Achieving Dynamic Sparsity in Time-Varying Parameter Regressions

Zhongfang He*

January 8, 2022

Abstract

Time-varying parameter regressions are investigated in a framework where the conditional variances of parameter innovations are modeled as deterministic functions of latent continuous variables and equal zero when these functions of latent variables are below fixed thresholds. Thus parameter innovations can be turned on and off locally to achieve dynamic sparsity while permitting flexible time variation patterns in the parameters. The use of latent continuous variables as the driver of sparsity allows efficient Bayesian model estimation by adapting the algorithm of Gerlach et al. (2000) while avoiding the curse of dimensionality. The new model is illustrated by examples of simulated and real economic data.

Keywords: TVP, Mixture Innovation, Adaptive MH, ASIS
JEL Codes: C11, C22, E37, G17

*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 useful for improved model inference and forecast (Clements and Hendry (1999)). Two critical issues face the researchers when tackling the varying parameter problem. First, how to avoid overfitting caused by forcing all parameters to be time varying? Moreover, for time varying parameters, what kind of time variation should be specified? 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 unnecessary parameter shifts.

Recent Bayesian studies have applied shrinkage methods to automatically differentiate constant and varying parameters but often require the researchers to pre-specify if the time variation in parameters is continuous (time-varying parameter or TVP approach) or infrequent (change-point approach). Examples include Fruhwirth-Schnatter and Wagner (2010), Bitto and Fruhwirth-Schnatter (2019) and Dufays and Rombouts (2020)) etc. One exception is the mixture innovation approach of Giordani and Kohn (2008) where the innovations of time-varying parameters follow spike-and-slab distributions and hence can be switched on and off at each time t to permit local parameter shift or constancy. As a result, both continuous and infrequent parameter shifts as well as their mixtures can be accommodated in the mixture innovation model, while the spike component in the spike-and-slab distributions of the parameter innovations can turn off unnecessary parameter changes to encourage model parsimony.

Despite its theoretical appeal, estimating the mixture innovation model suffers from the curse of dimensionality in that simulating the mixture indicators of the spike-and-slab distributions requires to evaluate the model likelihood under all 2^K combinatorial scenarios of the mixture indicators at each time t for a model with K regressors and quickly becomes impractical when K grows[1]. As an alternative, this paper proposes a new model framework

[1]See He (2021) for an approach that mimics the mixture innovation model but is computationally more efficient to be applied to larger regression problems.

where spike-and-slab-like distributions are specified for the innovations of time-varying parameters to obtain the theoretical benefits of the mixture innovation model while being computationally efficient for practical applications.

The critical ingredient of the proposed model is that the conditional variances of parameter innovations are time varying and are modeled as deterministic functions of auxiliary latent continuous variables. When a latent variable falls into certain range at time t, the conditional variance of the corresponding parameter innovation is truncated to be zero at time t and hence leads to local parameter constancy, which effectively plays the role of the spike component in the spike-and-slab distributions of the mixture innovation model.

Using latent continuous variables to drive the sparsity in parameter innovations is computationally more efficient than directly modeling the discrete mixture indicators as in the mixture innovation model. Instead of evaluating the model likelihood under all combinatorial scenarios of the discrete mixture indicators at each time t, sampling the latent continuous variables of time t in the proposed model needs only to evaluate the model likelihood twice in Metropolis-Hastings (MH) steps and hence greatly reduces the computation burden, rendering it feasible to estimate larger TVP regression models with spike-and-slab-like parameter innovations.

The latent continuous variables are modeled as independent Gaussian random walks to facilitate estimation and to permit serial persistence in the conditional variances of parameter innovations. Given the truncation mechanism, it is flexible to choose the link function mapping the latent continuous variables to conditional variances of parameter innovations. In this paper, I examine the two simple cases of the linear and square functions, both of which are truncated to be zero when below certain fixed thresholds.

An Markov chain Monte Carlo (MCMC) scheme is developed to estimate the proposed model. The algorithm of Gerlach et al. (2000) is adapted to sample the latent continuous variables efficiently by integrating out the time-varying regression coefficients. For improved sampling efficiency, the MCMC sampler is further boosted via the ancilarity-sufficiency interweaving strategy (ASIS) of Yu and Meng (2011).

The proposed approach is illustrated in a simulation study and 2 predictive exercises for the U.S. equity premium and inflation rate where the number of regressors ranges from

14 to 22 and it would be computationally prohibitive if the mixture innovation model is applied without any further restrictions. Out-of-sample forecast performance of the proposed approach is compared with 3 alternative TVP models that allow dynamic shrinkage of time-varying parameters: a restricted version of the mixture innovation model, the dynamic horseshoe model of Kowal et al. (2019), and the logistic mixture innovation model of He (2021). Similar to the proposed approach, the restricted mixture innovation model permits local constancy in time varying parameters but it limits the set of possible parameter movements at each time t for computational practicality. The latter two alternative TVP models shrink the time-varying conditional variances of parameter innovations to be approximately zero, but not exactly zero as in the proposed approach, when time variations in regression coefficients are unnecessary in certain time periods. The proposed approach is found to perform well and is competitive against these alternative models.

The remainder of the paper is organized as follows. Section 2 presents the proposed model. Section 3 describes the estimation algorithm and the computation of predictive likelihood for evaluating model forecasts. Section 4 and 5 illustrate the proposed approach in a simulated example and two empirical applications. Section 6 concludes. Additional details are provided in appendices.

2 The Model

The model framework for TVP regressions is as follows:

$$\begin{aligned} &y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ &\Delta\beta_{j,t} \sim N(0, f(z_{j,t})), \quad \beta_{j,0} \sim N(0, v_{j,0}^2) \\ &\Delta z_{j,t} \sim N(0, a_j^2), \quad z_{j,0} \sim N(0, a_j^2 a_{j,0}^2) \end{aligned} \tag{1}$$

where y_t is a scalar dependent variable, x_t is a K-dimensional vector of regressors, $\beta_t = [\beta_{1,t} \ ... \ \beta_{K,t}]'$ is the corresponding time-varying coefficients, $z_t = [z_{1,t} \ ... \ z_{K,t}]'$ is a vector of auxiliary latent variables that drive the conditional variances of β_t through the link function $f(\cdot)$ and the symbol Δ denotes the difference operator such that $\Delta\beta_{j,t} = \beta_{j,t} - \beta_{j,t-1}$ and $\Delta z_{j,t} = z_{j,t} - z_{j,t-1}$ for $j = 1, ..., K$ and $t = 1, ..., n$. The variables $\beta_0 = [\beta_{1,0} \ ... \ \beta_{K,0}]'$ and $z_0 = [z_{1,0} \ ... \ z_{K,0}]'$ are the initial values of the regression coefficients and the latent variables

respectively and are assume to follow zero-mean Gaussian distributions with the variances $v_0^2 = [v_{1,0}^2 \ ... \ v_{K,0}^2]'$ and $a^2 \odot a_0^2$ where $a^2 = [a_1^2 \ ... \ a_K^2]'$ collects the conditional variances of z_t, $a_0^2 = [a_{1,0}^2 \ ... \ a_{K,0}^2]'$ and the symbol $\odot$ denotes the component-wise product of two matrices of the same size.

The variable ϵ_t is the innovation of the dependent variable and is allowed to have time varying conditional variance σ_t^2 by a stochastic volatility (SV) model:

$$\log(\sigma_t^2) = (1-\rho)\mu + \rho \log(\sigma_{t-1}^2) + \epsilon_{y,t}, \ \ \epsilon_{y,t} \sim N(0, \sigma_y^2), \tag{2}$$

where $\log(\sigma_1^2) \sim N(\mu, \sigma_y^2/(1-\rho^2))$. The homoskedastic case can be accommodated by imposing $\sigma_t^2 = \sigma^2$.

To permit dynamic sparsity for each regression coefficient $\beta_{j,t}$, this paper considers the link function of the form:

$$f(z_{j,t}) = \max\{g(z_{j,t}) - c_j, 0\} \tag{3}$$

where $g(\cdot)$ is a function with support on the real line and c_j is a fixed threshold. Truncation of the conditional variance of $\beta_{j,t}$ to zero occurs when $g(z_{j,t}) \leq c_j$, which leads to the local constancy $\beta_{j,t} = \beta_{j,t-1}$ at time t. If $g(z_{j,t}) \leq c_j$ holds for all t, one obtains the global constancy $\beta_{j,t} = \beta_{j,0}$ over $t = 1, ..., n$. Thanks to the truncation mechanism in the link function $f(\cdot)$, virtually any function $g(\cdot)$ with support on the real line could be used in Equation (3) in principle, though the resulting model efficiency may differ. In this paper, I consider the two simple cases $g(z_{j,t}) = z_{j,t}$ and $g(z_{j,t}) = z_{j,t}^2$. Note that for the former, truncation occurs from below when $z_{j,t} \leq c_j$, while for the square function, truncation occurs in the center when $-\sqrt{c_j} \leq z_{j,t} \leq \sqrt{c_j}$ with positive c_j.

For ease of reference, the model of Equation (1) and (3) will be labeled as the *dynamic truncation* or DT model hereafter with the shorthand DTL and DTS referring to the DT model with the link function of the truncated linear and square one respectively.

2.1 Connection with Existing Models

A typical form of the mixture innovation model of Giordani and Kohn (2008) would specify the dynamics of the time-varying regression coefficients as $\Delta\beta_{j,t} \sim N(0, d_{j,t}\sigma_{\beta,j}^2)$ where $d_{j,t} \in \{0, 1\}$ is the mixture indicator and $\sigma_{\beta,j}^2$ is the time-invariant part of the conditional

variance of $\beta_{j,t}$ for $j = 1, ..., K$ and $t = 1, ..., n$. The mixture indicator $d_{j,t}$ is a Bernoulli variable and is able to turn on and off the innovation of $\beta_{j,t}$ locally at each time t. To see the connection with the proposed DT model, the max function in Equation (3) can be equivalently re-written as $f(z_{j,t}) = I\{g(z_{j,t}) \geq c_j\}(g(z_{j,t}) - c_j)$ where the indicator function $I\{g(z_{j,t}) \geq c_j\}$ plays the same role as the the mixture indicator $d_{j,t}$ in the mixture innovation model to control the activation of local parameter shift. In this sense, the DT model also uses a spike-and-slab-like distribution for the innovation of $\beta_{j,t}$.

