = 2, ..., n-1$. For the terminal $t = 1$, one has $p(z_1|\Theta, z_{-1}) \propto p(z_1|\rho)p(z_2|z_1, \rho)$ where $z_{j,1} \sim N\left(0, \frac{1}{1-\rho^2}\right)$ for $j = 1, ..., K$. It is straightforward to derive $z_1|z_{-1}, \rho \sim N(b_{z,1}, B_{z,1})$ with $B_{z,1} = I_K$ and $b_{z,1} = \rho z_2$. When $t = n$, the posterior is simply $p(z_n|\Theta, z_{-n}) \propto p(z_n|z_{n-1}, \rho)$ which is a normal distribution $N(\rho z_{n-1}, I_K)$.

Denote $y^t = \{y_1, ..., y_t\}$ and $y^{t,n} = \{y_t, ..., y_n\}$. The algorithm of Gerlach et al. (2000) is applied to compute the parts of the likelihood $p(y|z, \Theta)$ that is relevant to sampling $z_t$:

$$\begin{aligned} p(y|z,\Theta) &\propto p(y^{t,n}|y^{t-1}, z, \Theta) \\ &\propto r_t^{-\frac{1}{2}} \det(Q_t)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} m_t' \Omega_t m_t + \mu_t' m_t + \frac{1}{2}\phi_t' Q_t^{-1}\phi_t - \frac{(\tilde{y}_t - v_t)^2}{2r_t}\right) \end{aligned} \tag{D1}$$

where $m_t = E(\beta_t^*|y^t, z, \Theta)$ and $M_t = V(\beta_t^*|y^t, z, \Theta)$ are computed by a Kalman filter, $r_t = q_t + \tilde{x}_t' M_{t-1}\tilde{x}_t$ with $\tilde{x}_t \equiv x_t \odot v$, $q_t = \sigma_t^2 + \tilde{x}_t' W_t \tilde{x}_t$, $W_t = \text{diag}(\omega_t)$ with $\omega_t = [\omega_{1,t} \ ... \ \omega_{K,t}]'$ and $\omega_{j,t} = \text{I}\left\{z_{j,t} > \frac{\Phi^{-1}(q_j)}{\sqrt{1-\rho^2}}\right\}$ for $j = 1, ..., K$, $Q_t = I_K + T_t'\Omega_t T_t$, $T_t T_t' = M_t$, $\phi_t = T_t'(\mu_t - \Omega_t m_t)$, $v_t = \tilde{x}_t' m_{t-1}$ and $\tilde{y}_t = y_t - x_t'\beta_0$. The quantities of $\mu_t$ and $\Omega_t$ are computed by a backward recursion:

$$\begin{aligned} &\Omega_n = 0, \ \ \mu_n = 0 \\ &\Omega_{t-1} = A_t'(\Omega_t - \Omega_t C_t D_t^{-1} C_t' \Omega_t) A_t + \frac{\tilde{x}_t \tilde{x}_t'}{q_t} \\ &\mu_{t-1} = A_t'(I_K - \Omega_t C_t D_t^{-1} C_t')(\mu_t - \Omega_t b_t \tilde{y}_t) + \frac{\tilde{x}_t \tilde{y}_t}{q_t} \end{aligned} \tag{D2}$$

where $b_t = \frac{W_t \tilde{x}_t}{q_t}$, $A_t = I_K - b_t \tilde{x}_t'$, $C_t' C_t = W_t - \frac{W_t \tilde{x}_t \tilde{x}_t' W_t}{q_t}$ and $D_t = I_K + C_t' \Omega_t C_t$. The derivation details of the algorithm can be found in the original paper of Gerlach et al. (2000) and are not repeated here to save space. The Kalman filter is a standard technique for linear Gaussian state space systems for which a concise description can be found in Gerlach et al. (2000) and a textbook treatment can be found in Hamilton (1994).

# E Sampling the Parameters $q$ and $\rho$

The target is to sample from $p(\theta_q|y, z, \Theta_{-\theta_q})$ with $\theta_q = \{q, \rho\}$ by an MH-within-Gibbs sampler over the two blocks of $q$ and $\rho$.

