# Time Dependent Shrinkage of Time-Varying Parameter Regression Models

Zhongfang He*

March 20, 2021

**Abstract**

This paper studies the time-varying parameter (TVP) regression model in which the regression coefficients are random walk latent states with time dependent conditional variances. This TVP model is flexible to accommodate a wide variety of time variation patterns but requires effective shrinkage on the state variances to avoid over-fitting. A Bayesian shrinkage prior is proposed based on reparameterization that translates the variance shrinkage problem into a variable shrinkage one in a conditionally linear regression with fixed coefficients. The proposed prior offers strong shrinkage for the state variances while maintaining the flexibility to accommodate local signals. A Bayesian estimation method is developed that avoids intractable large-scale matrix operations and employs component-wise ancilarity-sufficiency interweaving strategies to boost sampling efficiency. Simulation study and an empirical application to forecast inflation rate illustrate the benefits of the proposed approach.

*Keywords:* TVP, Bayesian shrinkage, MCMC, Horseshoe, ASIS
*JEL Codes:* C01, C11, C22, 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

The time-varying parameter (TVP) regression model is a linear regression model that allows the regression coefficients to vary over time and has been widely applied in econometric studies of time series data (Cogley and Sargent (2005), Primiceri (2005), Dangl and Halling (2012), Belmonte et al. (2014)). The commonly used configuration of the TVP model specifies the time varying coefficients as latent states that follow independent random walk processes with constant variances. Bitto and Fruhwirth-Schnatter (2019) is a recent contribution that shrinks the variances of the time varying coefficients and is able to automatically discriminate constant coefficients from time varying ones. However the inherent limitation in the assumption of homoskedastic latent states restricts the types of time variations this TVP model is able to accommodate. Timely adaptation to episodical combinations of constant and time-varying coefficients remains a challenge for the TVP model with homoskedastic latent states.

Recent developments in the literature has begun to allow time varying variances in the latent states of TVP models. By allowing the latent state variance to take different values from approximately zero to very large ones at each time $t$, both constant parameters, continuous and discrete parameter shifts as well as their episodical combinations could be accommodated in TVP models. To allow for heteroskedastic latent states, one strand of the literature directly applies Bayesian global-local shrinkage priors to the differenced latent states. The variance of each differenced latent state is specified as the product of a "global"' shrinkage parameter that remains constant and a "local"' shrinkage parameter that can vary over time. The global shrinkage parameter pushes the variance of the latent state towards zero and hence favors a constant coefficient, while the local shrinkage parameter adapts to local signals at each time $t$ and could result in a large overall variance at time $t$ to accommodate possible local parameter shifts. For example, Kowal et al. (2019) and Huber and Pfarrhofer (2021) contain TVP models where the horseshoe prior (Carvalho et al. (2010)) is applied to each differenced latent state. Kowal et al. (2019) further proposes a dynamic horseshoe prior that allows an autoregressive structure in the local shrinkage parameter of each latent state variance and shows good performance of the

resulting TVP models in capturing various patterns of parameter changes[1].

Shrinking the state variances towards zero amounts to the "variance selection" problem in the literature (Harvey (1989)). Pioneering work on the variance selection problem in the Bayesian framework includes Fruhwirth-Schnatter (2004), Fruhwirth-Schnatter and Wagner (2010), Nakajima and West (2013) and Kalli and Griffin (2014) etc. An attractive idea from Fruhwirth-Schnatter and Wagner (2010) is that by appropriate reparameterization of the original TVP model, the variance selection problem can be recast as a variable selection problem in a conditionally linear regression with the signed square root of the latent state variance being the constant coefficients. Hence the vast literature on variable selection by shrinkage priors can be borrowed on for efficient model estimation. I build on this insight from Fruhwirth-Schnatter and Wagner (2010) and proposes a new hierarchical prior to model the heteroskedasticity in the latent states of TVP models.

Rather than directly applying shrinkage priors to the time dependent state variances, I consider reparametrizing the random walk TVP model and apply horseshoe shrinkage priors to the signed square root of the state variances. In a nutshell, the resulting prior on each latent state variance is a gamma distribution in which the scale parameter is the product of a time-invariant global shrinkage parameter and a time-varying local shrinkage parameter. While the local shrinkage parameters follow independent inverted beta distributions as in the horseshoe prior, the global shrinkage parameter is a scale-mixture gamma distribution. I label this new shrinkage prior as a *gamma horseshoe prior*. I show that the gamma horseshoe prior results in a prior distribution of the state variances that allows more probability mass near zero than directly placing horseshoe priors on the state variances and hence offers stronger shrinkage while retaining a heavy tail for flexibility. Dynamic structure of the local shrinkage parameter as in Kowal et al. (2019) can be easily added to the gamma horseshoe prior to produce a dynamic version.

With the use of the horseshoe prior, an advantage of the proposed gamma horseshoe prior is that few user-specified hyper-parameter or manual intervention is needed and hence

[1]See Hauzenberger et al. (2020) for similar strategies for versions of TVP models where latent states follow independent Gaussian distributions rather than random walks. Kalli and Griffin (2014) is another example of dynamic shrinkage for a version of TVP models where the time-varying coefficients follow normal-gamma autoregressive processes.

allows for convenient automated model estimation. It is possible to use a standard Gibbs sampler to estimate the TVP model with the gamma horseshoe prior, though the resulting posterior draws are slow to converge and mix poorly. To boost sampling efficiency, the ancilarity-sufficiency interweaving strategy (ASIS) of Yu and Meng (2011) is adopted in this paper. A straightforward implementation of the ASIS, however, involves large-scale matrix operations within each MCMC iteration and would be intractable in realistic applications. I develop an alternative MCMC scheme that avoids repeated large-scale matrix operations and instead breaks the parameters of the state variances into blocks and apply the ASIS to each block in a component-wise fashion. The resulting sampler is efficient and computationally tractable.

For evaluating out-of-sample forecast performance, the Kalman filter is used conditional on simulations of relevant parameters to approximate the one-step-ahead predictive likelihood. A similar strategy was used in Bitto and Fruhwirth-Schnatter (2019) for the random walk TVP model with homoskedastic latent states and found that the resulting estimate of the one-step-ahead predictive likelihood tends to be more accurate than a pure simulation-based approximation.

The proposed approach is investigated in two applications. In a simulation study where the regression coefficients of the data generating process include both constants, continuous and discrete parameter shifts and their episodical combinations, the in-sample estimation accuracy of the coefficients under the gamma horseshoe priors is found to be comparable to that under the dynamic horseshoe prior of Kowal et al. (2019) and outperforms the horseshoe prior. In contrast to the horseshoe prior, the benefit of adding an autoregressive structure to the local parameters is diminished in the gamma horseshoe prior. The dynamic version of the gamma horseshoe prior leads to stronger shrinkage than the dynamic horseshoe prior while the static version of the gamma horseshoe prior provides the strongest shrinkage among these priors. It is also found that posterior draws from the gamma horseshoe priors mix better than those from the dynamic horseshoe prior.

In an empirical illustration, the TVP model is applied to forecast quarterly U.S. inflation rates. The regressors include autoregressive lags and lagged interest rate and unemployment rate. The relative strength of shrinkage provided by the gamma horseshoe and horseshoe

priors is similar to the findings in the simulation study. In particular, the gamma horseshoe prior leads to stronger shrinkage than both its dynamic version and the dynamic horseshoe prior. The predictive gain through the strong shrinkage by the gamma horseshoe prior is evident in the out-of-sample forecast comparison by cumulative log predictive likelihoods.

The remainder of the paper is organized as follows. Section 2 provides the details of the TVP model and the shrinkage priors. Estimation details are provided in Section 3. Section 4 and 5 present the simulation study and the empirical application respectively. Section 6 concludes. Additional details of the shrinkage priors and estimation algorithm are provided in appendices.

# 2 The Model

The TVP model under study is a linear regression model where the regression coefficients follow independent random walk processes:

$$\begin{aligned} y_t &= x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\ \beta_t &= \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \text{diag}(w_t)), \\ \beta_0 &\sim N(0, \text{diag}(w_0)) \end{aligned} \tag{1}$$

where $y_t$ is a scalar dependent variable, $x_t$ is a $K$-dimensional vector of regressors and $\beta_t$ is the corresponding time-varying coefficients for $t = 1, 2, ..., n$. The initial value $\beta_0$ plays the role of fixed regression coefficients[2]. For the moment, I will focus on the case of univariate time series. Extension to the multivariate case will be discussed in Section 2.3.

Viewed as a state space system, $\epsilon_t$ and $\eta_t$ are the Gaussian disturbances in the measurement and state equations respectively. For economic time series, the measurement equation disturbance $\epsilon_t$ is often allowed to have a time dependent variance. In empirical studies of this paper, I use the stochastic volatility (SV) specification for the measurement equation variance $\sigma_t^2$:

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

[2]To see this, let $\beta_t^*$ = $\beta_t$ - $\beta_0$. The TVP model can be rewritten as $y_t = x_t'\beta_0 + x_t'\beta_t^* + \epsilon_t$, $\beta_t^* = \beta_{t-1}^* + \eta_t$ and $\beta_0^* = 0$.

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

The state variance $w_t$ is allowed to be time dependent. Different types of time variations in the regression coefficients $\beta_t$ can be captured by varying the value of $w_t$. For example, a constant non-zero value $w_t = w$ leads to the commonly used specification of TVP models (Bitto and Fruhwirth-Schnatter (2019)). The change-point specification that assumes discrete shifts in the regression coefficients can be accommodated by setting $w_t > 0$ at the break points while $w_t = 0$ at other times. Constant regression coefficients can be accommodated by $w_t = 0$ for $t = 1, 2, ..., n$. As argued in Hauzenberger et al. (2020), time variations in TVP models often occur only episodically and only for a subset of the regression coefficients in economic time series data. Such empirical patterns amount to episodical combinations of constant and time-varying parameters, which can be easily accommodated in the TVP model of Equation (1).

## 2.1 Prior for State Variance

The key parameter in the TVP model of Equation (1) is the state variance $w_t$. In a Bayesian framework, the task is to develop a suitable prior for $w_t$ that should, on one hand, allow significant likelihoods for $w_t$ close to zero to reduce the risk of overfitting and, on the other hand, contain a heavy tail for $w_t$ to retrieve possible local signals of each time $t$. Recent developments in the field of Bayesian shrinkage priors provide an ideal toolkit for this task. In this paper, I focus on the horseshoe prior (Carvalho et al. (2010), Polson and Scott (2012)) that is computationally convenient and has shown good performance in many applications (Bhadra et al. (2019)).