In fact, if one chooses the link function $f(z_{j,t}) = I\{g(z_{j,t}) \geq c_j\}\sigma^2_{\beta,j}$ in Equation (1), the resulting DT model would resemble an mixture innovation model with the mixture indicator $d_{j,t} = I\{g(z_{j,t}) \geq c_j\}$. However, in this case the model likelihood will be essentially flat with respect to the latent variable $z_{j,t}$ except for occasional jumps when $z_{j,t}$ crosses the thresholds. This can be seen by integrating out $\beta_{j,t}$ by a Kalman filter. As a result, observations of y_t are not informative of $z_{j,t}$ and the posterior draws of model parameters tend to mix rather poorly. As an alternative, He (2021) provides a logistic mixture innovation (*LMI* hereafter) model that sets $g(z_{j,t}) = z_{j,t}$ and replaces the indicator function $I\{z_{j,t} \geq c_j\}$ by a logistic one to solve the estimation difficulty while closely approximating the original mixture innovation model. Relative to the LMI model, the proposed DT model is able to set the conditional variances of unnecessary parameter innovations to be exactly, rather than approximately, zero while allowing for efficient model estimation.

The proposed DT model also connects with the recent literature of Bayesian dynamic shrinkage priors for TVP regressions. Examples include Kalli and Griffin (2014), Kowal et al. (2019), Uribe and Lopes (2020) and Rockova and McAlinn (2021)) etc. 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 proposed DT model. Hauzenberger et al. (2020) offers an approach that employs a sparse approximation to posterior draws of time-varying regression parameters from dynamic shrinkage priors. The proposed DT model differs from Hauzenberger et al. (2020) by directly allowing for sparsity in the model specification and thus no approximation is required.

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)). In particular, Nakajima and West (2013) provides an approach that 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 is able to accommodate episodical combinations of smooth variations and constant of zero in regression coefficients. These studies apply the indicator function to truncate the level of the regression coefficients. In contrast, the indicator function in the proposed DT model operates on the innovation of regression coefficients and thus allows a much wider range of possible time variation patterns in the coefficients.

3 Bayesian Estimation and Forecast

3.1 Reparameterization

In the DTL specification, the threshold c_j and the initial latent variable $z_{j,0}$ can not be jointly identified. For identification, it is set $c_j = 0$ in this paper. A benefit of this strategy is that the remaining parameter $z_{j,0}$ can be integrated out to simplify model estimation when sampling the latent variables $\{z_{j,t}\}_{t=1}^{n}$. By setting $c_j = 0$ and normalizing the scale factor of the latent variable $z_{j,t}$, Equation (1) under the DTL specification can be re-written as:

$$\begin{aligned} & y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ & \Delta\beta_{j,t} \sim N(0, v_j^2 \max\{z_{j,t}^*, 0\}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2) \\ & \Delta z_{j,t}^* \sim N(0,1), \quad z_{j,0}^* \sim N(0, a_{j,0}^2) \end{aligned} \tag{4}$$

where $z_{j,t}^* = \frac{z_{j,t}}{\sqrt{a_j^2}}$ is the rescaled latent variable and $v_j^2 = \sqrt{a_j^2}$ for $j = 1, ..., K$ and $t = 1, ..., n$. The benefit of the normalization is that the parameter a_j^2 is moved into the conditional variance of $\beta_{j,t}$ that will directly connect with the dependent variable y_t after integrating out β_t by Kalman filter and hence can be sampled more efficiently.

Similarly, substituting $z_{j,t}^* = \frac{z_{j,t}}{\sqrt{a_j^2}}$ into Equation (1) under the DTS specification leads

to:

$$
\begin{aligned}
&y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
&\Delta\beta_{j,t} \sim N(0, v_j^2 \max\{\left(z_{j,t}^*\right)^2 - c_j^*, 0\}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2) \\
&\Delta z_{j,t}^* \sim N(0,1), \quad z_{j,0}^* \sim N(0, a_{j,0}^2)
\end{aligned} \tag{5}
$$

where $v_j^2 = a_j^2$ and $c_j^* = \frac{c_j}{a_j^2}$ is the rescaled threshold for $j = 1, ..., K$.

The reparameterization in Equation (4) and (5) shares a two-component multiplicative structure for the conditional variance of each regression coefficient $\beta_{j,t}$ where a time-invariant component applies to all t while a time-varying component adapts to the local feature at each time t. Equation (4) and (5) will be the basis to specify priors and design MCMC schemes for the DTL and DTS models in this paper.

3.2 Prior Specification

The parameter v_j^2 in Equation (4) and (5) is the time invariant part of the conditional variance of $\beta_{j,t}$. Borrowing from Bitto and Fruhwirth-Schnatter (2019), I specify a prior for the signed square root $v_j = \pm\sqrt{v_j^2} \sim N(0, \tau_{v,j})$, which is equivalent to a gamma prior[2] $G(0.5, 2\tau_{v,j})$ for v_j^2. In estimation, the hyper-parameter $\tau_{v,j}$ is fixed at 10 to form weakly diffused priors for v_j over $j = 1, ..., K$.

The initial regression coefficient $\beta_{j,0}$ plays the role of fixed regression coefficients[3]. The horseshoe prior of Carvalho et al. (2010) is employed to shrink insignificant $\beta_{j,0}$. Specifically the prior variance of $\beta_{j,0}$ is $v_{j,0}^2 = \tau_0\tau_j$ with $\tau_0 \sim IB(0.5, 0.5)$ and $\tau_j \sim IB(0.5, 0.5)$ for $j = 1, ..., K$, where IB denotes the inverted beta distribution[4].

For the SV specification of the variance σ_t^2 of the dependent variable (Equation (2)), the priors are $\mu \sim N(0, 10)$, $\rho \sim N(0.95, 0.04)I_{\{-1<\rho<1\}}$, $\sigma_y^2|s_y \sim G(0.5, 2s_y)$ and $s_y \sim IB(0.5, 0.5)$. In the homoskedastic case $\sigma_t^2 = \sigma^2$, the Jeffery's prior $\sigma^2 \propto \frac{1}{\sigma^2}$ is used.

[2]The gamma distribution $G(\alpha, \beta)$ for a generic variable x has the density $\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp(-\frac{x}{\beta})$

[3]To see this, note that $y_t = x_t'\beta_t + \epsilon_t \Rightarrow y_t = x_t'\beta_0 + x_t'\beta_t^* + \epsilon_t$ with $\beta_t^* = \beta_t - \beta_0$.

[4]The density of an inverted beta distribution $IB(a, b)$ is $p(x) = \frac{x^{a-1}(1+x)^{-a-b}}{B(a,b)}I\{x > 0\}$ for a generic random variable x where $B(\cdot, \cdot)$ is the beta function and a and b are positive real numbers. If $x \sim IB(0.5, 0.5)$, then $\sqrt{x} \sim C^+(0, 1)$ and vice versa, where $C^+(0, 1)$ is a standard half-Cauchy distribution with the density $p(z) = \frac{2}{\pi(1+z^2)}I\{z > 0\}$.

For the rescaled initial latent variable $z^*_{j,0}$, its prior variance $a^2_{j,0}$ is set to be 10 in estimation and leads to a weekly diffused prior.

In the DTS specification, the prior for the rescaled threshold c^*_j is $N(0, 10)$. It is possible to enforce positive c^*_j by specifying a prior with the support of positive real numbers, though in my experiments the normal prior tends to produce posterior draws that mix better than alternatives such as log normal or gamma prior specifications. The normal prior for c^*_j is innocuous for model interpretation since a negative c^*_j simply implies that no truncation is warranted and a continuously time-varying $\beta_{j,t}$ is supported by the data.

3.3 MCMC Sampler

A Gibbs sampler is developed to estimate the DTL and DTS models and is described in this section[5]. For notational ease, let $y = \{y_t\}^n_{t=1}$, $x = \{x_t\}^n_{t=1}$, $\beta = \{\beta_t\}^n_{t=1}$, $z^*_t = [z^*_{1,t} \ ... \ z^*_{K,t}]'$, $z^* = \{z^*_t\}^n_{t=1}$, $v = [v_1 \ ... \ v_K]'$, $c^* = [c^*_1 \ ... \ c^*_K]'$, $\theta_0 = \{\tau_0, \tau_1, ..., \tau_K\}$, θ_σ include σ^2 in the homoskedastic case or $\{\sigma^2_t\}^n_{t=1}$ and associated other parameters in the SV model, and $\theta = \{\theta_0, \theta_\sigma\}$.

DTL Specification: For the DTL model, the target is to sample from the posterior distribution $p(\beta, z^*, v, \beta_0, \theta|y, x)$. Note that the rescaled initial latent variable $z^*_0 = [z^*_{1,0} \ ... \ z^*_{K,0}]'$ is not sampled because it can be integrated out and there is no inherent interest in z^*_0 per se. A Gibbs sampler divides the model parameters into 3 blocks θ_0, θ_σ and $\{\beta, z^*, v, \beta_0\}$ and is described below.

Conditional on β_0, the posterior of each hyper-parameter in θ_0 is an inverse gamma distribution based on the hierarchical representation of the inverted beta distribution in Makalic and Schmidt (2016). The details of sampling θ_0 are provided in Appendix A.

In the case of the SV model, the sampler of Kastner and Fruhwirth-Schnatter (2014) is applied to sample θ_σ with the details provided in Appendix B. In the homoskedastic case $\sigma^2_t = \sigma^2$, the posterior is $\sigma^2|y, x, \beta \sim \text{IG}\left(\frac{n}{2}, \frac{1}{2}\sum^n_{t=1}\epsilon^2_t\right)$ where $\epsilon_t = y_t - x'_t\beta_t$ and IG denotes the inverse gamma distribution.

[5]A similar MCMC sampler appeared in He (2021) for estimating the LMI model of time-varying parameter regressions.

For the block $\{\beta, z^*, v, \beta_0\}$, the posterior is decomposed as:

$$p(\beta, z^*, v, \beta_0|y, x, \theta) = p(z^*, v, \beta_0|y, x, \theta)p(\beta|y, x, \theta, z^*, v, \beta_0)$$

The part $p(\beta|y, x, \theta, z^*, v, \beta_0)$ is sampled as the latent states in a linear Gaussian state space system by the simulation smoother of Durbin and Koopman (2002). For the part $p(z^*, v, \beta_0|y, x, \theta)$ that integrates out β, a nested Metropolis-within-Gibbs sampler is applied to iterate over the following sub-blocks:

1. $z^*|y, x, \theta, v, \beta_0$: A single-move Gibbs sampler is applied to sample from:

$$p(z_t^*|y, x, \theta, v, \beta_0, z_{-t}^*) \propto p(z_t^*|z_{-t}^*)p(y|x, \theta, z^*, v, \beta_0)$$

   The prior part $p(z_t^*|z_{-t}^*)$ is a normal distribution thanks to the random walk specification of z_t^*. For the likelihood part, note that only the components relevant to z_t^* need to be computed for the MH step at time t. Let $y^t = \{y_1, ..., y_t\}$ and y^0 be an empty set. One can apply the factorization $p(y|x, \theta, z^*, v, \beta_0) \propto p(y_t, ..., y_n|y^{t-1}, x, \theta, z^*, v, \beta_0)$ for sampling z_t^*. The right-hand side of the factorization can be computed efficiently by the algorithm of Gerlach et al. (2000) (*GCK* hereafter). The details are provided in Appendix C.