A common ingredient of the conditional posteriors is the likelihood function $p(y|z, \Theta)$ that is computed by running a Kalman filter to integrate out $\beta^*$. Specifically, decompose the likelihood as $p(y|z, \Theta) = p(y_1|z, \Theta) \prod_{t=1}^{n-1} p(y_{t+1}|y^t, z, \Theta)$ where $y^t = \{y_1, ..., y_t\}$. Each component in the likelihood can be written as an integral:

$$p(y_{t+1}|y^t, z, \Theta) = \int p(y_{t+1}|\beta^*_{t+1}, y^t, z, \Theta) p(\beta^*_{t+1}|y^t, z, \Theta) d\beta^*_{t+1}$$

It can be derived that the first component inside the integral is a normal density of the form $N(x'_{t+1}\beta_0 + (x_{t+1} \odot \beta^*_{t+1})'v, \sigma^2_{t+1})$ while the second component

$$p(\beta^*_{t+1}|y^t, z, \Theta) = \int p(\beta^*_{t+1}|\beta^*_t, z, \Theta) p(\beta^*_t|y^t, z, \Theta) d\beta^*_t$$

is also a normal density $N(m_t, M_t + W_{t+1})$, where the mean $m_t = E(\beta^*_t|y^t, z, \Theta)$ and the covariance matrix $M_t = V(\beta^*_t|y^t, z, \Theta)$ can be computed through a Kalman filter and $W_{t+1} = \text{diag}(\omega_{t+1})$ is the conditional covariance matrix of $\beta^*_{t+1}$ with $\omega_t = [\omega_{1,t} \ ... \ \omega_{K,t}]'$ and $\omega_{j,t} = \text{I}\left\{ z_{j,t} > \frac{\Phi^{-1}(q_j)}{\sqrt{1-\rho^2}} \right\}$ for $j = 1, ..., K$. It follows that $p(y_{t+1}|y^t, z, \Theta)$ is a normal density with the mean $x'_{t+1}\beta_0 + (x_{t+1} \odot v)'m_t$ and the variance $\sigma^2_{t+1} + (x_{t+1} \odot v)'(M_t + W_{t+1})(x_{t+1} \odot v)$.

The initial component $p(y_1|z, \Theta) = \int p(y_1|\beta^*_1, z, \Theta) p(\beta^*_1|z, \Theta) d\beta^*_1$ is computed by inserting $p(\beta^*_1|z, \Theta) = N(0, W_1)$ and is normal with the mean $x'_1\beta_0$ and the variance $\sigma^2_1 + (x_1 \odot v)'W_1(x_1 \odot v)$.

With the formula of the likelihood at hand, the conditional posteriors in the MH-within-Gibbs sampler can be readily calculated by combining the priors specified in Section 2.1. There are two minor complications in the sampling. Because the parameters are constrained, reparameterization is needed when using Gaussian random walk proposals in the MH steps. The priors of the transformed parameters can be derived by the

change-of-variables formula. Specifically, the parameter $\rho \in (-1, 1)$ is reparametrized as $\rho^* = \log\left(\frac{1+\rho}{1-\rho}\right)$ with the prior $p(\rho^*) \propto (1+\exp(-\rho^*))^{-\tau_{\rho,1}}(1+\exp(\rho^*))^{-\tau_{\rho,2}}$. Each probability $q_j$ is reparametrized as $q_j^* = \log\left(\frac{q_j}{1-q_j}\right)$ with the prior $p(q_j^*) \propto \frac{1}{2+\exp(-q_j^*)+\exp(q_j^*)}$. Another minor complication is that the parameter $\rho^*$ enters the equation for $z_t$ and thus the conditional posterior is $p(\rho^*|y, z, \Theta_{-\rho^*}) \propto p(\rho^*)p(z|\rho^*)p(y|z, \Theta)$. The component $p(z|\rho^*)$ can be calculated through the factorization $p(z|\rho^*) = p(z_1|\rho^*)\prod_{t=2}^{n} p(z_t|z_{t-1}, \rho^*)$ and is proportional to $(1-\rho^2)^{\frac{K}{2}} \exp\left(-\frac{(1-\rho^2)\sum_{j=1}^{K} z_{j,1}^2 + \sum_{j=1}^{K}\sum_{t=2}^{n}(z_{j,t}-\rho z_{j,t-1})^2}{2}\right)$ where $\rho = \frac{\exp(\rho^*)-1}{\exp(\rho^*)+1}$.