A straightforward approach would be to directly impose a horseshoe prior on the differenced latent states $\Delta\beta_{j,t} = \beta_{j,t} - \beta_{j,t-1}$ for $j = 1, 2, ..., K$. That amounts to specify $w_{j,t} = r_j s_{j,t}$ with the "global"component $r_j$ and the "local" component $s_{j,t} \sim IB(0.5, 0.5)$, where IB denotes the inverted beta distribution[3]. This is the "static horseshoe" specifica-

[3]The density of an inverted beta distribution is $p(x) = \frac{x^{a-1}(1+x)^{-a-b}}{B(a,b)} I_{\{x>0\}}$ 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 $p(\sqrt{x}) = \frac{2a}{\pi(a^2+(\sqrt{x})^2)}$.

tion investigated in recent studies such as Kowal et al. (2019) and Huber and Pfarrhofer (2021).

This paper considers an alternative that translates such a variance shrinkage problem into a variable shrinkage problem in a (conditionally) linear regression model by reparameterization. A similar strategy has been adopted in Fruhwirth-Schnatter and Wagner (2010) and Bitto and Fruhwirth-Schnatter (2019) for the case of TVP models with homoskedastic latent states and has shown good performance. Let $\eta_t^* = \text{diag}(\frac{1}{\tilde{w}_t})\eta_t$ denote the normalized state disturbance where $\tilde{w}_t = \pm\sqrt{w_t}$ is the signed square root of the state variance $w_t$. Substituting the normalized state disturbance $\eta_t^*$ into the TVP model of Equation (1) gives:

$$\begin{aligned} y_t &= x_t'\beta_0 + \tilde{x}_t'\tilde{w} + \epsilon_t, \\ \eta_t^* &\sim N(0, I_K) \end{aligned} \tag{3}$$

where $\tilde{x}_t$ is a $nK$-by-1 vector with the first $tK$ elements being $x_t \odot \eta_1^*$, $x_t \odot \eta_2^*$, ..., $x_t \odot \eta_t^*$ and the remaining elements being zero. The notation $\odot$ denotes the element-wise product of matrices of the same size. $\tilde{w} = [\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_n]$ is a $nK$-by-1 vector stacking the signed square roots of the state variances $w_1$, ..., $w_n$. Equation (3) is the "non-centered" parametrization of the TVP model of Equation (1) in the spirit of Fruhwirth-Schnatter and Wagner (2010). Conditional on the normalized state disturbance $\eta_t^*$, the measurement equation in the reparametrized TVP model of Equation (3) becomes a linear regression model with the signed square roots $\tilde{w}$ being the fixed coefficients. Hence the task of specifying a shrinkage prior for the variance parameter $w_t$ is transformed into one of placing a shrinkage prior on fixed coefficients of a linear regression, which connects with the vast literature on variable selection in linear regression models. Indeed the horseshoe prior was originally proposed for shrinking fixed coefficients in a linear regression form.

With these considerations, I specify a hierarchical prior for the signed square roots $\tilde{w}$ as $\tilde{w}_{j,t}|v_j, d_{j,t} \sim N(0, v_j d_{j,t})$ with the *global parameter* $v_j$ and the *local parameter* $d_{j,t} \sim IB(0.5, 0.5)$ for $j = 1$, ..., $K$ and $t = 1$, ..., $n$. The resulting prior for the state variance is a gamma distribution $w_{j,t}|v_j, d_{j,t} \sim G(0.5, 2v_j d_{j,t})$[4].

[4]If $\pm\sqrt{x} \sim N(0, a)$, then $x \sim G(0.5, 2a)$ and vice versa, where the gamma distribution $G(\alpha, \beta)$ has the density $\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}\exp(-\frac{x}{\beta})$.

The remaining piece is to specify the prior for the global parameter $v_j$. Again I use the fact that the signed square root $\tilde{v}_j = \pm\sqrt{v_j}$ is the fixed coefficient in a conditionally linear regression model by reparametrizing the TVP model of Equation (1):

$$\begin{aligned} y_t &= x_t'\beta_0 + (x_t \odot \beta_t^*)'\tilde{v} + \epsilon_t, \\ \beta_t^* &\sim N(\beta_{t-1}^*, \text{diag}(\phi_t)), \\ \beta_0^* &= 0 \end{aligned} \tag{4}$$

where the transformed latent state is $\beta_{j,t}^* = (\beta_{j,t} - \beta_{j,0})/\tilde{v}_j$, the *scaled state variance* is $\phi_{j,t} = \frac{w_{j,t}}{v_j}$, $\beta_t^* = [\beta_{1,t}^*, \beta_{2,t}^*, ..., \beta_{K,t}^*]$, $\phi_t = [\phi_{1,t}, \phi_{2,t}, ..., \phi_{K,t}]$ and $\tilde{v} = [\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_K]$. A horseshoe prior for the fixed coefficients $\tilde{v}$ results in $\tilde{v}_j|\tau_0, \tau_j \sim N(0, \tau_0\tau_j)$ with the common component $\tau_0 \sim IB(0.5, 0.5)$ and the individual component $\tau_j \sim IB(0.5, 0.5)$. The equivalent prior for the global parameter $v_j = \tilde{v}_j^2$ is $v_j|\tau_0, \tau_j \sim G(0.5, 2\tau_0\tau_j)$ for $j = 1, ..., K$.

In summary, the proposed prior for the state variance $w_t$ is $w_{j,t}|v_j, d_{j,t} \sim G(0.5, 2v_j d_{j,t})$ with $d_{j,t} \sim IB(0.5, 0.5)$, $v_j|\tau_0, \tau_j \sim G(0.5, 2\tau_0\tau_j)$, $\tau_0 \sim IB(0.5, 0.5)$ and $\tau_j \sim IB(0.5, 0.5)$ for $j = 1, ..., K$ and $t = 1, ..., n$. I label this prior as a *gamma horseshoe* (GHS) prior since the local parameter $d_{j,t}$ in the prior for the signed square root $\tilde{w}_{j,t}$ of the state variance follows an inverted beta distribution as in a horseshoe prior while the global parameter $v_j$ follows a gamma distribution conditionally. Appendix A contains a comparison of the proposed gamma horseshoe prior with the horseshoe one. It is found that the gamma horseshoe prior has a more concentrated mass around the point of zero than the horseshoe prior while containing a heavy tail.

A dynamic version of the gamma horseshoe prior in the vein of Kowal et al. (2019) can be easily incorporated by specifying $d_{j,t} = d_{j,t-1}^{\rho_j} e_{j,t}$ where $d_{j,0} = 1$, $e_{j,t} \sim IB(0.5, 0.5)$ captures the serially uncorrelated component of $d_{j,t}$ and $\rho_j$ is the autoregressive coefficient. I will refer to the dynamic version as the dynamic gamma horseshoe (DGHS) prior to differentiate from the static one with $\rho_j = 0$.

Using the framework of Equation (1), Table 1 compares the priors for the state variances that appear in the literature for random-walk TVP models with heteroskedastic latent states. There are different ways to present the priors. I choose to write them in a common format in Table 1 to facilitate comparison. Relative to the horseshoe (HS) and dynamic horseshoe (DHS) priors, the proposed GHS and DGHS priors add an extra layer of gamma

distributions for both the global and local parameters and hence offers greater flexibility.

## 2.2 Prior for Other Model Parameters

Besides the state variance $w_t$, other parameters in the TVP model of Equation (1) include the measurement equation variance $\sigma_t^2$ and the initial state $\beta_0$.

For the SV specification of the measurement equation variance $\sigma_t^2$ (Equation (2)), the priors are $\mu \sim N(0, 10)$, $\rho \sim N(0.95, 0.04)I_{\{-1<\rho<1\}}$. Note that an informative prior is used for the autoregressive coefficient $\rho$ to reflect the empirical regularity that volatilities of economic time series tend to be highly persistent. For the variance parameter $\sigma_y^2$ in the SV model, I follow Kastner and Fruhwirth-Schnatter (2014) and specify a Gaussian prior for its signed square root $\sigma_y = \pm\sqrt{\sigma_y^2}$ as $\sigma_y|s_y \sim N(0, s_y)$ to facilitate estimation. It is possible to specify a fixed number (e.g. 10) for the variance $s_y$ in the prior of $\sigma_y$. In this paper, I use a prior $s_y \sim IB(0.5, 0.5)$ to determine $s_y$ in a data-driven way.

In the case where the measurement equation disturbance is homoskedastic $\sigma_t^2 = \sigma^2$, the Jeffery's prior $\sigma^2 \propto \frac{1}{\sigma^2}$ is used.

For the initial state $\beta_0$, a horseshoe prior is imposed to push insignificant elements towards zero. Specifically $\beta_{j,0}|\tau_{0,0}, \tau_{j,0} \sim N(0, \tau_{0,0}\tau_{j,0})$ with $\tau_{0,0} \sim IB(0.5, 0.5)$ and $\tau_{j,0} \sim IB(0.5, 0.5)$ for $j = 1, ..., K$.

## 2.3 Extension to Multivariate Time Series

The univariate TVP model of Equation (1) can be extended to the multivariate case in a straightforward way. The challenging part is how to handle the resulting covariance matrix of the measurement equation disturbance if the covariance matrix is time dependent and contains non-zero off-diagonal elements. A computationally convenient solution is offered by the time-varying Cholesky decomposition introduced in Lopes et al. (2018) and Carriero et al. (2019).

Suppose that the measurement equation of the multivariate TVP model is $y_t = X_t\beta_t + \epsilon_t$ with $\epsilon_t \sim N(0, \Sigma_t)$, where $y_t$ is a $P$-dimensional vector of the dependent variables, $X_t$ is a $P$-by-$K$ matrix of the regressors, $\beta_t$ is a $K$-by-1 vector of the regression coefficients and $\Sigma_t$ is the covariance matrix of the disturbance $\epsilon_t$.

The time-varying Cholesky decomposition factorizes the covariance matrix as $\Sigma_t = A_t^{-1} H_t (A_t')^{-1}$ where $A_t^{-1}$ is a lower triangular matrix with ones in the diagonal and $H_t$ is a diagonal matrix with each diagonal elements following an independent SV process. Multiplying $A_t$ to both sides of the measure equation leads to $A_t y_t = A_t x_t \beta_t + A_t \epsilon_t$ and $A_t \epsilon_t \sim N(0, H_t)$. By unfolding the elements in matrix products $A_t y_t$ and $A_t x_t \beta_t$, one obtains a system of $P$ independent linear regression models with univariate dependent variables and time-varying coefficients. The univariate TVP model of Equation (1) can be then applied independently to each of the $P$ regressions[5].