   Efficient sampling of the rescaled latent variable z_t^* is a critical component of the MCMC sampler. It is worth noting that, to sample z_t^* and hence computing the conditional variances of β_t in the DTL model, the likelihood $p(y|x, \theta, z^*, v, \beta_0)$ only needs to be evaluated twice via the GCK algorithm in an MH step. In contrast, directly sampling the mixture indicator in the conditional variances of β_t from the mixture innovation model of Giordani and Kohn (2008) would require to evaluate the model likelihood by the GCK algorithm over all 2^K combinatorial scenarios of the mixture indicator and is not generally practical.

2. $v|y, x, \theta, z^*, \beta_0$: The posterior is $p(v|y, x, \theta, z^*, \beta_0) \propto p(v)p(y|x, \theta, z^*, v, \beta_0)$. The prior $p(v)$ is normal and is specified in Section 3.2. The likelihood part $p(y|x, \theta, z^*, v, \beta_0)$ can be written as:

$$p(y|x, \theta, z^*, v, \beta_0) = \prod_{t=0}^{n-1} p(y_{t+1}|y^t, x, \theta, z^*, v, \beta_0)$$

Each component $p(y_{t+1}|y^t, x, \theta, z^*, v, \beta_0)$ in the likelihood function can be derived as a normal distribution by applying a Kalman filter to integrate out the time-varying coefficients β_t. Details of computing the likelihood $p(y|x, \theta, z^*, v, \beta_0)$ are provided in Appendix D.

3. $\beta_0|y, x, \theta, z^*, v$: The posterior is $p(\beta_0|y, x, \theta, z^*, v) \propto p(\beta_0|\theta_0)p(y|x, \theta, z^*, v, \beta_0)$. Following the discussions of drawing v, the prior $p(\beta_0|\theta_0)$ is normal as specified in Section 3.2 while the likelihood $p(y|x, \theta, z^*, v, \beta_0)$ is computed as in Appendix D.

MH steps are heavily used in the sampler. To avoid manual intervention, the adaptive optimal scaling method of Garthwaite et al. (2016) is adopted to automatically tune the scale factor of MH steps that employ the Gaussian random walk as the proposal distributions and target acceptance rates around 25%. The details of MH tuning are provided in Appendix E.

DTS Specification: For the DTS model, the target is to sample from the posterior distribution $p(\beta, z^*, c^*, v, \beta_0, \theta|y, x)$. The Gibbs sampler is similar to the one for the DTL model except for the new block of the rescaled thresholds c^*. Drawing c^* is conducted together with z^*, v and β_0 in the block $p(z^*, c^*, v, \beta_0|y, x, \theta)$ through a nested Metropolis-within-Gibbs sampler, within which sampling z^*, v and β_0 are similar to the steps for the DTL model. The conditional posterior for c^* is $p(c^*|y, x, \theta, z^*, v, \beta_0) \propto p(c^*)p(y|x, \theta, z^*, c^*, v, \beta_0)$. The prior $p(c^*)$ is a normal distribution as specified in Section 3.2. Computing the likelihood $p(y|x, \theta, z^*, v, \beta_0)$ is similar to the steps for drawing v and β_0 and can be found in Appendix D.

ASIS Boosting: In my experiments, the Gibbs sampler for the DT models could sometimes produce posterior draws that mix unsatisfactorily. For example, when a regression coefficient contains occasional structural breaks as in the simulation example of this paper, draws of the corresponding parameter v could mix poorly. To boost the Gibbs sampler, the *ancillarity-sufficiency interweaving strategy* (ASIS) of Yu and Meng (2011) is adopted that connects the sampling from different parametrizations of the DT models and thus allows the sampler to explore the parameter space more efficiently. Details of the ASIS can be found in the original paper of Yu and Meng (2011).

To implement the ASIS for the DT models, Equation (1) is reparametrized as:

$$
\begin{aligned}
&y_t = x_t'\beta_0 + (x_t \odot \beta_t^*)'v + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
&\Delta\beta_{j,t}^* \sim N(0, w_{j,t}), \quad \beta_{j,0}^* = 0
\end{aligned} \tag{6}
$$

where $\beta_{j,t}^* = \frac{\beta_{j,t}-\beta_{j,0}}{v_j}$ is the normalized regression coefficient, $\beta_t^* = [\beta_{1,t}^* \ ... \ \beta_{K,t}^*]'$, and $w_{j,t}$ equals $\max\{z_{j,t}^*, 0\}$ in the DTL model and $\max\{\left(z_{j,t}^*\right)^2 - c_j^*, 0\}$ in the DTS model for $j = 1$, ..., K[6]. It is clear that β_0 and v become the fixed coefficients of a linear regression model conditional on β_t^* and σ_t^2 in Equation (6). The specific steps of the ASIS boosting for the DT models are as follows:

1. In each MCMC sweep after running the Gibbs sampler, compute $\beta_t^* = (\beta_t - \beta_0) \odot \frac{1}{v}$ for $t = 1$, ..., n.

2. Let α be a $2K$-by-1 vector stacking β_0 and v. Conditional on β_t^*, σ_t^2 and the hyper-parameters θ_0, draw α from a linear regression with the posterior $N(b_\alpha, B_\alpha)$, where $B_\alpha^{-1} = B_0^{-1} + \sum_{t=1}^n \frac{1}{\sigma_t^2}\tilde{x}_t\tilde{x}_t'$, $B_\alpha^{-1}b_\alpha = \sum_{t=1}^n \frac{1}{\sigma_t^2}\tilde{x}_t y_t$, B_0 is a $2K$-by-$2K$ diagonal matrix with the diagonal elements $\tau_0\tau_1$, ..., $\tau_0\tau_K$, $\tau_{v,1}$, ..., $\tau_{v,K}$, and $\tilde{x}_t$ stacks x_t and $x_t \odot \beta_t^*$.

3. Compute back $\beta_t = \beta_t^* \odot v + \beta_0$ for $t = 1$, ..., n.

The resulting β_0, v and β_t from the ASIS step are used as their final draw in an MCMC sweep. With a marginal computation cost, the ASIS step noticeably improves the sampling efficiency of the DT models in my experiments.

3.4 Predictive Likelihood

For model comparison, this paper uses the predictive likelihood that integrates out any latent variables and fixed parameters in each model under study. See Geweke and Whiteman (2006) and Geweke and Amisano (2010) for reviews that advocate the use of the predictive likelihood as a coherent approach for model evaluation. Specifically, let $\sigma^{2,t} = \{\sigma_1^2, ..., \sigma_t^2\}$, $\theta_{-\sigma}$ denote the parameters in θ except $\sigma^{2,t}$, Θ denote the fixed parameters $\{\theta_{-\sigma}, v, \beta_0\}$ for the

[6]Such a reparameterization of a TVP model appeared in Fruhwirth-Schnatter and Wagner (2010) and has been shown to improve the sampling quality of model parameters (Bitto and Fruhwirth-Schnatter (2019)).

DTL model and $\{\theta_{-\sigma}, c^*, v, \beta_0\}$ for the DTS one. The one-step ahead predictive likelihood for the observation y_{t+1} is $p(y_{t+1}|y^t)$ by treating the regressors x_1, ..., x_{t+1} as pre-determined and integrating out the fixed parameters Θ and variables $\sigma^{2,t+1}$, $\beta^{t+1} = \{\beta_1, ..., \beta_{t+1}\}$ and $z^{*,t+1} = \{z_1^*, ..., z_{t+1}^*\}$.

In my experiments, the numerical approximation of $p(y_{t+1}|y^t)$ directly based on posterior draws of β^t tends to be unstable. A more stable numerical integration is obtained by using a Kalman filter to integrate out β^t. Denote the filtering distribution of β_t by a Kalman filter as $\beta_t|y^t, z^{*,t}, \sigma^{2,t}, \Theta \sim N(m_t, M_t)$. Substituting $\beta_{t+1} = \beta_t + \Delta\beta_{t+1}$ into the equation for y_{t+1} leads to $y_{t+1} = x'_{t+1}\beta_t + x'_{t+1}\Delta\beta_{t+1} + \epsilon_{t+1}$. It follows that:

$$y_{t+1}|y^t, z^{*,t+1}, \sigma^{2,t+1}, \Theta \sim N(x'_{t+1}m_t, \;\; x'_{t+1}M_t x_{t+1} + \sigma^2_{t+1} + x'_{t+1}W_{t+1}x_{t+1}) \tag{7}$$

where W_{t+1} is the conditional covariance matrix of β_{t+1} and is a diagonal matrix with each diagonal element $v_j^2 \max\{z^*_{j,t+1}, 0\}$ in the DTL model and $v_j^2 \max\{(z^*_{j,t+1})^2 - c_j^*, 0\}$ in the DTS model for $j = 1$, ..., K. A numerical approximation for the predictive likelihood can be computed as:

$$\begin{aligned} p(y_{t+1}|y^t) &= \int p(y_{t+1}|y^t, z^{*,t+1}, \sigma^{2,t+1}, \Theta) p(z^{*,t+1}, \sigma^{2,t+1}, \Theta|y^t) dz^{*,t+1} d\sigma^{2,t+1} d\Theta \\ &\approx \frac{1}{M}\sum_{i=1}^{M} p(y_{t+1}|y^t, z^{*,t+1}_{(i)}, \sigma^{2,t+1}_{(i)}, \Theta_{(i)}) \end{aligned} \tag{8}$$

where the integrand $p(y_{t+1}|y^t, z^{*,t+1}, \sigma^{2,t+1}, \Theta)$ is computed per Equation (7) based on posterior draws $z^{*,t}_{(i)}$, $\sigma^{2,t}_{(i)}$ and $\Theta_{(i)}$ as well as simulations of z^*_{t+1} and σ^2_{t+1} conditional on $z^{*,t}_{(i)}$, $\sigma^{2,t}_{(i)}$ and $\Theta_{(i)}$ for $i = 1$, ..., M.