# F Tuning Metropolis-Hastings Steps

The adaptive optimal scaling method of Garthwaite et al. (2016) allows for automatic scaling of random walk MH steps towards a target acceptance probability. The method is effective and has been successfully applied in studies such as Gunawan et al. (2019) and Kreuzer and Czado (2020).

To show the use of the adaptive optimal scaling method for the LTVP model, consider the example of drawing the latent variable $z_t$. The proposal for its $i+1^{\text{th}}$ draw is a random walk $z_t(i+1) \sim N(z_t(i), \omega_i^2 A)$ where $A$ equals $I_K$ when $i \leq i^*$ and the sample covariance matrix $\frac{1}{i}\sum_{j=1}^{i} z_t(j)z_t(j)' - \frac{1}{i^2}\sum_{j=1}^{i} z_t(j)\sum_{j=1}^{i} z_t(j)'$ when $i > i^*$.[12] The scalar $i^*$ is a fixed threshold to avoid unstable sample covariance matrix of $z_t$ when $i$ is small. The scalar $\omega_i^2$ is updated according to $\log(\omega_{i+1}) = \log(\omega_i) + \frac{c}{d_{c,i}}(p_i - p^*)$ where $p_i$ is the MH acceptance probability in the $i^{\text{th}}$ draw of $z_t$ and $p^*$ is the target acceptance probability. The scalar $c$ for updating $\omega_i$ is determined as:

$$c = \frac{1}{Kp^*(1-p^*)} + \left(1 - \frac{1}{K}\right)\frac{\sqrt{2\pi}\exp\left(\frac{\alpha_\omega^2}{2}\right)}{2\alpha_\omega} \tag{F1}$$

where $\alpha_\omega$ satisfies $\Phi(-\alpha_\omega) = \frac{p^*}{2}$. As suggested in Garthwaite et al. (2016), the scalar $d_{c,i}$ for

[12] To avoid the risk of near-singular sample covariance matrix, one can add $\frac{\epsilon}{i}I_K$ to $A$ in the $i+1^{\text{th}}$ draw where $\epsilon$ is a small positive number (e.g. 1e-6).

updating $\omega_i$ is set as $\max(\frac{i}{K}, d_c^*)$, where $d_c^*$ is a fixed threshold to avoid that $\omega_i$ converges before the sample covariance matrix of $z_t$ stabilizes. The update of the scalar $\omega_i$ is re-started whenever $\log(\omega_i)$ changes more than $\log(3)$ from its value at the start or the most recent re-start in order to reduce the impact of a poor starting value of $\omega_i$. In this paper, I set $p^* = 0.25$, $i^* = 100$ and $d_c^* = 200$. The configuration of the adaptive MH step for the parameter block $q^*$ is the same as in the example of $z_t$.

For the univariate parameter block $\rho^*$, the proposal for its $i + 1^{\text{th}}$ draw is a random walk $\rho^*(i+1) \sim N(\rho^*(i), \omega_i^2)$ where the scalar $\omega_i^2$ is updated as $\log(\omega_{i+1}) = \log(\omega_i) + \frac{c}{i}(p_i - p^*)$ where $p_i$ is the MH acceptance probability in the $i^{\text{th}}$ draw of $\rho^*$ and $p^*$ is the target acceptance probability. The scalar $c$ is determined as in Equation (F1) by setting $K = 1$.

# G ASIS Boosting

At the end of each MCMC sweep, two ASIS steps are performed to boost the Gibbs sampler described in Section 3. Theoretical details of the ASIS can be found in the original paper of Yu and Meng (2011).