## 2.4 Comparison with The Mixture Innovation Model

The TVP model of Equation (1) with heteroskedastic latent states resembles the mixture innovation model of Giordani and Kohn (2008) where the state variance is also allowed to be time dependent. The key difference is that the mixture innovation model amounts to specify a discrete "spike-and-slab" mixture for the state variance while the gamma horseshoe prior leads to a continuous mixture. Using the notation of Equation (1), a mixture innovation model would specify $w_{j,t} \sim \kappa_{j,t}\delta_0 + (1 - \kappa_{j,t}) f(w_{j,t})$ where $\delta_0$ denotes the point mass at zero and $f(\cdot)$ denotes a distribution with positive support (e.g. a gamma or inverse gamma distribution) or simply a point mass at a user-selected positive value. The indicator variable $\kappa_{j,t} \in \{0, 1\}$ determines if a change occurs in the time varying parameter $j$ at time $t$ and could either be specified as an independent Bernoulli variable or be driven by a Markov process.

The main hurdle of the mixture innovation model is the combinatorial complexity to jointly sample the mixture indicators $\kappa_{j,t}$. Gerlach et al. (2000) and Giordani and Kohn (2008) develop Gibbs samplers for the mixture indicators that integrate out the latent states but sample the mixture indicators in an element-by-element fashion. As explained in Carvalho et al. (2010) and Bhadra et al. (2017), the horseshoe prior offers substantial computational advantage over the discrete mixture approach for posterior exploration.

[5]A subtle issue of the time-varying Cholesky decomposition is the "prior ordering" problem described in Carriero et al. (2019). Since priors are placed on the parameters under a triangular decomposition, changing the ordering of the dependent variables could imply different priors on the original model parameters which in turn could affect posterior exploration. More details can be found in Carriero et al. (2019).

Being continuous mixtures, the computational efficiency of the horseshoe prior extends to the gamma horseshoe prior.

# 3 Estimation

Given the priors described in Section 2, developing a standard Gibbs sampler for the TVP model of Equation (1) is straightforward. However the resulting posterior draws suffer from slow convergence and poor mixing, particularly for the state variances and associated hyper-parameters. In this paper, I adopt the ancillarity-sufficiency interweaving strategy (ASIS) of Yu and Meng (2011) to boost the efficiency of the standard Gibbs sampler. The ASIS boosting strategy has shown good performance in many applications of state space models (e.g. Kastner and Fruhwirth-Schnatter (2014), Simpson et al. (2017) and Bitto and Fruhwirth-Schnatter (2019)). A discussion of applying the ASIS is provided in Section 3.1. The details of the proposed Gibbs sampler are presented in Section 3.2.

## 3.1 ASIS Boosting

The ASIS considers two parametrizations of a hierarchical latent variable model: a sufficient augmentation (SA) in which the latent variable is a sufficient statistic for the relevant parameters, and an ancillary augmentation (AA) in which the same parameters are all in the distribution of the observation and hence the latent variable is an ancillary statistic for them. By moving between a SA-AA pair, Yu and Meng (2011) found that the resulting Markov chain Monte Carlo (MCMC) algorithm is often more efficient and, under certain conditions, provides the fastest converging algorithm within a general class of data augmentation schemes.

The ASIS can be implemented in a component-wise fashion. That is, a *conditional* SA-AA pair for a subset of model parameters can be formed by treating other parameters as fixed, while the conditional SA-AA pair can be different for different subsets of model parameters. Hence the ASIS is quite flexible for use in practical applications.

For the TVP model of Equation (1), the key to implementing the ASIS is to find a suitable (conditional) SA-AA pair or SA-AA pairs of model parametrizations for the

parameters that are difficult to estimate in a vanilla Gibbs sampler, namely, the state variances and associated hyper-parameters.

Return to the TVP model of Equation (1). It is clear that the latent state $\beta_t$ is a sufficient statistic for the state variance $w_t$ and for that matter the signed square root $\tilde{w}_t$. Since the reparametrization of Equation (3) moves the signed square roots $\tilde{w}$ to the measurement equation and leaves the new latent state $\eta_t^*$ free of $\tilde{w}_t$, one may be tempted to directly apply the ASIS to the signed square roots $\tilde{w}$ by treating Equation (1) and Equation (3) as a SA-AA pair. The drawback of this strategy however is that drawing the signed square roots $\tilde{w}$ in Equation (3) will be working with an enormous linear regression with $nK$ regressors. Even with the state-of-art algorithm in Bhattacharya et al. (2016), the computational complexity is in the order of $\mathcal{O}(n^3K)$ in each MCMC iteration and could be overwhelming in typical economic studies with hundreds of data points[6].

This paper proposes an alternative strategy. Instead of directly boosting the state variance, two separate component-wise ASISs are applied to the global parameter $v_j$ and the local parameter $d_{j,t}$ in the prior of the state variance.

Recall from Equation (4) the auxiliary variable of the scaled state variance $\phi_{j,t} = \frac{w_{j,t}}{v_j}$. It follows $w_{j,t} = v_j\phi_{j,t}$ with $\phi_{j,t}|v_j, d_{j,t} \sim G(0.5, 2d_{j,t})$. Conditional on the scaled state variance $\phi_{j,t}$, the latent state $\beta_{j,t}$ is a sufficient statistic for the global parameter $v_j$ and hence the TVP model of Equation (1) can be viewed as a SA representation for $v_j$. On the other hand, the reparametrization of Equation (4) has the signed squared root $\tilde{v}_j$ of the global parameter $v_j$ in its measurement equation, while its state equation $\beta_{j,t}^* \sim N(\beta_{j,t-1}^*, \phi_{j,t})$ is conditionally free of the global parameter $v_j$. Therefore the reparametrization of Equation (4) constitutes a conditional AA representation for $v_j$. An ASIS boosting for the global parameter $v_j$ can be conducted based on this conditional SA-AA pair.

For the local parameter $d_{j,t}$, consider the subset of Equation (4) along with the conditional distribution of the scaled state variance $\phi_{j,t}$:

$$\begin{aligned} \Delta\beta_{j,t}^* &= \beta_{j,t}^* - \beta_{j,t-1}^* \sim N(0, \phi_{j,t}) \\ \phi_{j,t}|v_j, d_{j,t} &\sim G(0.5, 2d_{j,t}) \end{aligned} \tag{5}$$

[6]See Johndrow et al. (2020) and Hauzenberger et al. (2020) for methods that use approximations to the exact algorithm of Bhattacharya et al. (2016).

Equation (5) can be viewed as a nested latent variable model with $\Delta\beta^*_{j,t}$ being the observation, $\phi_{j,t}$ being the latent variable and $d_{j,t}$ being the parameter of interest. In this nested model, the latent variance $\phi_{j,t}$ is a sufficient statistic for $d_{j,t}$ while the observation $\Delta\beta^*_{j,t}$ is independent of $d_{j,t}$ conditional on $\phi_{j,t}$. Hence Equation (5) constitutes a conditional SA representation for $d_{j,t}$. Now let $\phi^*_{j,t} = \frac{\phi_{j,t}}{d_{j,t}}$. It follows that the conditional distribution $\phi^*_{j,t}|v_j, d_{j,t} \sim G(0.5, 2)$ is free of $d_{j,t}$. Substituting $\phi_{j,t} = \phi^*_{j,t} d_{j,t}$ into Equation (5) gives a conditional AA representation for $d_{j,t}$:

$$\begin{aligned} \Delta\beta^*_{j,t} &\sim N(0, \phi^*_{j,t} d_{j,t}) \\ \phi^*_{j,t}|v_j, d_{j,t} &\sim G(0.5, 2) \end{aligned} \tag{6}$$

A second ASIS boosting can be applied to this conditional SA-AA pair for the local parameter $d_{j,t}$.

## 3.2 MCMC Estimation

The proposed Gibbs sampler divides the model parameters into the following blocks:

1. Measurement equation variance:
   (a) SV specification: $\sigma_t^2$, $\mu$, $\rho$, $\sigma_y^2$ and associated hyper-parameters including $s_y$,
   (b) Homoskedastic specification: $\sigma^2$,
2. Hyper-parameters $\{\tau_0, \tau_j\}$ and $\{\tau_{0,0}, \tau_{j,0}\}$ for the global parameter $v_j$ and the initial state $\beta_{j,0}$ respectively,
3. Hyper-parameters for the local parameter $d_{j,t}$,
4. Transformed latent states $\beta^*_{j,t}$, global parameter $v_j$ and initial state $\beta_0$,
5. Scaled state variance $\phi_{j,t}$ and local parameter $d_{j,t}$,

where $j = 1, 2, ..., K$ and $t = 1, 2, ..., n$.

In block 1, sampling of the measurement equation variance in the SV specification is similar to Kastner and Fruhwirth-Schnatter (2014) by adopting the log linearization strategy of Kim et al. (1998) and Omori et al. (2007). The details are given in Appendix

B. In the homoskedastic case $\sigma_t^2 = \sigma^2$ with the Jeffery's prior $\sigma^2 \propto \frac{1}{\sigma^2}$, the posterior is $\sigma^2|y, x, \beta \sim \text{IG}\left(\frac{n}{2}, \frac{1}{2}\sum_{t=1}^{n}\epsilon_t^2\right)$ where $\epsilon_t = y_t - x_t'\beta_t$, $y = [y_1, ..., y_n]$, $x = [x_1, ..., x_n]$, $\beta = [\beta_1, ..., \beta_n]$ and IG denotes the inverse gamma distribution.

Block 2 is the hyper-parameters for two horseshoe priors. By adopting the hierarchical inverse gamma representation of Makalic and Schmidt (2016), the posteriors of these hyper-parameters are inverse gamma distributions. Appendix C provides the details.

Block 3 contains the hyper-parameters for the local parameter $d_{j,t}$. In the static case, an auxiliary variable $e_{j,t}$ is introduced to represent $d_{j,t} \sim IB(0.5, 0.5)$ as a hierarchical inverse gamma distribution $d_{j,t} \sim IG\left(0.5, \frac{1}{e_{j,t}}\right)$ and $e_{j,t} \sim IG(0.5, 1)$(Makalic and Schmidt (2016)). The posterior for the auxiliary variable is $e_{j,t}|d_{j,t} \sim IG\left(1, 1 + \frac{1}{d_{j,t}}\right)$. In the dynamic case, use the log linearization $\log(d_{j,t}) = \rho_j \log(d_{j,t-1}) + \log(\xi_{j,t})$ where $\xi_{j,t} \sim IB(0.5, 0.5)$. Following Polson et al. (2013) and Kowal et al. (2019), $\log(\xi_{j,t})$ can be represented as a precision mixture of normals $\log(\xi_{j,t}) \sim N\left(0, \frac{1}{e_{j,t}}\right)$, where $e_{j,t} \sim PG(1, 0)$ is an auxiliary variable and $PG$ denotes the polya-gamma distribution (Barndorff-Nielsen et al. (1982)). The posteriors of the parameters $\rho_j$ and $e_{j,t}$ are derived in Appendix D.