Given one-step ahead predictive likelihoods $p(y_t|y^{t-1}, \mathcal{M}_1)$ and $p(y_t|y^{t-1}, \mathcal{M}_2)$ from two competing models $\mathcal{M}_1$ and $\mathcal{M}_2$, a predictive Bayes factor for observation y_t can be computed as $\text{BF}_t = \frac{p(y_t|y^{t-1}, \mathcal{M}_1)}{p(y_t|y^{t-1}, \mathcal{M}_2)}$ where a positive BF_t indicates the data evidence favoring $\mathcal{M}_1$ over $\mathcal{M}_2$ for observation y_t and vice versa when BF_t is negative. The cumulative sum $\sum_{j=1}^{h} \log(\text{BF}_{t+j})$ over a prediction sample $h = 1$, ..., H reveals the entire evolution of the relative forecast performance of $\mathcal{M}_1$ over $\mathcal{M}_2$ and is used in this paper to compare competing models.

4 Simulation Example

A simulation study is conducted to test the ability of the proposed DT models to estimate regression coefficients that exhibit different yet plausible time variation patterns. A time series of length $n = 300$ is simulated from a linear regression model with 6 coefficients as follows:

1. Random walk: $\beta_{1,t} = \sum_{j=1}^{t} u_j$ with $u_j \sim N(0, 0.01)$.
2. Change point: $\beta_{2,t} = I_{\{t_1 < t \leq t_2\}} + 0.5\, I_{\{t > t_2\}}$.
3. Mixture of constant and slanted line:
$$\beta_{3,t} = \frac{t - t_1}{t_2 - t_1} I_{\{t_1 < t \leq t_2\}} + I_{\{t > t_2\}}$$
4. Mixture of constant and random walk:
$$\beta_{4,t} = \left(\sum_{j=1}^{t} u_j \right) I_{\{t_1 < t \leq t_2\}} + I_{\{t > t_2\}}$$
with $u_j \sim N(0, 0.01)$.
5. Ones: $\beta_{5,t} = 1$.
6. Zeros: $\beta_{6,t} = 0$.

where $t_1 = 100$ and $t_2 = 200$. Coefficient 3 and 4 are designed to approximate the situation where the relevance of a regressor moves gradually from one equilibrium to a new one. The regressors in the simulation are drawn from standard normal distributions. The dependent variable is generated by adding a noise from $N(0, \sigma_0^2)$ where $\sigma_0 = 1.1227$ is calibrated to reach a ratio of σ_0^2 to the variance of the dependent variable around 0.2[7].

In estimation, the dependent variable is taken to be homoskedastic, i.e. $\sigma_t^2 = \sigma^2$. Posterior analysis is conducted based on 10,000 draws by discarding 5,000 burn-ins and keeping every 5$^{\text{th}}$ iteration. Producing 1,000 posterior draws from the DT models takes about 100 seconds on a standard desktop computer with a 3.0 GHz Intel Core i5 CPU

[7]Two other cases where the ratio of σ_0^2 to the variance of the dependent variable is 0.5 and 0.8 respectively are also examined in the experiments. The estimation results are qualitatively similar.

running in MATLAB R2020b. The thinned posterior draws mix reasonably well. For example, the inefficiency factor (*IF* hereafter) of v_j^2 is less than 100 in the DTL and DTS models[8]. For the DTS model, the IF of the rescaled threshold c_j^* is less than 30. In both DT models, the acceptance rates of the MH steps for v, c^*, β_0 and z_t^* over $t = 1, ..., 300$ are all between 22% and 26% and are close to the target value of 25%.

Figure 1 plots the point-wise posterior median and 90% credible set of the regression coefficient β_t by the DTL model, along with the true coefficient value. It can be seen that β_t is well estimated by the DTL model. The point-wise 90% credible set properly covers the true value of β_t despite the heterogeneous shapes of the coefficients. The point-wise posterior median and credible set of β_t by the DTS model is visually similar to those by the DTL model and is not shown to save space.

To compare the coefficient estimates by the DTL and DTS models, Figure 2 shows the point-wise root mean squared error (RMSE) of β_t by the two DT models:

$$\text{RMSE}_{j,t} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\hat{\beta}_{j,t}^{(i)} - \beta_{j,t}\right)^2} \tag{9}$$

where $\hat{\beta}_{j,t}^{(i)}$ denotes the i^{th} posterior draw of the coefficient from a given model and $\beta_{j,t}$ the true coefficient value for $j = 1, ..., 6$, $t = 1, ..., 300$, and $i = 1, ..., M$. The DTL model produces slightly smaller RMSEs for the two constant coefficients (coefficient 5 and 6) over some time points than the DTS model, though the overall estimation accuracy of the 6 coefficients in β_t by the two DT models is on the same level.

The point-wise posterior median and 90% credible set of the conditional standard deviation of β_t by the two DT model are shown in Figure 3 and are broadly consistent with the shapes of β_t. For example, estimates of the conditional standard deviation for the two constant coefficients (coefficient 5 and 6) are uniformly close to zero over the time points while for the change point coefficient (coefficient 2), there are pronounced spikes in the conditional standard deviations around the change points. On the other hand, the point-wise 90% credible sets of the conditional standard deviations are relatively wide and reflect

[8]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. As the sign of v_j is unidentified and hence its posterior draws could exhibit bi-modal shapes around zero, using the IF to evaluate the autocorrelations for v_j^2 is more meaningful than directly for v_j.

a large amount of uncertainty about the local movement of the coefficients at each time t, which renders it difficult to infer local parameter changes precisely. It is notable that the point-wise 5^{th} percentile of the conditional standard deviations is mostly zero across t thanks to the sparsification mechanism in the DT models. Relative to the DTS model, the DTL model produces more zero conditional standard deviations and hence more frequent local constancy in the coefficients but on the other hand appears to be more sensitive to the impact of local noise.

5 Empirical Examples

In this section, the proposed DT models are investigated in two predictive exercises for the U.S. equity premium and inflation rate[9]. The results of in-sample estimates are described in Section 5.1 and 5.2. An out-of-sample comparison of the DT models with alternative TVP regression models that apply dynamic shrinkage is presented in Section 5.3.

5.1 Equity Premium

As a first example, the proposed DT models are applied to predict the U.S. equity premium one quarter ahead with the predictors examined in Welch and Goyal (2008). The dependent variable in the DT models 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). The sample runs from Q1 1947 to Q4 2020 with a total of 296 quarterly observations. The data is kindly provided by Amit Goyal in his website[10].

Besides the economic predictors, the estimated models also include an intercept and an first-order autoregressive lag of the equity premium. The conditional variance of the dependent variable is modeled by the SV specification of Equation (2). In estimation, all

[9]Similar predictive studies were conducted in Kalli and Griffin (2014) and He (2021).

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

non-constant regressors are normalized by subtracting their sample means and dividing by their sample standard deviations. A total of 10,000 draws are kept for analysis with a burn-in length of 5,000 and a thinning frequency of 1-in-5. Sampling 1,000 posterior draws takes about 120 and 135 seconds from the DTL and DTS models respectively.

Estimated regression coefficients β_t from the two DT models are similar. Among the 12 economic predictors, only the *dividend price ratio* is clearly significant throughout the data sample with the zero point consistently outside the point-wise 90% credible set of its coefficient. The point-wise median of the coefficient of the *dividend price ratio* is positive and appears to gradually increase over time, reminiscent of the finding in Dangl and Halling (2012) that the importance of the dividend-lagged-price ratio grows after the initiation of Rule 10b-18 permitting share buybacks in the 1980s. The coefficients of two other predictors *long term yield* and *investment-to-capital ratio* have their point-wise 95^{th} percentiles near zero for most of the time periods and hence could be considered as being marginally significant. Figure 4 illustrates the point-wise posterior median and 90% credible set of the coefficients of these 3 predictors from the two DT models. The other predictors either go from significant to insignificant (*term spread*, *default return spread*) or are insignificant for most of the time periods.

Figure 4 also compares the point-wise posterior median and 90% credible set of the conditional variance of equity premium, which is evidently volatile over time. It can be seen that the two DT models produce comparable model fit and lead to very similar conditional variance estimate for equity premium.

5.2 Inflation Rate

As a second example, I fit the DT models to the U.S. inflation rate. The dependent variable is the quarter-to-quarter change of the quarterly inflation rate which is computed as the difference of log quarterly GDP deflator. The predictors are the 1-quarter lag of 20 economic variables including real activity variables, interest rates and other macroeconomic indicators as well as an intercept and an first-order autoregressive lag of the dependent variable. Table 2 describes the economic variables.

The data source is the FRED database of the U.S. Federal Reserve Bank of St. Louis

except for the S&P500 index which is from Robert Shiller's website[11]. The data sample is from Q2 1966 to Q1 2021 with 220 quarterly observations. The quarterly value of monthly variables are constructed as the monthly averages within each quarter. Similar to the equity premium prediction model, the SV model of Equation (2) is used for the conditional variance of the dependent variable. Non-constant regressors are normalized in estimation. Posterior analysis is based on 10,000 draws after discarding 5,000 burn-ins and retaining every 5^{th} iteration. For this application, it takes about 140 and 160 seconds to produce 1,000 draws from the DTL and DTS models respectively.

The point-wise 90% credible set of β_t is used to compare the statistical significance of predictors. Parameter estimates from the two DT models are close. Based on the point-wise credible set, the autoregressive lag is the only non-constant predictor that is significant throughout the data sample. The point-wise median of the autoregressive coefficient is negative and gradually decreases over time, broadly consistent with the finding of declining persistence in the U.S. inflation rate dynamics in studies such as Cogley and Sargent (2002) and Cogley et al. (2010). The predictors *GDP*, *employee* and *unemployment* are insignificant at the beginning of the sample but gradually become significant. To illustrate the time variation patterns, Figure 5 shows the point-wise posterior median and 90% credible set of the coefficients of the autoregressive lag, *GDP*, *employee* and *unemployment* from the two DT models. The other predictors either become insignificant in recent years (*investment*, *expenditure*, *producer price*) or have nearly flat coefficients around zero throughout the data sample.

The point-wise posterior median and 90% credible set of the conditional variance of the inflation rate change by the two DT models are also shown in Figure 4. The volatility of the inflation rate change is high in the 1970s and, to a lesser extent, the time periods since 2014-2015 while staying relatively low between the two episodes. Similar to the equity premium study, conditional variance estimates of the inflation rate change by the two DT models are rather close.

[11]The web address is: http://www.econ.yale.edu/ shiller/data.htm.

5.3 Out-of-Sample Forecast

This section compares the out-of-sample forecast performance of the proposed DT models and 3 alternative TVP specifications that permit dynamic shrinkage.