The first ASIS step involves the normalized time-varying coefficients $\beta^*$ and the parameters $\{v, \beta_0\}$ and is based on replacing each $\beta_{j,t}^*$ with $\beta_{j,t} = \beta_{j,0} + v_j \beta_{j,t}^*$ such that $\Delta\beta_{j,t} \sim N(0, v_j^2 \omega_{j,t})$ with $\omega_{j,t} \equiv \mathrm{I}\left\{ z_{j,t} > \frac{\Phi^{-1}(q_j)}{\sqrt{1-\rho^2}} \right\}$ for $j = 1, ..., K$ and $t = 1, ..., n$. The specific steps are as follows:

1. Compute $\beta_{j,t} = \beta_{j,0} + v_j \beta_{j,t}^*$ for $j = 1, ..., K$ and $t = 1, ..., n$.

2. Find the set $\mathcal{T}_j = \{t : \omega_{j,t} = 1\}$ for each $j = 1, ..., K$. If $\mathcal{T}_j$ is non-empty, keep the sign of $v_j$ and then sample $v_j^2 \sim GIG\left(\frac{1-n_j}{2}, \frac{1}{\tau_v}, \sum_{t \in \mathcal{T}_j} \Delta\beta_{j,t}^2\right)$, where GIG stands for the generalized inverse Gaussian distribution and $n_j$ is the length of $\mathcal{T}_j$.[13] Update $v_j$

[13]Sampling from the GIG distribution is by adapting the Matlab function **gigrnd** written by Enes Makalic and Daniel Schimidt that implements an algorithm from Devroye (2014).

as $\sqrt{v_j^2}$ times the sign of $v_j$ before entering the ASIS step.

3. If $\omega_{j,1} = 1$, update $\beta_{j,0} \sim N(b_{j,0}, B_{j,0})$ with $B_{j,0}^{-1} = \tau_0^{-1}\tau_j^{-1} + v_j^{-2}$ and $B_{j,0}^{-1} b_{j,0} = v_j^{-2}\beta_{j,1}$ for $j = 1, ..., K$.

4. Compute back $\beta_{j,t}^* = (\beta_{j,t} - \beta_{j,0})/v_j$ for $j = 1, ..., K$ and $t = 1, ..., n$.

The resulting $\beta_0$, $v$ and $\{\beta_t^*\}_{t=1}^n$ from this ASIS step are used as their final draw in an MCMC sweep.

The second ASIS step involves the latent variable $z$ and the parameter $q$. Define $\mu_{q,j} \equiv \Phi^{-1}(q_j)$ for $j = 1, ..., K$. It follows that $\mu_{q,j} \sim N(0,1)$ since $q_j \sim U(0,1)$. The steps are as follows:

1. Compute $z_{j,t}^* = z_{j,t} - \frac{\mu_{q,j}}{\sqrt{1-\rho^2}}$ for $j = 1, ..., K$ and $t = 1, ..., n$. It follows that $z_{j,1}^* \sim N\left(\frac{-\mu_{q,j}}{\sqrt{1-\rho^2}}, \frac{1}{1-\rho^2}\right)$ and $z_{j,t}^* - \rho z_{j,t-1}^* \sim N\left(\frac{(\rho-1)\mu_{q,j}}{\sqrt{1-\rho^2}}, 1\right)$ for $t = 2, ..., n$.

2. Treat $\mu_{q,j}$ as the parameter of a linear regression based on $\{z_{j,t}^*\}_{t=1}^n$. Update $\mu_{q,j} \sim N(b_{q,j}, B_{q,j})$ where $B_{q,j}^{-1} = 2 + \frac{(n-1)(1-\rho)}{1+\rho}$ and $B_{q,j}^{-1} b_{q,j} = -\sqrt{1-\rho^2} z_{j,1}^* + \frac{\rho-1}{\sqrt{1-\rho^2}} \sum_{t=2}^n z_{j,t}^* - \rho z_{j,t-1}^*$ for $j = 1, ..., K$.