Block 4 and 5 implement two component-wise ASISs. For block 4, the steps are as follows:

i Conditional on $\phi_t$, $\beta_0$, $\tilde{v}$ and $\sigma_t^2$, draw the transformed latent states $\beta_t^*$ from the AA represention (Equation (4)) by the simulation smoother of Durbin and Koopman (2002).

ii Let $\alpha$ be a $2K$-by-1 vector stacking $\beta_0$ and $\tilde{v}$. Conditional on $\beta_t^*$, $\sigma_t^2$ and the hyper-parameters in block 2, draw $\alpha$ from a linear regression with the posterior $N(b, B)$, where $B^{-1} = B_0^{-1} + \sum_{t=1}^{n} \frac{1}{\sigma_t^2}\tilde{x}_t\tilde{x}_t'$, $B^{-1}b = \sum_{t=1}^{n} \frac{1}{\sigma_t^2}\tilde{x}_t y_t$, $B_0$ is a diagonal matrix with the diagonal elements $\tau_{0,0}\tau_{1,0}$, ..., $\tau_{0,0}\tau_{K,0}$, $\tau_0\tau_1$, ..., $\tau_0\tau_K$, and $\tilde{x}_t = [x_t, x_t \odot \beta_t^*]'$.

iii Keep the sign of $\tilde{v}$. Calculate the latent state $\beta_{j,t} = \beta_{j,0} + \tilde{v}_j\beta_{j,t}^*$ of the SA representation (Equation (1)).

iv Update $v$ and $\beta_0$ in the SA representation of Equation (1):

$$v_j|\tau_0,\tau_j,\phi_{j,t},\beta_{j,t},\beta_{j,0} \sim GIG\left(\frac{1-n}{2},\frac{1}{\tau_0\tau_j},\sum_{t=1}^{n}\frac{1}{\phi_{j,t}}\left(\Delta\beta_{j,t}\right)^2\right)$$

$$\beta_{j,0}|\beta_{j,1},v_j,\tau_{0,0},\tau_{j,0},\phi_{j,1} \sim N\left(\frac{\tau_{0,0}\tau_{j,0}\beta_{j,1}}{\tau_{0,0}\tau_{j,0}+v_j\phi_{j,1}},\frac{\tau_{0,0}\tau_{j,0}v_j\phi_{j,1}}{\tau_{0,0}\tau_{j,0}+v_j\phi_{j,1}}\right)$$

where GIG denotes the generalized inverse Gaussian distribution[7].

v Update $\tilde{v}_j$ as the positive square root of $v_j$ times the sign from Step iii. Update the transformed latent state $\beta^*_{j,t}=\frac{\beta_{j,t}-\beta_{j,0}}{\tilde{v}_j}$.

Sampling the parameters of block 5 has the following steps:

i Draw the latent variable $\phi_{j,t}$ from the SA representation of Equation (5):

$$\phi_{j,t}|d_{j,t},\beta^*_{j,t} \sim GIG\left(0,\frac{1}{d_{j,t}},\left(\Delta\beta^*_{j,t}\right)^2\right)$$

ii Based on the SA representation of Equation (5), draw $d_{j,t}$ from the following posterior in the static case:

$$d_{j,t}|e_{j,t},\phi_{j,t} \sim IG\left(1,\frac{1}{e_{j,t}}+\frac{\phi_{j,t}}{2}\right)$$

In the dynamic case, draw $d_{j,t}$ based on the algorithm in Kowal (2019). The key is to treat $d_{j,t}$ as the stochastic volatility of a SV model and use the log linearization strategy of Kim et al. (1998) and Omori et al. (2007) to sample $\log(d_{j,t})$. The details are given in Appendix D.

iii Calculate the transformed latent variable $\phi^*_{j,t}=\frac{\phi_{j,t}}{d_{j,t}}$ in the AA representation of Equation (6).

iv Based on the AA representation of Equation (6), update

$$d_{j,t}|e_{j,t},\phi^*_{j,t},\beta^*_{j,t} \sim IG\left(1,\frac{1}{e_{j,t}}+\frac{1}{2\phi^*_{j,t}}\left(\Delta\beta^*_{j,t}\right)^2\right)$$

in the static case. In the dynamic case, update $d_{j,t}$ based on the algorithm in Kowal et al. (2019) with the details in Appendix D.

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

v Update the latent variable $\phi_{j,t} = \phi^*_{j,t} d_{j,t}$.

This completes the description of the MCMC algorithm for estimating the TVP model with the proposed gamma horseshoe priors. It should be noted that the proposed algorithm mostly samples from closed-form distributions and requires neither manual intervention nor difficult user-specified hyper-parameters. Thus model estimation can be performed in an automated fashion and facilitates practical applications.

## 3.3 Predictive Likelihoods

In this paper, the predictive likelihood is used to compare different prior specifications of TVP models. See Geweke and Amisano (2010) for a reviews of Bayesian predictive analysis. Specifically, given a model specification $\mathcal{M}$ and a data sample of the dependent variable $y^n = \{y_t\}_{t=1}^n$ and regressors $x^{n+1} = \{x_t\}_{t=1}^{n+1}$, the one-step-ahead predictive likelihood is $p(y_{n+1}|y^n, x^{n+1}, \mathcal{M})$ that integrates out all the parameters in model $\mathcal{M}$. For expositional convenience, the regressors in $x_t$ are treated as exogenous. It should be understood that in the case of predictive regressions, $x_t$ are variables observed at time $t-1$.

To compute the one-step-ahead predictive likelihood $p(y_{n+1}|y^n, x^{n+1}, \mathcal{M})$, I use the strategy that is labeled as the *conditionally optimal Kalman mixture approximation* in Bitto and Fruhwirth-Schnatter (2019). Let $\theta^n = \{\theta_t\}_{t=1}^n$ denote the parameters in the TVP model of Equation (1) over time points $t = 1, ..., n$. It is straightforward to derive the conditional distribution:

$$y_{n+1}|x_{n+1}, w_{n+1}, \sigma^2_{n+1}, b_n, B_n \sim N\left(x'_{n+1} b_n,\ \sigma^2_{n+1} + x'_{n+1}\left(B_n + \text{diag}(w_{n+1})\right) x_{n+1}\right) \tag{7}$$

where $b_n$ and $B_n$ are the mean and covariance matrix of the filtering distribution $\beta_n|y^n, x^n, \theta^n \sim N(b_n, B_n)$. One can approximate the one-step-ahead predictive likelihood as:

$$p\left(y_{n+1}|y^n, x^{n+1}, \mathcal{M}\right) \approx \frac{1}{M}\sum_{i=1}^{M} p\left(y_{n+1}|x_{n+1}, w^{(i)}_{n+1}, \left(\sigma^2_{n+1}\right)^{(i)}, b^{(i)}_n, B^{(i)}_n\right) \tag{8}$$

where $w^{(i)}_{n+1}$ and $\left(\sigma^2_{n+1}\right)^{(i)}$ are simulated based on posterior draws of the gamma horseshoe prior and the SV model for the measurement equation disturbance. The filtered mean $b^{(i)}_n$ and filtered covariance matrix $B^{(i)}_n$ are computed based on a Kalman filter with their derivations provided in Appendix E.

Given the one-step-ahead predictive likelihood $p(y_t|y^{t-1}, x^t, \mathcal{M})$, the cumulative log predictive likelihood $\sum_{i=1}^{m} \log p(y_{n+i}|y^{n+i-1}, x^{n+i}, \mathcal{M})$ over a prediction sample $t = n + 1, ..., n + m$ is the criterion for comparing the performance of model $\mathcal{M}$ with alternatives.

# 4 Simulation Study

A sample of 300 data points are simulated from a linear regression model with 6 coefficients exhibiting different types of time variation over $t = 1, ..., 300$:

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

2 Change point: $\beta_{2,t} = I_{\{100<t\leq 200\}} - I_{\{t>200\}}$.

3 Mixture of constant parameter, random walk and change point:

$$\beta_{3,t} = \left( \sum_{j=1}^{t} u_j \right) I_{\{100<t\leq 200\}} + I_{\{t>200\}}$$

with $u_j \sim N(0, 0.01)$.

4 Ones: $\beta_{4,t} = 1$.

5 Small constant: $\beta_{5,t} = 0.1$.

6 Zeros: $\beta_{6,t} = 0$.

The regressors are from standard normal distributions. The dependent variables is generated by adding a noise from $N(0, \sigma^2)$ where $\sigma^2$ is calibrated such that the ratio of $\sigma^2$ to the variance of the dependent variable is $0.2^8$. The 4 TVP priors in Table 1 are applied to the simulated data. In estimation, the measurement equation disturbance is taken to be homoskedastic. A sample of 10,000 posterior draws are kept for analysis after a burn-in of length 5,000.

[8] In experiments, another two sets of dependent variables are simulated where the ratio of $\sigma^2$ to the variance of the dependent variable is 0.5 and 0.8 respectively. The estimation results are qualitatively similar.

To compare the accuracy of coefficient estimates from the TVP priors, I compute the point-wise root mean square error (RMSE) for each coefficient $j$ at point $t$:

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

where $\hat{\beta}_{j,t}^{(i)}$ denotes a posterior draw of the coefficient from a given model and $\beta_{j,t}$ the true coefficient value. Box plots of the distributions of $\text{RMSE}_{j,t}$ across $t = 1, ..., 300$ for each coefficient $j = 1, ..., 6$ are shown in Figure 2. A concentrated distribution of $\text{RMSE}_{j,t}$ implies a stable level of estimation precision across points $t$ and hence is more desirable.

In Figure 2, the RMSEs from the HS prior clearly lag behind those from the other 3 priors. So I focus on comparing the two gamma horseshoe priors and the DHS prior. The DGHS and DHS priors show very similar levels of estimation precision. Interestingly the RMSEs of the GHS prior are comparable to those of the DGHS and DHS priors except in the case of the change-point coefficient where the dynamic priors tend to perform slightly better, suggesting that the stronger shrinkage from the GHS prior reduces the need for a dynamic structure in the state variance.

To compare the estimated shrinkage levels between the 4 TVP priors, Figure 3 shows their point-wise posterior medians and 90% credible sets of the log state variance $\log(w_{j,t})$ for $j = 1, ..., 6$ and $t = 1, ..., 300$. Estimates of $\log(w_{j,t})$ from the HS prior are clearly larger than those from the other 3 priors. Shrinkage on $\log(w_{j,t})$ under the DGHS prior is slightly stronger than under the DHS prior, while the GHS prior produces the strongest shrinkage on $\log(w_{j,t})$, particularly for the constant coefficients. It is interesting to note that, in the case of the change-point coefficient, posteriors of $\log(w_{j,t})$ from both the DGHS and DHS priors show two pronounced spikes around the break points $t = 100$ and 200, suggesting the adaptivity of the shrinkage priors to local signals. Posterior of $\log(w_{j,t})$ from the GHS prior also exhibits two spikes around the break points, though less standout in magnitude than the two dynamic priors.