The first is a restricted version of the mixture innovation model of Giordani and Kohn (2008) (*RMI* hereafter). For computational practicality, the RMI model only allows 3 combinatorial scenarios of parameter change at each time t: all coefficients are constant at time t, all coefficients are time varying at time t, and only the coefficient of the intercept is time varying at time t[12]. Like the proposed DT models, the RMI model permits the parameters to change or stay constant locally at each time t. The other two alternative models permit locally adaptive shrinkage, though not exact sparsity, for time varying parameters. One is the dynamic horseshoe (*DHS* hereafter) model of Kowal et al. (2019) that specifies an autoregressive process for the logarithm of the conditional variance of each regression coefficient. The other one is the LMI model of He (2021) that closely approximates the mixture innovation model of Giordani and Kohn (2008) but is computationally more practical. Appendix F describes the details of the 3 alternative TVP models.

The predictions are recursive from Q1 2011 to Q4 2020 (equity premium prediction) and Q1 2021 (inflation rate prediction). For each quarter t in the forecast sample, the one-quarter ahead predictive likelihood is computed as in Section 3.4 based on the data up to time $t-1$ and the realization of the dependent variable at time t. In forecasts, the non-constant predictors are recursively normalized by the sample moments of the estimation sample only. Cumulative log predictive Bayes factor as described in Section 3.4 is used to compare the models.

For the two DT models, Figure 6 shows the cumulative log predictive Bayes factor of the DTL model versus the DTS one in the equity premium and inflation rate prediction exercises. The DTL model performs marginally better than the DTS one in the equity premium study with the maximum cumulative log predictive Bayes factor of 0.95. But overall speaking, the two DT models return the same level of forecast performance.

[12]For example, without any restriction in the mixture innovation model, sampling 1,000 draws from the equity premium prediction model (14 regressors) would require over 66 hours as there are $2^{14} = 16,384$ combinatorial scenarios of parameter change at each time t.

Turning to the comparison with alternative TVP models, Figure 7 shows the cumulative log predictive Bayes factor of the alternative TVP models versus the DTL model in the two prediction studies. The RMI model underperforms the DTL model throughout the prediction samples with a cumulative log predictive Bayes factor of -17 at the end of the prediction sample in the equity premium study and -1.9 in the inflation rate study. Relative to the DHS model, the DTL model steadily accumulates gains in predictive likelihoods in the two prediction studies. At the end of the prediction samples, the cumulative log predictive Bayes factor of the DHS model versus the DTL model is -5.9 in the equity premium study and -5.6 in the inflation rate study, providing consistent supportive evidence for the DTL model. The LMI model performs slightly better than the DTL model in the equity premium study with the maximum cumulative log predictive Bayes factor of 1.3 around the end of the prediction sample while being comparable with the DTL model in the inflation rate study with the cumulative log predictive Bayes factor fluctuating around zero within a range of ± 0.5.

Figure 8 illustrates the cumulative log predictive Bayes factor of the alternative TVP models versus the DTS model in the two prediction studies. Similar to the DTL model, the DTS model outperforms the RMI and DHS models and is overall comparable to the LMI model in terms of predictive likelihoods.

6 Conclusion

This paper presents a framework that allows the coefficients of a TVP regression model to vary or stay constant locally and achieves dynamic sparsity for the time varying parameters. Similar to the mixture innovation model of Giordani and Kohn (2008), the conditional variance of each parameter innovation in the proposed model has a dynamic spike-and-slab-like structure that controls whether or not there is a random parameter shift at each time t and therefore allows for very flexible time variation patterns in the parameters, relieving the practitioners of the duty to pre-determine a specific pattern for the time varying parameters. But unlike the mixture innovation model, the proposed model uses auxiliary latent continuous variables and a thresholding mechanism to drive the switches between the spike and slab components for parameter innovations and is able to adapt the

efficient algorithm of Gerlach et al. (2000) for estimating larger TVP regression models. Applications of the proposed model to simulated and real economic data show encouraging results.

Appendix

A Hyper-Parameters of Horseshoe Prior

The horseshoe prior of the initial regression coefficients is $\beta_{j,0} \sim N(0, \tau_0\tau_j)$ with $\tau_0 \sim IB(0.5, 0.5)$ and $\tau_j \sim IB(0.5, 0.5)$ for $j = 1, ..., K$. Following Makalic and Schmidt (2016), the inverted beta distributions are represented as hierarchical inverse gamma ones by introducing auxiliary variables::

$$\tau_0 \sim IB(0.5, 0.5) \Longleftrightarrow \tau_0|\kappa_0 \sim IG\left(0.5, \frac{1}{\kappa_0}\right), \;\; \kappa_0 \sim IG(0.5, 1)$$
$$\tau_j \sim IB(0.5, 0.5) \Longleftrightarrow \tau_j|\kappa_j \sim IG\left(0.5, \frac{1}{\kappa_j}\right), \;\; \kappa_j \sim IG(0.5, 1)$$

with the following posteriors:

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

B Estimating SV Model

The sampler of Kastner and Fruhwirth-Schnatter (2014) is adapted to estimate the SV model of Equation (2). The log linearization strategy of Omori et al. (2007) is first applied to approximate the logarithm of a $\chi^2(1)$-distributed variable by a mixture of normal distributions. A key ingredient of Kastner and Fruhwirth-Schnatter (2014) is applying the ASIS 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. The details of the method 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_y in the gamma prior $\sigma_y^2 \sim G(0.5, 2s_y)$, this paper specifies a prior $s_y \sim IB(0.5, 0.5)$ to determine s_y in a data driven way. The conditional posterior of s_y can be obtained by applying the hierarchical inverse gamma representation in Makalic and Schmidt (2016): $s_y|a_y, \sigma_y^2 \sim IG\left(1, \frac{1}{a_y} + \frac{\sigma_y^2}{2}\right)$ where a_y is an auxiliary variable with the prior $a_y \sim IG(0.5, 1)$ and the posterior $a_y|s_y \sim IG\left(1, 1 + \frac{1}{s_y}\right)$.

C Sampling the Latent Variable z_t^*

The target is to sample from the posterior $p(z_t^*|y, x, \theta_z, z_{-t}^*) \propto p(z_t^*|z_{-t}^*)p(y|x, z^*, \theta_z)$ where θ_z denotes $\{\theta, v, \beta_0\}$ in the DTL model and $\{\theta, c^*, v, \beta_0\}$ in the DTS model to simplify notation.

Given the random walk specification of z_t^* in Equation (4) and (5), the prior part can be written as $p(z_t^*|z_{-t}^*) \propto p(z_t^*|z_{t-1}^*)p(z_{t+1}^*|z_t^*)$ for $t = 2, ..., n-1$. One can derive $p(z_t^*|z_{-t}^*) = N(b_{z,t}, B_{z,t})$ where $B_{z,t}^{-1} = 2I_K$ and $B_{z,t}^{-1}b_{z,t} = z_{t-1} + z_{t+1}$ for $t = 2, ..., n-1$. When $t = 1$, one has $p(z_1^*|z_{-1}^*) \propto p(z_1^*)p(z_2^*|z_1^*)$ where $p(z_1^*) = N\left(0, \text{diag}\left(1 + a_0^2\right)\right)$ by integrating out z_0^*. It is straightforward to derive $p(z_1^*|z_{-1}^*) = N(b_{z,1}, B_{z,1})$ with $B_{z,1}^{-1} = I_K + \text{diag}\left(\frac{1}{1+a_0^2}\right)$ and $B_{z,1}^{-1}b_{z,1} = z_2^*$. When $t = n$, the posterior is simply $p(z_n^*|z_{-n}^*) \propto p(z_n^*|z_{n-1}^*) = N(z_{n-1}^*, I_K)$.

Denote $y^t = \{y_1, ..., y_t\}$ and $y^{t,n} = \{y_t, ..., y_n\}$. The GCK algorithm is applied to compute the components of the likelihood $p(y|x, z^*, \theta_z)$ that is relevant to sampling z_t^*:

$$\begin{aligned} p(y|x, z^*, \theta_z) &\propto p(y^{t,n}|y^{t-1}, x, z^*, \theta_z) \\ &\propto r_t^{-\frac{1}{2}} \det(Q_t)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}m_t'\Omega_t m_t + \mu_t' m_t + \frac{1}{2}\phi_t' Q_t^{-1}\phi_t - \frac{(y_t - v_t)^2}{2r_t}\right) \end{aligned} \quad \text{(C1)}$$

where $m_t = E(\beta_t|y^t, x, z^*, \theta_z)$ and $M_t = V(\beta_t|y^t, x, z^*, \theta_z)$ are computed by a Kalman filter, $r_t = q_t + x_t' M_{t-1} x_t$, $q_t = \sigma_t^2 + x_t' W_t x_t$, $Q_t = I_K + T_t'\Omega_t T_t$, $T_t T_t' = M_t$, $\phi_t = T_t'(\mu_t - \Omega_t m_t)$ and $v_t = x_t' m_{t-1}$. The quantity W_t is the conditional covariance matrix of β_t and is a diagonal matrix with each diagonal element $v_j^2 \max\{z_{j,t}^*, 0\}$ in the DTL model and $v_j^2 \max\{(z_{j,t}^*)^2 -$

$c_j^*, 0\}$ in the DTS model for $j = 1, ..., K$. The quantities of μ_t and Ω_t are computed by a backward recursion:

$$\begin{aligned}
&\Omega_n = 0, \;\; \mu_n = 0 \\
&\Omega_{t-1} = A_t'(\Omega_t - \Omega_t C_t D_t^{-1} C_t' \Omega_t) A_t + \frac{x_t x_t'}{q_t} \\
&\mu_{t-1} = A_t'(I_K - \Omega_t C_t D_t^{-1} C_t')(\mu_t - \Omega_t b_t y_t) + \frac{x_t y_t}{q_t}
\end{aligned} \tag{C2}$$

where $b_t = \frac{W_t x_t}{q_t}$, $A_t = I_K - b_t x_t'$, $C_t' C_t = W_t - \frac{W_t x_t x_t' W_t}{q_t}$ and $D_t = I_K + C_t' \Omega_t C_t$. The details of derivation of the GCK 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).

D Computing the Model Likelihood Function

The likelihood function $p(y|x, z^*, \theta_z)$ of the DT models can be computed by running a Kalman filter to integrate out the time-varying coefficients β_t, where θ_z denotes $\{\theta, v, \beta_0\}$ in the DTL model and $\{\theta, c^*, v, \beta_0\}$ in the DTS model. Specifically, decompose the likelihood as $p(y|x, z^*, \theta_z) = p(y_1|x, z^*, \theta_z) \prod_{t=1}^{n-1} p(y_{t+1}|y^t, x, z^*, \theta_z)$ where $y^t = \{y_1, ..., y_t\}$.