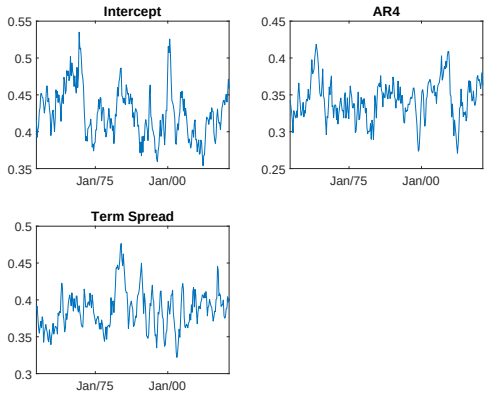
3. Compute back $q_j = \Phi(\mu_{q,j})$ and $z_{j,t} = z_{j,t}^* + \frac{\mu_{q,j}}{\sqrt{1-\rho^2}}$ for $j = 1, ..., K$ and $t = 1, ..., n$.

The resulting $q$ and $\{z_t\}_{t=1}^n$ from this ASIS step are used as their final draw in an MCMC sweep.

# H Additional Results: Industrial Output

Figure H1 shows the point-wise posterior mean of $\omega$ for the 3 regressors identified as being time varying: the intercept, the AR 4 lag and the term spread. As discussed in Section 4.2, the coefficients of these 3 regressors are found to have infrequent changes based on the posterior of the stationary probabilities $q$. Inspecting Figure H1, the point-wise posterior

Figure H1: Estimate of $\omega$ for Time-Varying Coefficients in the LTVP Model: Industrial Output




Note: This figure shows the point-wise posterior mean of $\omega$ of the time-varying coefficients in the LTVP model for predicting the industrial output.

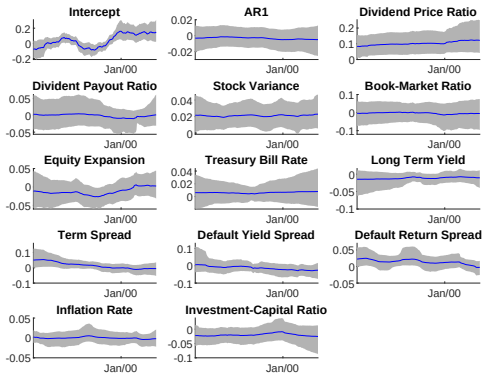
mean of $\omega$ for each coefficient clearly has frequent spikes of various magnitudes and would need substantial subjective judgement to identify the locations of parameter shifts.

# I Additional Results: Equity Premium

Figure I1 shows the point-wise posterior median and 90% credible set of $\beta_t$ while Figure I2 shows the posterior densities of $v^*$ for all the regressors. The description of the economic regressors is provided in Table 1.

Inspecting Figure I2, only the posterior density of $v^*$ for the intercept shows clear bi-modality, consistent with the finding in the main text based on the mode of $|v|$. Examining the point-wise 90% credible set of $\beta_t$ in Figure I1, the *dividend price ratio* and the *stock variance* are the two predictors that have consistently positive coefficients over the data sample. The value of zero is consistently around the point-wise $95^{\text{th}}$ percentile of the coefficients of the *long term yield* and the *investment-capital ratio*, supporting that these two predictors have weakly negative predictive relations with the equity premium. The

Figure I1: Estimate of $\beta_t$ in the LTVP Model: Equity Premium




Note: This figure shows the point-wise posterior median (solid line) and 90% credible set (gray area) of $\beta_t$ in the LTVP model for predicting the equity premium.

other economic predictors appear to be insignificant based on the estimates of $\beta_t$ and $v^*$.

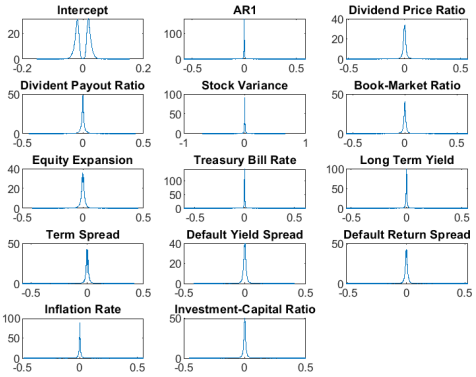
# J Alternative TVP Models

The alternative TVP models include the triple gamma (TG) model of Cadonna et al. (2020), the mixture innovation (MI) model of Giordani and Kohn (2008), the dynamic horseshoe (DH) model of Kowal et al. (2019) and the latent threshold (LT) model of Nakajima and West (2013). The models are briefly described by using the framework of Equation (1).