It may appear counter intuitive that the GHS prior produces stronger shrinkage on the state variance than the DGHS prior as the autoregressive structure in the DGHS prior is more flexible. However this observation ignores that the gamma distributions placed on the global and local components of the GHS and DGHS priors already allows an extra layer of

flexibility and hence the autoregressive structure becomes less important than in the case of the HS and DHS priors. To see this, note that, in both the GHS and DGHS priors, the state variance $w_{j,t} = v_j \phi_{j,t}$ can be re-written as $w_{j,t} = v_j d_{j,t} \phi^*_{j,t}$ where $\phi^*_{j,t} = \frac{\phi_{j,t}}{d_{j,t}}$ and $\phi^*_{j,t} | v_j, d_{j,t} \sim G(0.5, 2)$. Hence local behaviors in the GHS and DGHS priors are controlled by two parameters $d_{j,t}$ and $\phi^*_{j,t}$. In contrast, the HS and DHS priors have a single parameter $d_{j,t}$ to adjust local behavior and hence the extra flexibility rendered by an autoregressive structure on $d_{j,t}$ in the DHS prior makes larger difference relative to the HS prior.

In terms of the mixing efficiency of posterior draws, the DGHS prior performs slightly better than the DHS prior while the GHS prior is the best among the 3 priors. As an illustration, Figure 4 compares the point-wise effective sample size (ESS)[9] from the 3 priors for the log state variances $\log(w_{j,t})$ over $j = 1, ..., 6$ and $t = 1, ..., 300$. The ESSs of $\log(w_{j,t})$ from the GHS and DGHS priors are respectively over 2 times larger and about 1.2 to 1.5 times larger than those from the DHS prior. For both the DGHS and DHS priors, posterior draws of the autoregressive coefficient in the dynamic shrinkage process are found to mix unsatisfactorily, which hinders the sampling efficiency of these two dynamic priors.

# 5 Empirical Illustration

To illustrate the proposed methodology, I apply the TVP model of Equation (1) to forecast the quarterly U.S. inflation rate. The dependent variable is the quarter-to-quarter change of annualized log quarterly inflation rate. The regressors include a constant, lags 1 to 6 of the dependent variable, lagged quarter-to-quarter change of quarterly average 3-month U.S. treasury bill rate, and lagged quarter-to-quarter change of quarterly average U.S. unemployment rate[10]. The data sample runs from Q2 1957 to Q4 2020 with a total of 255

[9]The effective sample size of parameter draws $\{\theta^{(i)}\}_{i=1}^M$ is computed as the ratio of its sample variance $V_0 = \frac{1}{M} \sum_{i=1}^M \left(\theta^{(i)} - \bar{\theta}\right)^2$, where $\bar{\theta} = \frac{1}{M} \sum_{i=1}^M \theta^{(i)}$, to its long-run variance $V$. The long-run variance $V$ is computed via a Bartlett kernel as $V_0 + 2 \sum_{j=1}^{D-1} \left(1 - \frac{j}{D}\right) V_j$ where $D = \text{floor}\left(4(M/100)^{2/9}\right)$ and $V_j = \frac{1}{M} \sum_{i=j+1}^M \left(\theta^{(i)} - \bar{\theta}\right) \left(\theta^{(i-j)} - \bar{\theta}\right)$.

[10]The data source is the FRED database of the U.S. federal reserve bank of St. Louis. The series names are CPILFESL, TB3MS and UNRATE for consumer price index, 3-month treasury bill rate and unemployment rate respectively. Quarterly average is computed as the average monthly values within each quarter.

observations. Figure 5 shows the data used in the estimation. The high volatilities in late 1970s and 2020 are evident in the inflation rate data.

MCMC estimation is performed with 10,000 posterior draws after a burn-in of length 5,000. The time variation patterns of the estimated coefficients are similar across the TVP prior specifications. As an illustration, Figure 6 shows the point-wise posterior medians and 90% credible sets of the coefficients $\beta_t$ from the GHS prior, along with the OLS estimate for comparison. The intercept shows a small yet prolonged dip in the 1990s. For the autoregressive lags, coefficients on lag 1 and 3 trend upwards while coefficients on lag 2, 4, 5 and 6 decline gradually over time, though the magnitude of changes in the autoregressive coefficients is relatively small. The coefficient on the lagged interest rate variable is markedly smaller than the OLS estimate and contains zero within its 90% credit sets for most of the time periods. For the lagged unemployment rate, its coefficient shows a V shape over the data sample. The V bottom is reached around mid 1970s when the posterior of the coefficient on the lagged unemployment rate mostly falls in the negative region, suggesting a pattern consistent with the Phillips curve. Since then, the posterior median of the coefficient on the lagged unemployment rate steadily goes up towards zero.

Figures 7 and 8 show the point-wise posterior medians and 90% credit sets of the log state variance $\log(w_{j,t})$ from the 4 TVP priors for coefficient $j = 1, ..., 9$ and time $t = 1, ..., 248$. The comparison of the shrinkage strength between the TVP priors turns out similar to the findings in the simulation study. The shrinkage on $\log(w_{j,t})$ by the HS prior is much weaker than the other 3 priors. The DGHS prior produces slightly stronger shrinkage on $\log(w_{j,t})$ than the DHS prior, while the strongest shrinkage on $\log(w_{j,t})$ is produced by the GHS prior. There appears no pronounced spikes in the log state variance $\log(w_{j,t})$, suggesting that the time variations in the coefficients are closer to be gradual than being abrupt relative to the level of noise in inflation rate changes.

Turning to out-of-sample forecasts, the log predictive likelihoods discussed in Section 3.3 are used to compare the performance of the 4 TVP priors. An iterative prediction exercise is conducted. First the sample from Q2 1957 to Q4 2003 is used to estimate the TVP models under the 4 priors and predictive likelihoods for Q1 2004 are computed. Next the estimation sample is expanded by one observation to include Q1 2004 and generates

the forecast for Q2 2004. This procedure is repeated until Q4 2020 with a total of 68 out-of-sample forecasts.

Figure 9 shows the trend of the difference in the cumulative log predictive likelihoods of the GHS and DGHS priors relative to those of the HS and DHS priors. The dominance of the GHS and DGHS priors in predictive likelihoods over the HS prior is evident to the extent that the difference between the predictive likelihoods of the GHS and DGHS priors is hardly visible in the upper panel of Figure 9. Relative to the DHS prior, the GHS and DGHS priors show steady gains in predictive likelihoods over the forecast sample. The volatile movements of inflation rates in 2020 appear to result in large predictive gains for the GHS and DGHS priors relative to the DHS prior. Among the two gamma horseshoe priors, the GHS prior performs better than the DGHS prior and accumulates advantage in predictive likelihoods throughout the forecast sample. These forecast results show the benefits from the strong shrinkage of the gamma horseshoe priors.

# 6 Conclusion

The TVP model with time-dependent state variance offers a flexible framework to accommodate a wide variety of time variation patterns of parameter changes but is prone to over-parameterization. Effective shrinkage on the state variance is crucial to reduce the risk of model over-fitting. This paper introduces a hierarchical horseshoe prior named *gamma horseshoe prior* on the state variance that offers stronger shrinkage than the conventional horseshoe prior while maintaining flexibility to accommodate local signals. The new prior is motivated by reparametrizing the TVP model that translates a variance shrinkage problem into a variable shrinkage problem borrowing on the insight from Fruhwirth-Schnatter and Wagner (2010). Dynamic version of the gamma horseshoe prior in the vein of Kowal et al. (2019) is also provided.

A MCMC algorithm is developed to estimate the TVP model under the proposed gamma horseshoe priors that employs the ASIS of Yu and Meng (2011) as the critical ingredient to boost sampling efficiency. The novelty of the algorithm is that it avoids repeated large-scale matrix operations and instead applies the ASIS in a tractable component-wise fashion. In both simulation and an inflation rate forecast exercise, the gamma horseshoe priors are

shown to provide stronger shrinkage than the conventional horseshoe priors. It is also noted that the need for dynamic structure is reduced in the gamma horseshoe priors relative to the case of the conventional horseshoe priors. In out-of-sample forecasts, the gamma horseshoe priors are found to outperform the conventional horseshoe priors.

# Appendix

## A Comparing Gamma Horseshoe Prior with Horseshoe Prior

Consider applying the horseshoe and the gamma horseshoe priors to a parameter $\beta$ such that $\beta \sim N(0, w)$. The marginal distribution of the variance $w$ from the gamma horseshoe prior has no closed form. So the comparison is conducted via simulations. The number of simulations is set to be $m = 1,000,000$.

First it is simulated $\tau_0^{(i)} \sim IB(0.5, 0.5)$, $\tau^{(i)} \sim IB(0.5, 0.5)$ and $d^{(i)} \sim IB(0.5, 0.5)$ for $i = 1, ..., m$ where sampling from the inverted beta distribution $IB(0.5, 0.5)$ is conducted through the hierarchical inverse gamma representation of Makalic and Schmidt (2016). The prior variance of $\beta$ from the horseshoe prior is $w_{\text{HS}}^{(i)} = \tau_0^{(i)} \tau^{(i)} d^{(i)}$. The resulting draws of $\beta$ is $\beta_{\text{HS}}^{(i)} \sim N\left(0, w_{\text{HS}}^{(i)}\right)$.

For the gamma horseshoe prior, it is further simulated $v^{(i)} \sim G\left(0.5, 2\tau_0^{(i)} \tau^{(i)}\right)$ and $\phi^{(i)} \sim G\left(0.5, 2d^{(i)}\right)$. The resulting prior variance is $w_{\text{GHS}}^{(i)} = v^{(i)} \phi^{(i)}$ while draws of $\beta$ is $\beta_{\text{GHS}}^{(i)} \sim N\left(0, w_{\text{GHS}}^{(i)}\right)$.

Following Carvalho et al. (2010), define the *shrinkage parameter* as $\kappa = \frac{1}{1+w}$. $\kappa \to 1$ implies shrinking the parameter $\beta$ towards zero while $\kappa \to 0$ implies minimal shrinkage for $\beta$. Figure 1 compares the histograms of the simulated $\kappa$ from the horseshoe and gamma horseshoe priors. The U-shaped histogram of $\kappa$ from the gamma horseshoe prior is similar to that of the horseshoe prior but is tilted towards complete shrinkage of $\kappa = 1$. The empirical density mass of $\kappa$ located in the three disjoint regions $(0, 0.1]$, $(0.1, 0.9)$ and $[0.9, 1)$ is 33%, 34% and 33% under the horseshoe prior while being 22%, 27% and 51% under the gamma horseshoe prior. Hence the gamma horseshoe prior offers an appreciably

stronger shrinkage for the parameter $\beta$.

Checking the two tail regions of $\beta$: $(-\infty, -25] \cup [25, \infty)$ and $(-\infty, -100] \cup [100, \infty)$, the empirical density mass of the simulated $\beta$ in these regions under the gamma horseshoe prior is 5.6% and 2.3% respectively, which is lower than under the horseshoe prior (8.7% and 3.6% respectively) but remains appreciable to accommodate extreme values of $\beta$.