To compute $p(y_{t+1}|y^t, x, z^*, \theta_z)$, first consider the distribution $p(y_{t+1}|\beta_t, y^t, x, z^*, \theta_z)$. Let η_t be a vector collecting $\beta_{j,t} - \beta_{j,t-1}$ over $j = 1, ..., K$. By substituting $\beta_{t+1} = \beta_t + \eta_{t+1}$ into the equation $y_{t+1} = x_{t+1}'\beta_{t+1} + \epsilon_{t+1}$, it is straightforward to show $p(y_{t+1}|\beta_t, y^t, x, z^*, \theta_z) = N(x_{t+1}'\beta_t, \sigma_{t+1}^2 + x_{t+1}' W_{t+1} x_{t+1})$ where W_{t+1} is the conditional covariance matrix of β_{t+1} and is a diagonal matrix with each diagonal element $v_j^2 \max\{z_{j,t+1}^*, 0\}$ in the DTL model and $v_j^2 \max\{(z_{j,t+1}^*)^2 - c_j^*, 0\}$ in the DTS model for $j = 1, ..., K$. Next write $p(y_{t+1}|y^t, x, z^*, \theta_z) = \int p(y_{t+1}|\beta_t, y^t, x, z^*, \theta_z) p(\beta_t|y^t, x, z^*, \theta_z) d\beta_t$. The distribution $p(\beta_t|y^t, x, z^*, \theta_z)$ is normal where the mean $m_t = E(\beta_t|y^t, x, z^*, \theta_z)$ and the covariance matrix $M_t = V(\beta_t|y^t, x, z^*, \theta_z)$ can be computed through a Kalman filter. It can be shown that $p(y_{t+1}|y^t, x, z^*, \theta_z)$ is also normal with the mean $x_{t+1}' m_t$ and the variance $x_{t+1}' M_t x_{t+1} + \sigma_{t+1}^2 + x_{t+1}' W_{t+1} x_{t+1}$.

The initial component $p(y_1|x, z^*, \theta_z) = \int p(y_1|\beta_0, x, z^*, \theta_z) p(\beta_0|x, z^*, \theta_z) d\beta_0$ is computed by inserting the prior $p(\beta_0|v_0^2) = N(0, \text{diag}(v_0^2))$ and is normal with the mean of zero and

the variance $x_1'\text{diag}(v_0^2)x_1 + \sigma_1^2 + x_1'W_1x_1$.

E Tuning Metropolis-Hastings Steps

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

To illustrate the use of the adaptive optimal scaling method for the DT models, consider the example of drawing the rescaled latent variable z_t^* by an MH step. The proposal for its $i+1^{\text{th}}$ draw is a random walk $z_t^*(i+1) \sim N(z_t^*(i), w_i^2 A)$ 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 $i > i^*$[13]. i^* is a fixed threshold to avoid unstable sample covariance matrix of z_t^* when i is small. The scalar w_i^2 is updated according to $\log(w_{i+1}) = \log(w_i) + \frac{c}{d_i}(p_i - p^*)$ where p_i is the MH acceptance probability in the i^{th} draw of z_t^* and p^* is the target acceptance probability. The scalar c for updating w is determined as:

$$c = \frac{1}{Kp^*(1-p^*)} + \left(1 - \frac{1}{K}\right)\frac{\sqrt{2\pi}\exp\left(\frac{\alpha_w^2}{2}\right)}{2\alpha_w} \tag{E1}$$

where α_w satisfies $\Phi(-\alpha_w) = \frac{p^*}{2}$. As suggested in Garthwaite et al. (2016), the scalar d_i for updating w is set as $\max(\frac{i}{K}, d^*)$, where d^* is a fixed threshold to avoid that w converges before the sample covariance matrix of z_t^* stabilizes. The update of the scalar w is re-started whenever $\log(w)$ 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 w. In this paper, I set $p^* = 0.25$, $i^* = 100$ and $d^* = 200$. The configurations of adaptive MH steps for v, β_0 and c^* are similar to the example of z_t^*.

[13]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 ϵ is a small positive number (e.g. 1e-6).

F Alternative TVP Models

Three alternative TVP models with dynamic shrinkage features are compared to the proposed DT models in the applications of this paper.

- The RMI model that is a restricted version of the mixture innovation model of Giordani and Kohn (2008):

$$\begin{aligned} &y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ &\Delta\beta_{j,t} \sim N(0, v_j^2 d_{j,t}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2) \end{aligned} \tag{F1}$$

  where the individual mixture indicator $d_{j,t} \in \{0, 1\}$ is a Bernoulli variable. Let $d_t = [d_{1,t} \ ... \ d_{K,t}]'$ collects the individual mixture indicators. For computational practicality, the RMI model allows only 3 scenarios for d_t at each time t instead of the unrestricted 2^K scenarios: all $d_{j,t}$ equal zero (i.e. no parameter change at time t), all $d_{j,t}$ equal one (i.e. all parameters change at time t), and $d_{1,t} = 1$ and $d_{j,t} = 0$ for $j > 1$ (i.e. only the intercept changes at time t while other parameters remain constant). Denote the 3 scenarios as s_1, s_2 and s_3. The dynamics of d_t is Markovian following $p(d_t = s_j | d_{t-1} = s_j) = p_r$ and $p(d_t = s_i | d_{t-1} = s_j) = \frac{1-p_r}{2}$ for $i \neq j$. The prior for p_r is Beta$(50, 0.5)$ to favor persistent dynamics of d_t. Priors for v_j^2 and $v_{j,0}^2$ and the specification of σ_t^2 are the same as those in the DTL and DTS models.

- The DHS model of Kowal et al. (2019):

$$\begin{aligned} &y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ &\Delta\beta_{j,t} \sim N(0, \tau_{v,0}\tau_{v,j}\phi_{j,t}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2), \\ &\log(\phi_{j,t}) = \rho_j \log(\phi_{j,t-1}) + \xi_{j,t}, \quad \xi_{j,t} \sim Z(0.5, 0.5, 0, 1), \quad \phi_{j,0} = 1 \end{aligned} \tag{F2}$$

  where $\tau_{v,0} \sim IB(0.5, 0.5)$ and $\tau_{v,j} \sim IB(0.5, 0.5)$ for $j = 1, ..., K$. The logarithm of the local variance $\phi_{j,t}$ follows an autoregressive process with a Z-distributed innovation $\xi_{j,t}$ which is obtained as the logarithm of a inverted-beta distributed random variable. The Z distribution can be sampled as a scale mixture of normal distributions. See Kowal et al. (2019) for the motivation and details of the dynamic horseshoe model. The prior for ρ_j is $N(0.95, 1)I\{-1 < \rho_j < 1\}$. The Prior for $v_{j,0}^2$ and the specification of σ_t^2 are the same as those in the DTL and DTS models.

- The LMI model of He (2021) that closely approximates the mixture innovation model of Giordani and Kohn (2008) but is computationally more practical:

$$
\begin{aligned}
& y_t = x_t' \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
& \Delta \beta_{j,t} \sim N(0, v_j^2 d_{j,t}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2), \\
& d_{j,t} = \frac{1}{1 + \exp(-z_{j,t})}, \\
& \Delta z_{j,t} \sim N(0, a_j^2), \quad z_{j,0} \sim N(0, a_j^2 a_{j,0}^2)
\end{aligned} \tag{F3}
$$

  where priors for v_j^2, $v_{j,0}^2$ and the specification of σ_t^2, $a_{j,0}^2$ are the same as those in the DTL and DTS models. The parameter a_j has $N(0, 10)$ as its prior.

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.

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., G. Primiceri, and T. Sargent (2010). Inflation-gap persistence in the u.s. *American Economic Journal: Macroeconomics 2*, 43–69.

Cogley, T. and T. Sargent (2002). Evolving post-world war ii u.s. inflation dynamics. In B. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual*, Volume 16, pp. 331–373.

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.

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.

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 G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*, 216–230.

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.

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].

He, Z. (2021). A computationally efficient mixture innovation model for time-varying parameter regressions. SSRN: https://ssrn.com/abstract=4021201.

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.

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.

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.

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

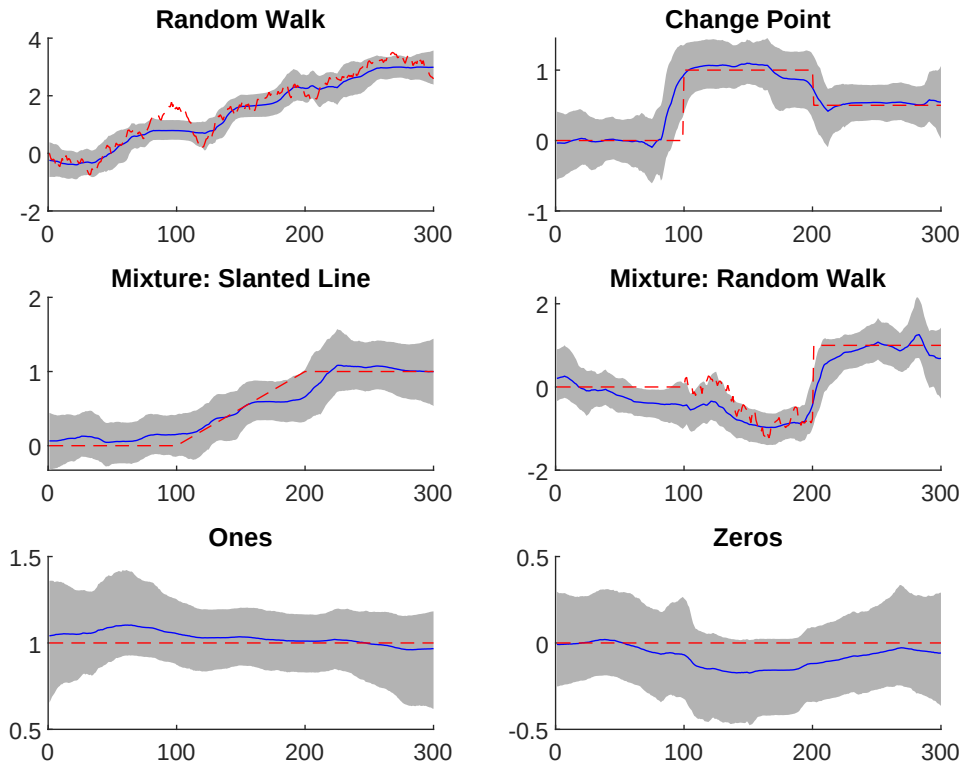
Tong, H. (1990). *Non-Linear Time Series: A Dynamical Systems Approach*. Oxford: Oxford University Press.

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.