Let $h_{j,t}$ be the conditional variance of $\Delta\beta_{j,t}$ for $j = 1, ..., K$ and $t = 1, ..., n$. The TG, MI and DH models differ by how they specify $h_{j,t}$:

- The TG model is a static shrinkage one that specifies $h_{j,t} = h_j$ with $h_j \sim G(0.5, 2\tau_0\tau_j)$ where the hyper-parameters $\tau_0$ and $\tau_j$ could follow rather general distributions. In this paper, I use the special version that amounts to using the horseshoe prior of Carvalho et al. (2010) for $\pm\sqrt{h_j}$: $\sqrt{\tau_0} \sim C^+(0, 1/n)$ and $\sqrt{\tau_j} \sim C^+(0, 1)$ for all $j$.

Figure I2: Estimate of $v^*$ in the LTVP Model: Equity Premium




Note: This figure shows the posterior distribution of the parameters $v^*$ in the LTVP model for predicting the equity premium.

- The MI model specifies $h_{j,t} = w_j c_{j,t}$ where the mixture indicator $c_{j,t} \in \{0, 1\}$. For computational feasibility, only 3 possible scenarios for the vector of mixture indicators $c_t = [c_{1,t} \ ... \ c_{K,t}]'$ are allowed: all coefficients being constant at time $t$ ($c_{j,t} = 0$ for all $j$), all coefficients changing at time $t$ ($c_{j,t} = 1$ for all $j$) and only the intercept changing at time $t$ ($c_{1,t} = 1$ while $c_{j,t} = 0$ for all $j > 1$). The prior probability of each scenario is $1/3$. The probability of staying in the same scenario is $\pi$ while transiting to each of the other 2 scenarios is $\frac{1-\pi}{2}$. In estimation the prior is $G(0.5, 20)$ for the slab variance $w_j$ for all $j$ and is $Beta(50, 0.5)$ for the probability $\pi$.

- The DH model specifies $h_{j,t} = \tau_0 \tau_j \tau_{j,t}$ where $\sqrt{\tau_0} \sim C^+(0, 1/n)$, $\sqrt{\tau_j} \sim C^+(0, 1)$ and $\log(\tau_{j,t}) = \rho_j \log(\tau_{j,t-1}) + \xi_{j,t}$ with $\xi_{j,t} \sim Z(0.5, 0.5, 0, 1)$ following a Z distribution for all $j$ and $t$. In estimation, the prior for $\rho_j$ is $N(0.95, 1)\mathrm{I}\{-1 < \rho_j < 1\}$ for all $j$.

The LT model takes the following form to dynamically reset "small" coefficients to zero:

$$y_t = \sum_{j=1}^{K} x_{j,t} \beta_{j,t} \mathrm{I}\{|\beta_{j,t}| > d_j\} + \epsilon_t, \ \epsilon_t \sim N(0, \sigma_t^2)$$

$$\beta_{j,t} = (1 - \rho_j)\mu_j + \rho_j \beta_{j,t-1} + \eta_{j,t}, \ \eta_{j,t} \sim N(0, h_j)$$

where $\sigma_t^2$ follows the SV process and $\beta_{j,1} \sim N(0, h_j/(1-\rho_j^2)$ for all $j$ and $t$. In estimation, the priors are set as $d_j \sim G(0.5, 2)$, $\mu_j \sim N(0, 1)$, $h_j \sim G(0.5, 2)$ and $\frac{1+\rho_j}{2} \sim Beta(20, 1.5)$ for all $j$. Given these priors, the ASIS boosting of Yu and Meng (2011) is applied to the block $\{\mu_j, h_j\}$ by transforming $\beta_{j,t}^* = \frac{\beta_{j,t}-\mu_j}{h_j^*}$ where $h_j^* = \pm\sqrt{h_j}$ for all $j$ and $t$.

# References


Belmonte, M., G. Koop, and D. Korobolis (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting 33*, 80–94.

Bitto, A. and S. Fruhwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*, 75–97.

Cadonna, A., S. Fruhwirth-Schnatter, and P. Knaus (2020). Triple the gamma - a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics 8*(2), 20.