# B SV Model Estimation

Estimating the SV model of Equation (2) follows Kastner and Fruhwirth-Schnatter (2014). There are two key ingredients in the method of Kastner and Fruhwirth-Schnatter (2014). The first is the log linearization strategy of Omori et al. (2007) that approximates the logarithm of a $\chi^2(1)$-distributed variable by a mixture of normal distributions and hence transforms a non-linear state space model into a linear one. The second is applying the ASIS strategy of Yu and Meng (2011) that directly boosts the sampling efficiency of the long-run mean and the variance parameter of the 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 volatility process. Use the notation of Equation (2). Kastner and Fruhwirth-Schnatter (2014) sets a fixed value for the scale parameter $s_y$ in the gamma prior of the variance parameter $\sigma_y^2 \sim G(0.5, 2s_y)$. This paper instead 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 easily derived 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 Hyper-Parameters of Horseshoe Prior

The horseshoe prior for the square root of the global parameter is $\tilde{v}_j = \pm\sqrt{v_j} \sim N(0, \tau_0\tau_j)$ 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:

$$\tau_0 \sim IB(0.5, 0.5) \Longleftrightarrow \tau_0 \sim IG\left(0.5, \frac{1}{a_0}\right), \quad a_0 \sim IG(0.5, 1)$$
$$\tau_j \sim IB(0.5, 0.5) \Longleftrightarrow \tau_j \sim IG\left(0.5, \frac{1}{a_j}\right), \quad a_j \sim IG(0.5, 1)$$

The posteriors can be derived as follows:

$$\tau_0 | a_0, \{\tilde{v}_j\}_{j=1}^K, \{\tau_j\}_{j=1}^K \sim IG\left(\frac{1+K}{2}, \frac{1}{a_0} + \frac{1}{2}\sum_{j=1}^K \frac{1}{\tau_j}\tilde{v}_j^2\right),$$
$$a_0 | \tau_0 \sim IG\left(1, 1 + \frac{1}{\tau_0}\right),$$
$$\tau_j | \tilde{v}_j, \tau_0, a_j \sim IG\left(1, \frac{1}{a_j} + \frac{1}{2\tau_0}\tilde{v}_j^2\right),$$
$$a_j | \tau_j \sim IG\left(1, 1 + \frac{1}{\tau_j}\right).$$

Similarly for the horseshoe prior of the initial state $\beta_{j,0} \sim N(0, \tau_{0,0}\tau_{j,0})$ with $\tau_{0,0} \sim IB(0.5, 0.5)$ and $\tau_{j,0} \sim IB(0.5, 0.5)$ for $j = 1, ..., K$, the hierarchical inverse gamma representation is:

$$\tau_{0,0} \sim IB(0.5, 0.5) \Longleftrightarrow \tau_{0,0} \sim IG\left(0.5, \frac{1}{b_0}\right), \quad b_0 \sim IG(0.5, 1)$$
$$\tau_{j,0} \sim IB(0.5, 0.5) \Longleftrightarrow \tau_{j,0} \sim IG\left(0.5, \frac{1}{b_j}\right), \quad b_j \sim IG(0.5, 1)$$

with the following posteriors:

$$\tau_{0,0} | b_0, \{\beta_{j,0}\}_{j=1}^K, \{\tau_{j,0}\}_{j=1}^K \sim IG\left(\frac{1+K}{2}, \frac{1}{b_0} + \frac{1}{2}\sum_{j=1}^K \frac{1}{\tau_{j,0}}\beta_{j,0}^2\right),$$
$$b_0 | \tau_{0,0} \sim IG\left(1, 1 + \frac{1}{\tau_{0,0}}\right),$$
$$\tau_{j,0} | \beta_{j,0}, \tau_{0,0}, b_j \sim IG\left(1, \frac{1}{b_j} + \frac{1}{2\tau_{0,0}}\beta_{j,0}^2\right),$$
$$b_j | \tau_{j,0} \sim IG\left(1, 1 + \frac{1}{\tau_{j,0}}\right).$$

# D Dynamic Gamma Horseshoe Prior

For convenience, reproduce the SA parametrization of Equation (5) and the equation for $\log(d_{j,t})$:

$$\begin{aligned}
\Delta\beta^*_{j,t} &\sim N(0, \phi_{j,t}) \\
\phi_{j,t} &\sim G(0.5, 2d_{j,t}) \\
\log(d_{j,t}) &= \rho_j \log(d_{j,t-1}) + \log(\xi_{j,t})
\end{aligned}$$

where $\log(d_{j,0}) = 0$, $\log(\xi_{j,t}) \sim N\left(0, \frac{1}{e_{j,t}}\right)$ and $e_{j,t} \sim PG(1,0)$ for $j = 1, ..., K$ and $t = 1, ..., n$.

The posterior of the scaled state variance can be derived as:

$$\phi_{j,t}|\beta^*_{j,t}, d_{j,t} \sim GIG\left(0, \frac{1}{d_{j,t}}, \left(\Delta\beta^*_{j,t}\right)^2\right)$$

The local parameter $d_{j,t}$ can be viewed as a stochastic volatility following a SV process with the long-run mean of zero. The strategy of Kastner and Fruhwirth-Schnatter (2014) can be applied here to estimate the SV model. Specifically, applying the log linearization strategy of Omori et al. (2007) approximates a $\log(\chi^2(1))$ distribution by a 10-component mixture of normals $\sum_{i=1}^{10} \pi_i N(\mu^*_i, h^*_i)$ and transforms the distribution $\phi_{j,t} \sim G(0.5, 2d_{j,t})$[11] into a normal one $\log(\phi_{j,t}) \sim N\left(\mu^*_{s_{j,t}} + \log(d_{j,t}), h^*_{s_{j,t}}\right)$ where $s_{j,t} \in \{1, 2, ..., 10\}$ is an indicator for the 10-component-mixture-normal distribution. The posterior for the indicator is:

$$p(s_{j,t} = i|\phi_{j,t}, d_{j,t}) \propto (h^*_i)^{-\frac{1}{2}} \exp\left(-\frac{1}{2h^*_i}\left(\log(\phi_{j,t}) - \mu^*_i - \log(d_{j,t})\right)^2\right)\pi_i$$

To sample $\log(d_{j,t})$, the precision-based algorithm of Rue (2001) and McCausland et al. (2011) is used. Let $\tilde{d}_j$ be a $n$-by-1 vector collecting $\log(d_{j,1})$, ..., $\log(d_{j,n})$, $\tilde{y}_j$ be a $n$-by-1 vector collecting $\log(\phi_{j,1}) - \mu^*_{s_{j,1}}$, ..., $\log(\phi_{j,n}) - \mu^*_{s_{j,n}}$, $H$ be a $n$-by-$n$ lower bi-diagonal matrix with ones in the diagonal and $-\rho_j$ in each $(i, i-1)$ element for $i = 2, ..., n$, $P$ be a $n$-by-$n$ diagonal matrix with diagonal elements $h^*_{s_{j,1}}$, ..., $h^*_{s_{j,n}}$ and $Q$ be a $n$-by-$n$ diagonal matrix with diagonal elements $\frac{1}{e_{j,1}}$, ..., $\frac{1}{e_{j,n}}$. It follows that the posterior for $\tilde{d}_j$ is $N(b_d, B_d)$ where $B_d^{-1} = H'Q^{-1}H + P^{-1}$ and $B_d^{-1}b_d = P^{-1}\tilde{y}_j$. The key for efficient computation is to note that the matrix $B_d^{-1}$ is a tri-diagonal band matrix. Hence a Cholesky decomposition of

[11] A Gamma distribution $G(0.5, 2)$ is equivalent to a $\chi^2(1)$ distribution.

$B_d^{-1}$ can be efficiently computed by back-substitution and be directly used to draw $\tilde{d}_j$ from the posterior $N(b_d, B_d)$.

To sample $\log(d_{j,t})$ from the AA representation of Equation (6), note that the system is:

$$\begin{aligned}
&\Delta\beta^*_{j,t} \sim N(0, \phi^*_{j,t} d_{j,t}) \\
&\phi^*_{j,t} \sim G(0.5, 2) \\
&\log(d_{j,t}) = \rho_j \log(d_{j,t-1}) + \log(\xi_{j,t})
\end{aligned}$$

Conditional on $\phi^*_{j,t}$, one can write $\frac{\Delta\beta^*_{j,t}}{\sqrt{\phi^*_{j,t}}} \sim N(0, d_{j,t})$. Hence the local parameter $d_{j,t}$ can be recast as the stochastic volatility in a SV process with $\frac{\left(\Delta\beta^*_{j,t}\right)^2}{\phi^*_{j,t}}$ being the observation variable. Sampling of $\log(d_{j,t})$ follows the same steps as in the SA representation by replacing the observation variable $\phi_{j,t}$ with the new one $\frac{\left(\Delta\beta^*_{j,t}\right)^2}{\phi^*_{j,t}}$.

Conditional on $d_{j,t}$, its hyper-parameters $\rho_j$ and $e_{j,t}$ can be sampled as follows. Given the prior $\rho_j \sim N(\mu_\rho, h_\rho) I_{\{-1<\rho_j<1\}}$, a Metropolis-Hasting step is used to sample $\rho_j$ by a proposal $N(a_\rho, b_\rho)$ where $b_\rho^{-1} = h_\rho^{-1} + \sum_{t=1}^n e_{j,t} \left(\log(d_{j,t-1})\right)^2$ and $b_\rho^{-1} a_\rho = h_\rho^{-1}\mu_\rho + \sum_{t=1}^n e_{j,t} \log(d_{j,t-1}) \log(d_{j,t})$. This paper sets $\mu_\rho = 0.95$ to reflect the prior belief that $d_{j,t}$, being the stochastic volatility of a SV process, is highly persistent and at the same time sets $h_\rho = 1$ to allow a large degree of uncertainty for this prior belief.

Following Kowal et al. (2019), the posterior for $e_{j,t}$ is $e_{j,t}|d_{j,t}, \rho_j \sim PG(1, \log(d_{j,t}) - \rho_j \log(d_{j,t-1}))$[12]. This completes the sampling steps for the dynamic gamma horseshoe prior.

# E Kalman Filter

The iterations to derive the conditional filtering distribution $p(\beta_n|y^n, x^n, \theta^n)$ are adapted from Bitto and Fruhwirth-Schnatter (2019) and are provided below using the notation of Equation (1):

- Start with the initial state $\beta_0 \sim N(b_0, B_0)$ where $b_0 = 0$ and $B_0 = \text{diag}(\tau_0\tau_1, ..., \tau_0\tau_K)$.

[12]Sampling from the polya-gamma distribution is by the Matlab function **pgdraw** written by Enes Makalic and Daniel Schimidt that implements an algorithm from Windle (2013).