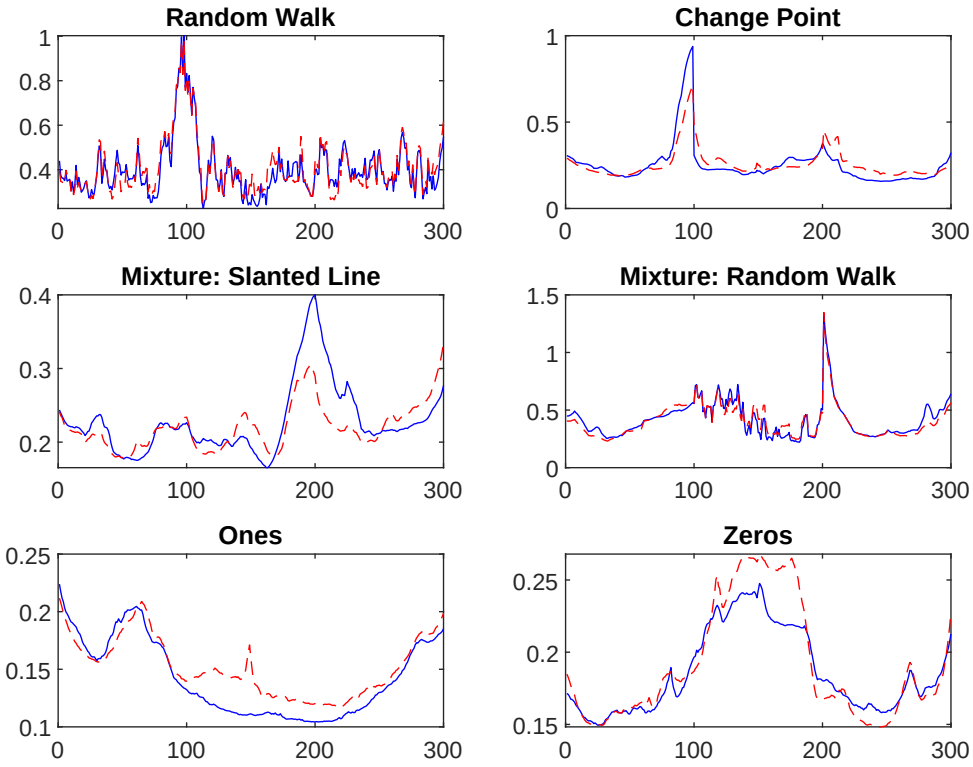
Yu, Y. and X. Meng (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.

Figure 1: Estimate of β_t: Simulation Study



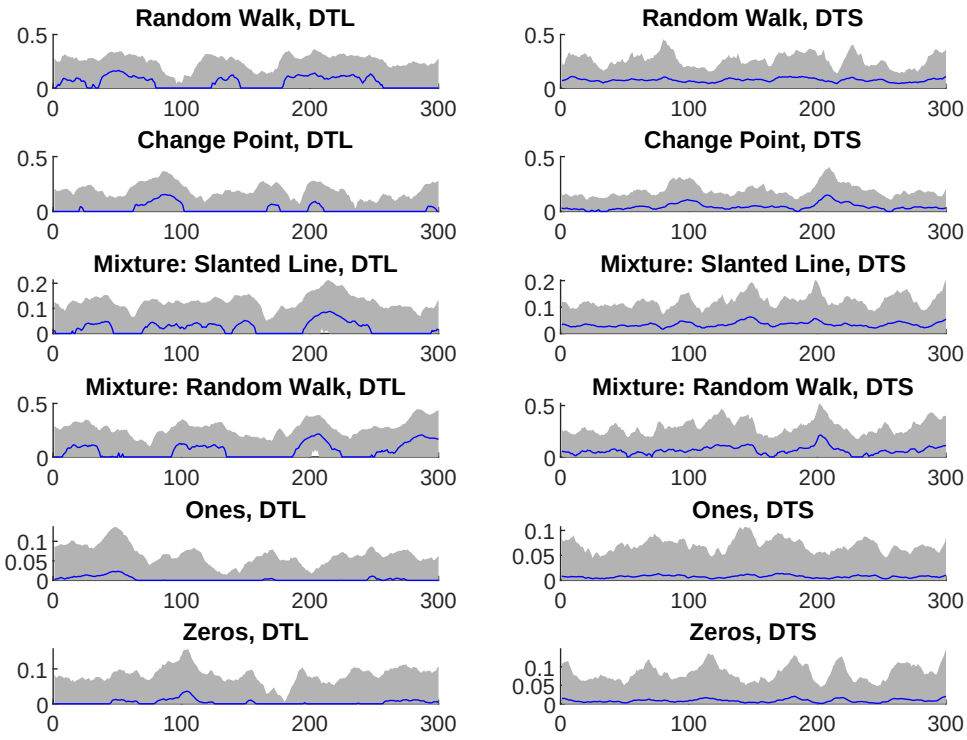
Note: The figure shows the point-wise posterior median (blue solid line) and 90% credible set (gray area) of estimated coefficients β_t for $t = 1, ..., 300$ by the DTL model. The red dashed line is the true coefficient.

Figure 2: Comparing Root Mean Squared Error of β_t Estimates: Simulation Study



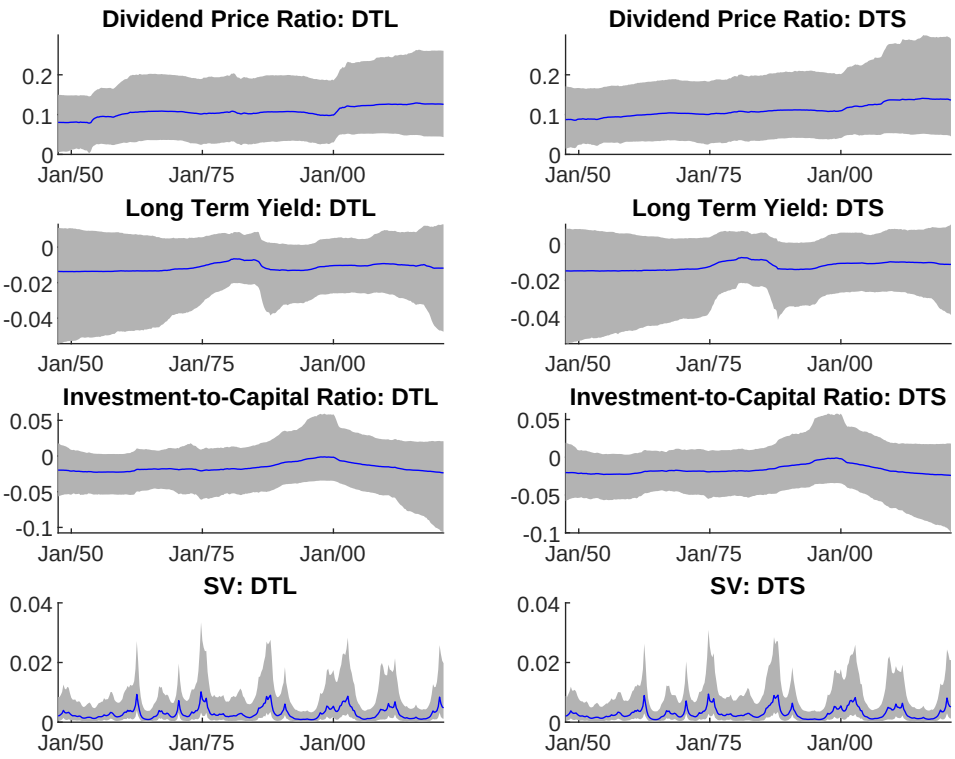
Note: The figure compares the point-wise root mean squared error (Equation (9)) of estimated coefficients β_t for $t = 1, ..., 300$ by the DTL (blue solid line) and DTS (red dashed line) models.

Figure 3: Conditional Standard Deviation Estimates: Simulation Study



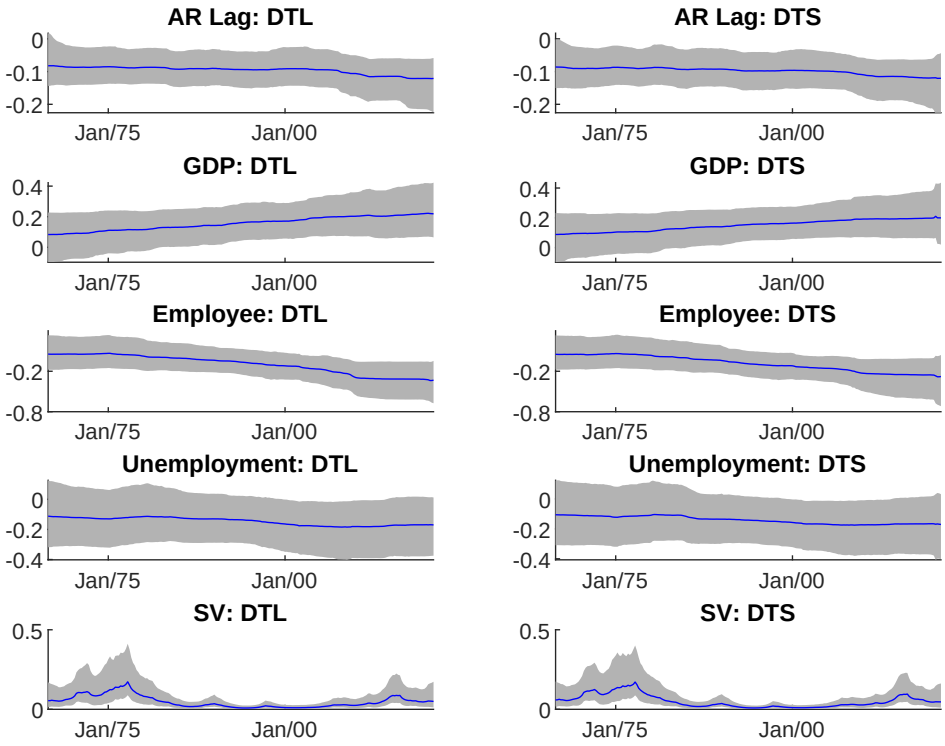
Note: The figure shows the point-wise posterior median (blue solid line) and 90% credible set (gray area) of estimated conditional standard deviations of β_t for $t = 1, ..., 300$ by the DTL (left panels) and DTS (right panels) models.