Carvalho, C., N. Polson, and J. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*, 465–480.

Chang, Y., Y. Choi, and J. Park (2017). A new approach to model regime switching. *Journal of Econometrics 196*(1), 127–143.

Clements, M. and D. Hendry (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass., MIT Press.

Cogley, T. and T. Sargent (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics 8*, 262–302.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*, 157–181.

Devroye, L. (2014). Random variate generation for the generalized inverse Gaussian distribution. *Statistics and Computing 24*, 239–246.

Dufays, A., Z. Li, J. Rombouts, and Y. Song (2021). Sparse change-point VAR models. *Journal of Applied Econometrics 36*, 703–727.

Dufays, A. and J. Rombouts (2020). Relevant parameter changes in structural break models. *Journal of Econometrics 217*, 46–78.

Durbin, J. and S. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–615.

Estrella, A., A. Rodrigues, and S. Schich (2003). How stable is the predictive power of the yield curve? evidence from Germany and the United States. *The Review of Economics and Statistics 85*(3), 629–644.

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

Fruhwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for Gaussian and partially non-Gaussian state space models. *Journal of Econometrics 154*, 85–100.

Garthwaite, P., Y. Fan, and S. Sisson (2016). Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. *Communications in Statistics - Theory and Methods 45*(17), 5098–5111.

Gerlach, R., C. Carter, and R. Kohn (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association 95*, 819–828.

Geweke, J. and C. Whiteman (2006). Bayesian forecasting. In J. Geweke and C. Whiteman (Eds.), *Handbook of Economic Forecasting*, Volume 1. Elsevier.

Geyer, C. (1992). Practical Markov chain Monte Carlo. *Statistical Science 7*, 473–483.

Giordani, P. and R. Kohn (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics 26*, 66–77.

Gunawan, D., M. Tran, K. Suzuki, J. Dick, and R. Kohn (2019). Computationally efficient Bayesian estimation of high-dimensional Archimedean copulas with discrete and mixed margins. *Statistics and Computing 29*, 933–946.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Haubrich, J. (2020). Does the yield curve predict output? Federal Reserve Bank of Cleveland, Working Paper No. 20-34.

Hauzenberger, N., F. Huber, and G. Koop (2020). Dynamic shrinkage priors for large time-varying parameter regressions using scalable Markov chain Monte Carlo methods. arXiv:1810.09004v1 [stat.ME].

Huber, F., G. Kastner, and M. Feldkircher (2019). Should I stay or should I go? a latent threshold approach to large-scale mixture innovation models. *Journal of Applied Econometrics 34*(5), 621–640.

Huber, F., G. Koop, and L. Onorante (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business and Economic Statistics 39*(3), 669–683.

Huber, F. and M. Pfarrhofer (2021). Dynamic shrinkage in time-varying parameter stochastic volatility in mean models. *Journal of Applied Econometrics 36*, 262–270.

Kalli, M. and J. Griffin (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics 178*, 779–793.

Kastner, G. and S. Fruhwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis 76*, 408–423.

Kowal, D., D. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*, 781–804.

Kreuzer, A. and C. Czado (2020). Efficient Bayesian inference for nonlinear state space models with univariate autoregressive state equation. *Journal of Computational and Graphical Statistics 29*, 523–534.

Makalic, E. and D. Schmidt (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*(1), 179–182.

McCausland, W., S. Miller, and D. Pelletier (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis 55*(1), 199–212.

McCulloch, R. and R. Tsay (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association 88*, 968–978.

Nakajima, J. and M. West (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics 31*, 151–164.

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

Piironen, J. and A. Vehtari (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics 11*(2), 5018–5051.

Primiceri, G. (2005). Time varying structural autoregressions and monetary policy. *Review of Economic Studies 72*(3), 821–852.

Rockova, V. and K. McAlinn (2021). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis 16*(1), 233–269.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*, 325–338.

Stock, J. and M. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*, 11–30.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical Systems Approach*. Oxford: Oxford University Press.

Uribe, P. and H. Lopes (2020). Dynamic sparsity on dynamic regression models. arXiv:2009.14131v1 [stat.ME].

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Yu, Y. and X. Meng (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.