- Iterate forward over a prediction step and a filtering step for $t = 1, ..., n$:
  - $\beta_t|\mathcal{I}_{t-1} \sim N(b_{t-1}, R_t)$ with $R_t = B_{t-1} + \text{diag}(w_t)$;
  - $\beta_t|\mathcal{I}_t \sim N(b_t, B_t)$ with $b_t = b_{t-1} + K_t(y_t - \hat{y}_t)$ and $B_t = (I_K - K_t x_t')R_t$, where $\hat{y}_t = x_t' b_{t-1}$, $K_t = R_t x_t S_t^{-1}$ and $S_t = x_t' R_t x_t + \sigma_t^2$.

where $\mathcal{I}_t$ denotes the information set $\{y^t, x^t, \theta^t\}$ at time $t$.

# References


Barndorff-Nielsen, O., J. Kent, and M. Sorensen (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review 50*, 145–159.

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

Bhadra, A., J. Datta, N. Polson, and B. Willard (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis 12*(4), 1105–1131.

Bhadra, A., J. Datta, N. Polson, and B. Willard (2019). Lasso meets horseshoe: A survey. *Statistical Science 34*(3), 405–427.

Bhattacharya, A., A. Chakraborty, and B. Mallick (2016). Fast sampling with gaussian scale mixture priors in high-dimensional regression. *Biometrika 103*(4), 985–991.

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

Carriero, A., T. Clark, and M. Marcellino (2019). Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics 212*(1), 137–154.

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

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.

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. (2004). Efficient bayesian parameter estimation. In A. Harvey, S. Koopman, and N. Shephard (Eds.), *State Space and Unobserved Component Models: Theory and Applications*, pp. 123–151. Cambridge University Press.

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.

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.

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.

Harvey, C. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal 45*(5), 38–45.

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:2005.03906v1 [econ.EM].

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

Johndrow, J., P. Orenstein, and A. Bhattacharya (2020). Scalable approximate mcmc algorithms for the horseshoe prior. *Journal of Machine Learning Research 21*(73), 1–61.

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.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies 65*, 361–393.

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

Lopes, H., R. McCulloch, and R. Tsay (2018). Parsimony inducing priors for large scale state space models. Technical report, Booth School of Business, University of Chicago.

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.

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.

Polson, N. and J. Scott (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis 7*(4), 887–902.

Polson, N., J. Scott, and J. Windle (2013). Bayesian inference for logistic models using polya-gamma latent variables. *Journal of the American Statistical Association 108*(504), 1339–1349.

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

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

Simpson, M., J. Niemi, and V. Roy (2017). Interweaving markov chain monte carlo strategies for efficient estimation of dynamic linear models. *Journal of Computational and Graphical Statistics 26*(1), 152–159.

Windle, J. (2013). Forecasting high-dimensional, time-varying variance-covariance matrices with high-frequency data and sampling polya-gamma random variates for posterior distributions derived from logistic likelihoods. University of Texas at Austin, PhD Thesis.

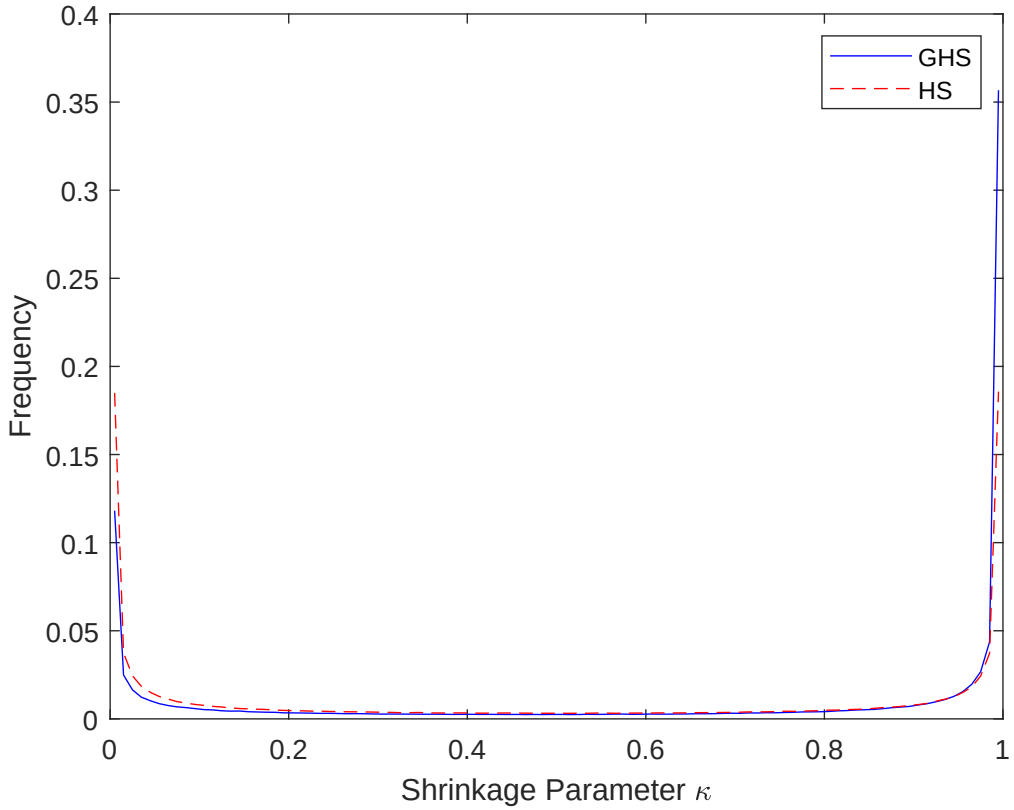
Yu, Y. and X. Meng (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.

Table 1: Comparing Priors for State Variances

| | State Variance | Hierarchical Prior |
|---|---|---|
| Gamma horseshoe (GHS) | $w_{j,t} = v_j \phi_{j,t}$ | $v_j \sim G(0.5, 2\tau_0\tau_j)$, $\phi_{j,t} \sim G(0.5, 2d_{j,t})$ |
| | | $\tau_0 \sim IB(0.5, 0.5)$, $\tau_j \sim IB(0.5, 0.5)$ |
| | | $d_{j,t} \sim IB(0.5, 0.5)$ |
| Dynamic gamma horseshoe (DGHS) | $w_{j,t} = v_j \phi_{j,t}$ | $v_j \sim G(0.5, 2\tau_0\tau_j)$, $\phi_{j,t} \sim G(0.5, 2d_{j,t})$ |
| | | $\tau_0 \sim IB(0.5, 0.5)$, $\tau_j \sim IB(0.5, 0.5)$ |
| | | $d_{j,t} = d_{j,t}^{\rho_j} e_{j,t}$, $e_{j,t} \sim IB(0.5, 0.5)$ |
| Horseshoe (HS) | $w_{j,t} = \tau_0\tau_j d_{j,t}$ | $\tau_0 \sim IB(0.5, 0.5)$, $\tau_j \sim IB(0.5, 0.5)$ |
| | | $d_{j,t} \sim IB(0.5, 0.5)$ |
| Dynamic horseshoe (DHS) | $w_{j,t} = \tau_0\tau_j d_{j,t}$ | $\tau_0 \sim IB(0.5, 0.5)$, $\tau_j \sim IB(0.5, 0.5)$ |
| | | $d_{j,t} = d_{j,t}^{\rho_j} e_{j,t}$, $e_{j,t} \sim IB(0.5, 0.5)$ |

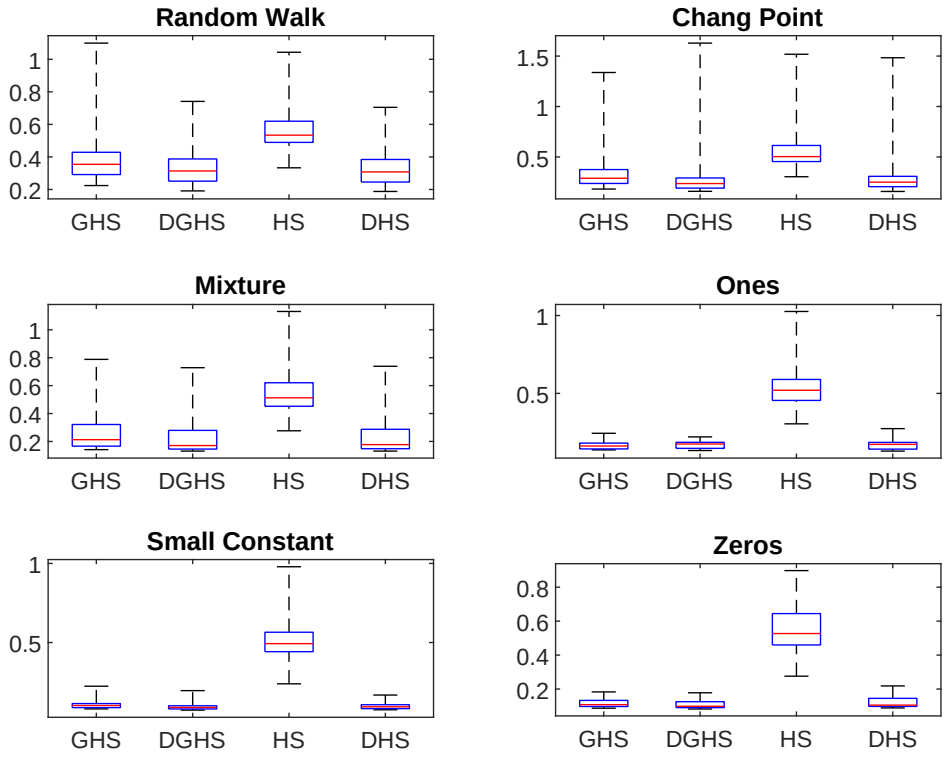
Note: This table compares the priors for states variances in TVP models with heteroskedastic latent states as in Equation (1). When writing the hierarchical priors, the conditioning variables in conditional distributions are omitted to save space.