Figure 4: Estimates of Equity Premium Prediction



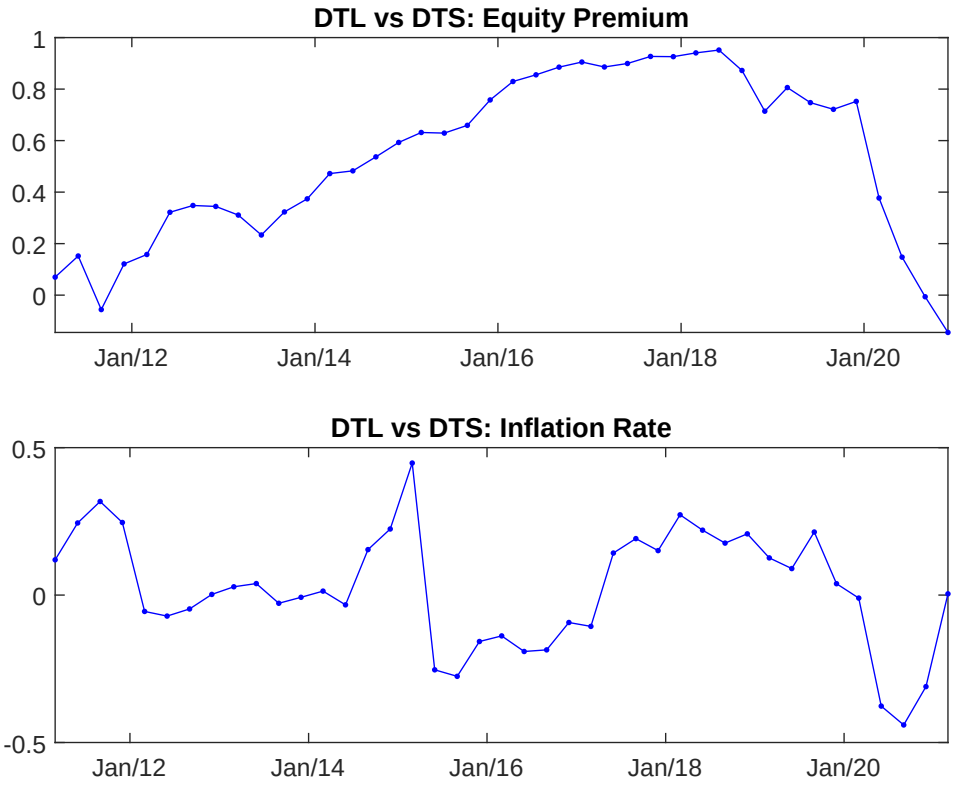
Note: The figure shows the point-wise posterior median (blue solid line) and 90% credible set (gray area) of estimated coefficient β_t by the DTL and DTS model for the 3 regressors in the equity premium prediction model that appear significant for most of the time periods (*dividend price ratio*, *long term yield*, *investment-to-capital ratio*). Description of the regressors can be found in Table 1. The last row in the figure plots the point-wise posterior median (blue solid line) and 90% credible set (gray area) of the conditional variance of the dependent variable by the DTL and DTS models.

Figure 5: Estimates of Inflation Rate Prediction



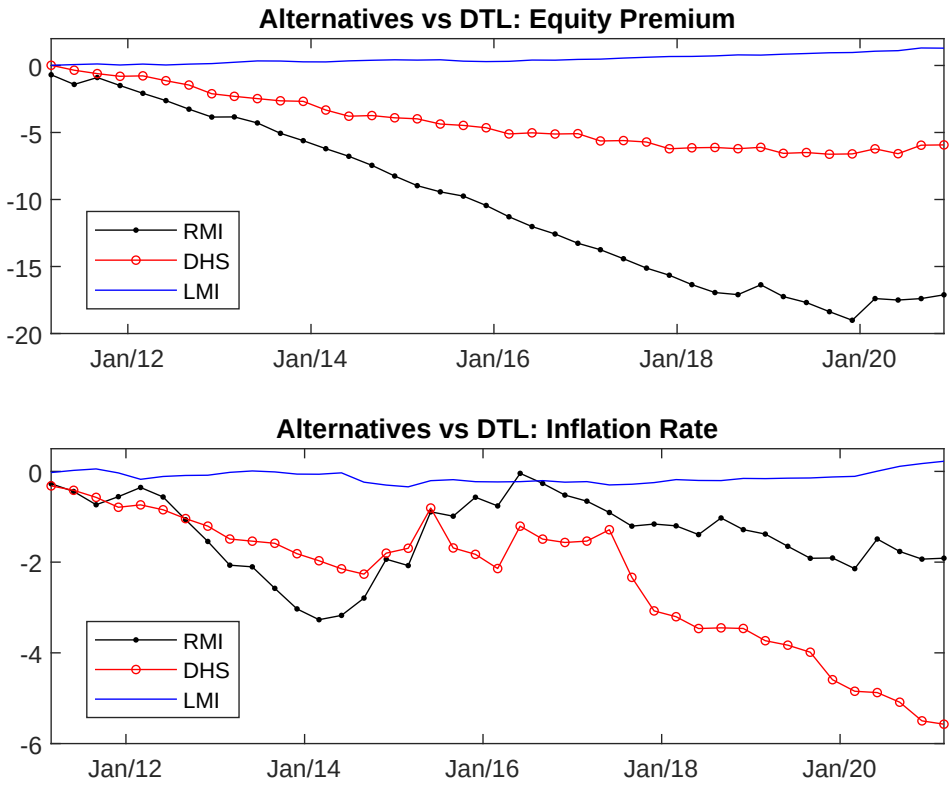
Note: The figure shows the point-wise posterior median (blue solid line) and 90% credible set (gray area) of estimated coefficient β_t by the DTL and DTS models for the 4 regressors in the inflation rate prediction model that are either consistently significant (*AR lag*) or are significant towards the end of the data sample (*GDP*, *employee*, *unemployment*). Description of the regressors can be found in Table 2. The last row in the figure plots the point-wise posterior median (blue solid line) and 90% credible set (gray area) of the conditional variance of the dependent variable by the DTL and DTS models.

Figure 6: Cumulative Log Predictive Bayes Factor: DTL versus DTS



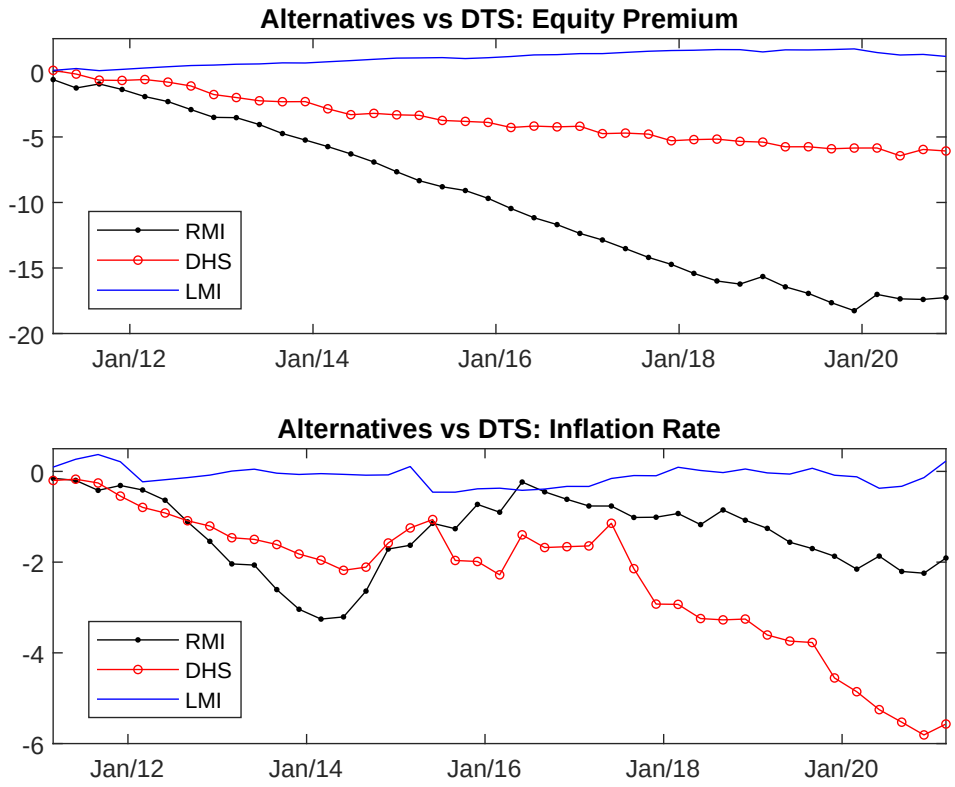
Note: The figure shows the cumulative log predictive Bayes factor of the DTL model over the DTS one in the equity premium prediction study (upper panel) and the inflation rate prediction study (lower panel) over their respective prediction sample. A positive value is in favor of the DTL model.

Figure 7: Cumulative Log Predictive Bayes Factor: Alternative TVP Models versus the DTL Model



Note: The figure shows the cumulative log predictive Bayes factors of 3 alternative TVP models over the DTL model in the equity premium prediction study (upper panel) and the inflation rate prediction study (lower panel) over their respective prediction sample. The alternative TVP models are the restricted mixture innovation (RMI) model, the dynamic horseshoe (DHS) model and the logistic mixture innovation (LMI) model. Descriptions of the alternative TVP models can be found in Appendix F. A positive value is in favor of the alternative TVP model.

Figure 8: Cumulative Log Predictive Bayes Factor: Alternative TVP Models versus the DTS Model



Note: The figure shows the cumulative log predictive Bayes factors of 3 alternative TVP models over the DTS model in the equity premium prediction study (upper panel) and the inflation rate prediction study (lower panel) over their respective prediction sample. The alternative TVP models are the restricted mixture innovation (RMI) model, the dynamic horseshoe (DHS) model and the logistic mixture innovation (LMI) model. Descriptions of the alternative TVP models can be found in Appendix F. A positive value is in favor of the alternative TVP model.

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	Sum of squared daily returns on the S&P500 index
Book-to-market ratio	Ratio of book to market value for the Dow Jones Industrial Average index
Net 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-to-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 Predictors for Inflation Rate

Name	Description
GDP	Log change of real GDP
Investment	Log change of real gross private domestic investment
Expenditure	Log change of real government consumption expenditures and gross investment
Imports	Log change of imports of goods and services
Potential GDP	Log change of real potential GDP
Employee	Log change of total non-farm employees
Unemployment	Change of unemployment rate
Wage	Log change of average hourly earnings of production and non-supervisory employees
House start	Log change of new privately-owned housing unites started
House supply	Change of the ratio of houses for sale to houses sold
Public debt	Change of the ratio of public debt to GDP
Consumer debt	Log change of consumer credit to households and non-profit organizations
Mortgage	Log change of one-to-four-family residential mortgages
Energy price	Log change of consumer price index for energy in U.S. city average
Producer price	Log change of producer price index for all commodities
Short rate	Change of 3-month Treasury bill rate
Term spread	Difference between 10-year Treasury constant maturity rate and 3-month Treasury bill rate
S&P500	Log change of average daily closing price
M1	Log change of M1 money stock
M2	Log change of M2 money stock

Note: Data on the S&P500 index is from Robert Shiller's website http://www.econ.yale.edu/ shiller/data.htm. Data on all other variables are from the FRED database of the U.S. Federal Reserve Bank of St. Louis.