Figure 1: Histogram of Shrinkage Parameter



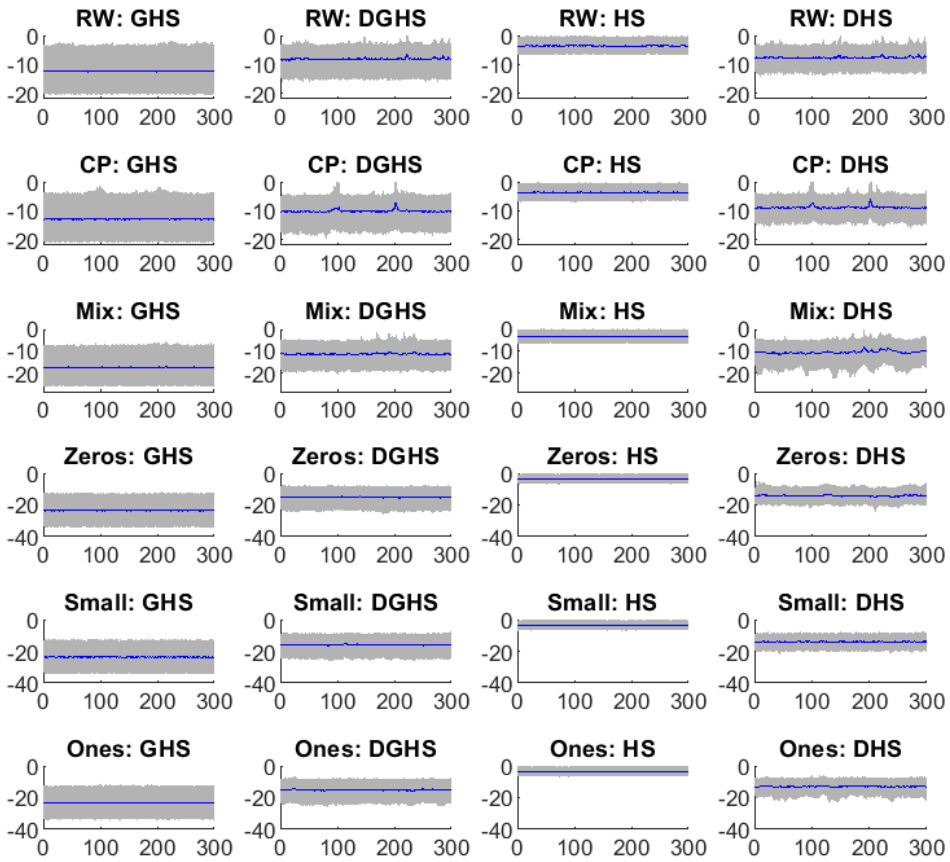

Note: The shrinkage parameter is $\kappa = \frac{1}{1+w}$ where $w$ is simulated separately from the horseshoe (HS, dash line) and gamma horseshoe (GHS, solid line) priors.

Figure 2: Root Mean Squared Error of Coefficient Estimates: Simulation



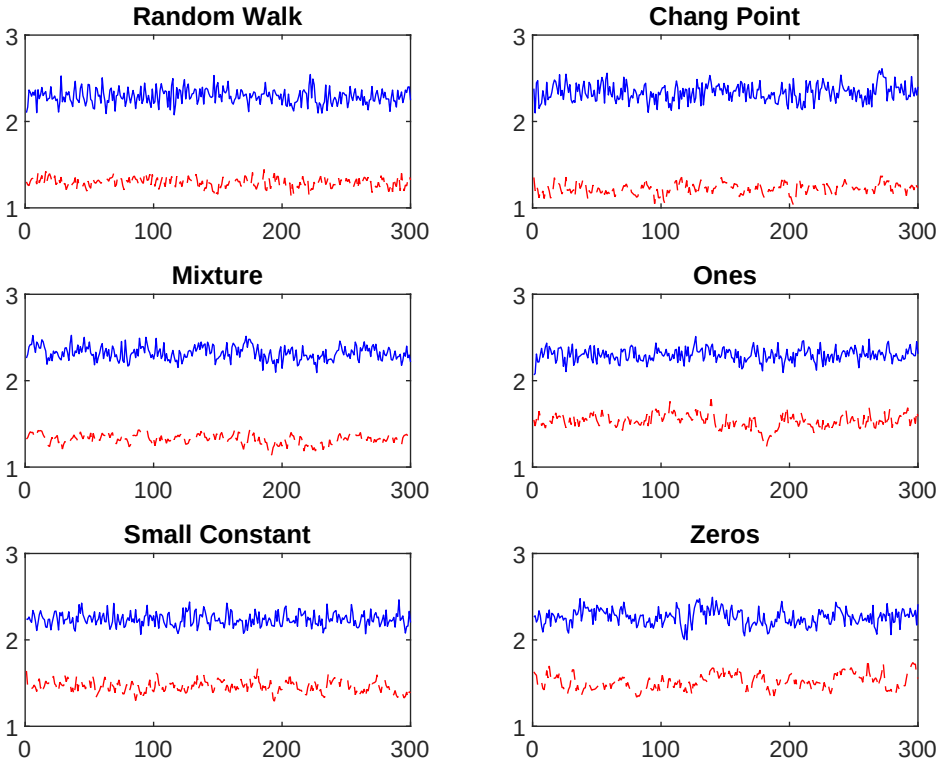

Note: The figure shows the box plots of the root mean squared errors (Equation (9)) of estimated coefficient $\beta_{j,t}$ over time points $t = 1, ..., 300$ for each coefficient $j = 1, ..., 6$ under the 4 TVP priors: gamma horseshoe (GHS), dynamic gamma horseshoe (DGHS), horseshoe (HS) and dynamic horseshoe (DHS). On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the $25^{\text{th}}$ and $75^{\text{th}}$ percentiles respectively. The whiskers extend to the most extreme data points.

Figure 3: Posterior of Log State Variance: Simulation



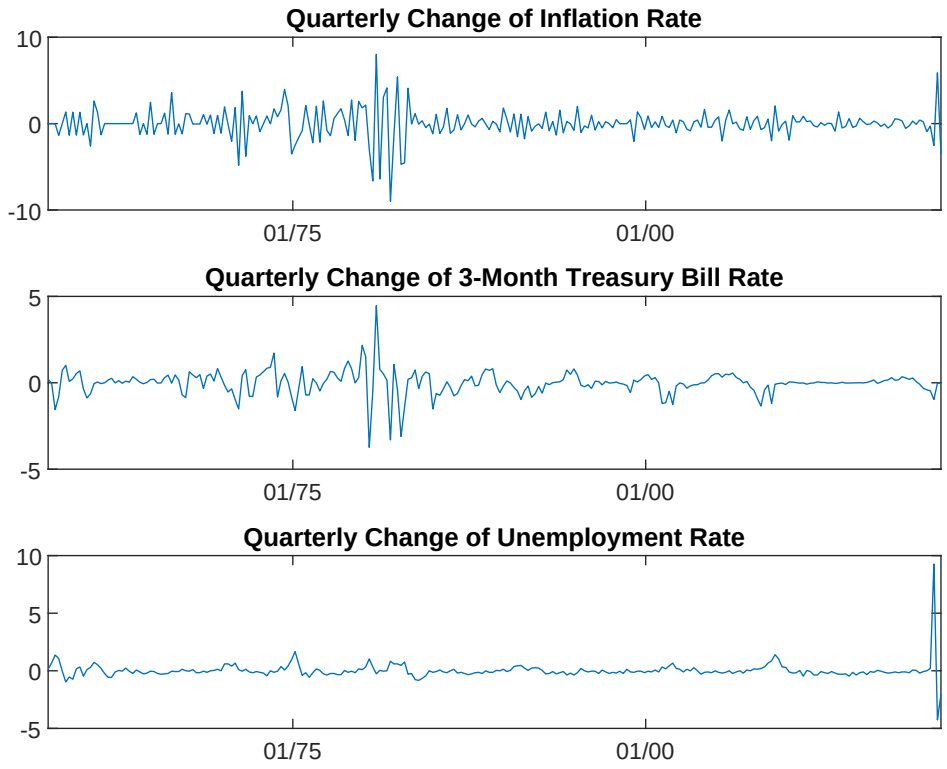

Note: The figure shows the point-wise posterior median (solid line) and 90% credible sets (grey shade) of the log state variance $\log(w_{j,t})$ over time points $t = 1, ..., 300$ for each coefficient $j = 1, ..., 6$ under the 4 TVP priors: gamma horseshoe (GHS), dynamic gamma horseshoe (DGHS), horseshoe (HS) and dynamic horseshoe (DHS). The coefficients "random walk", "change point", "mixture" and "small constant" are abbreviated as "RW", "CP", "Mix" and "Small" respectively to save space.

Figure 4: Ratio of Point-Wise Effective Sample Size of Log State Variance: Simulation



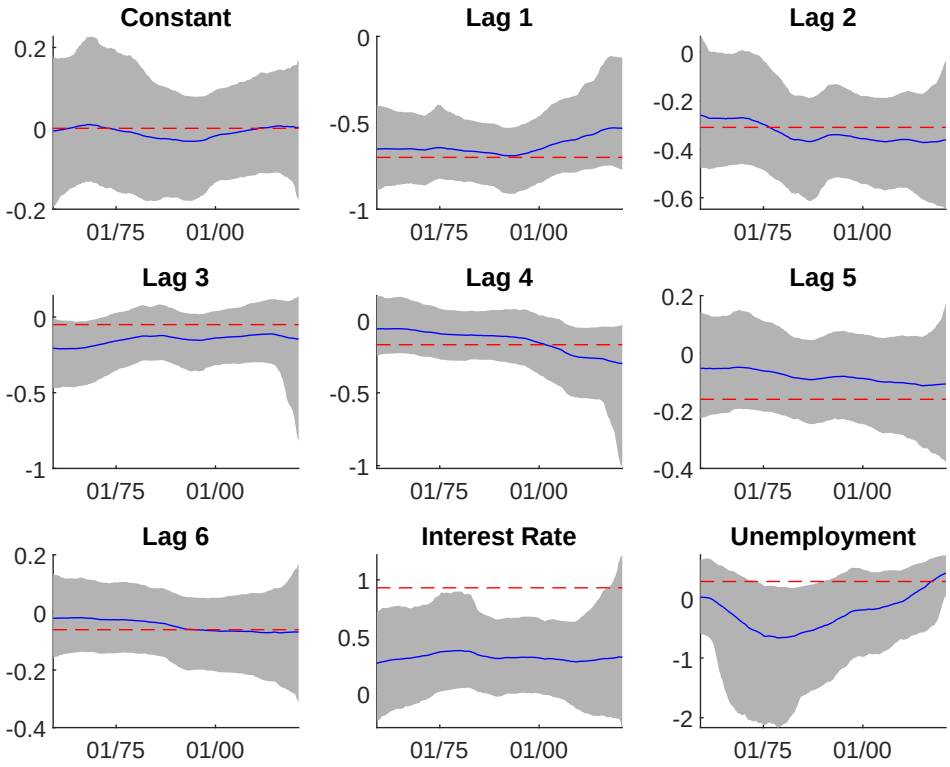

Note: The figure shows the ratio of the point-wise effective sample size of the posterior draws of the log state variance $\log(w_{j,t})$ under the gamma horseshoe (GHS, solid line) and dynamic gamma horseshoe (DGHS, dash line) priors to that under the dynamic horseshoe (DHS) prior over time points $t = 1, ..., 300$ for each coefficient $j = 1, ..., 6$. Calculation of the effective sample size follows the discussion in Section 4. A larger effective sample size implies better mixing behavior of the posterior draws.

Figure 5: Data for Empirical Application: Inflation Rate



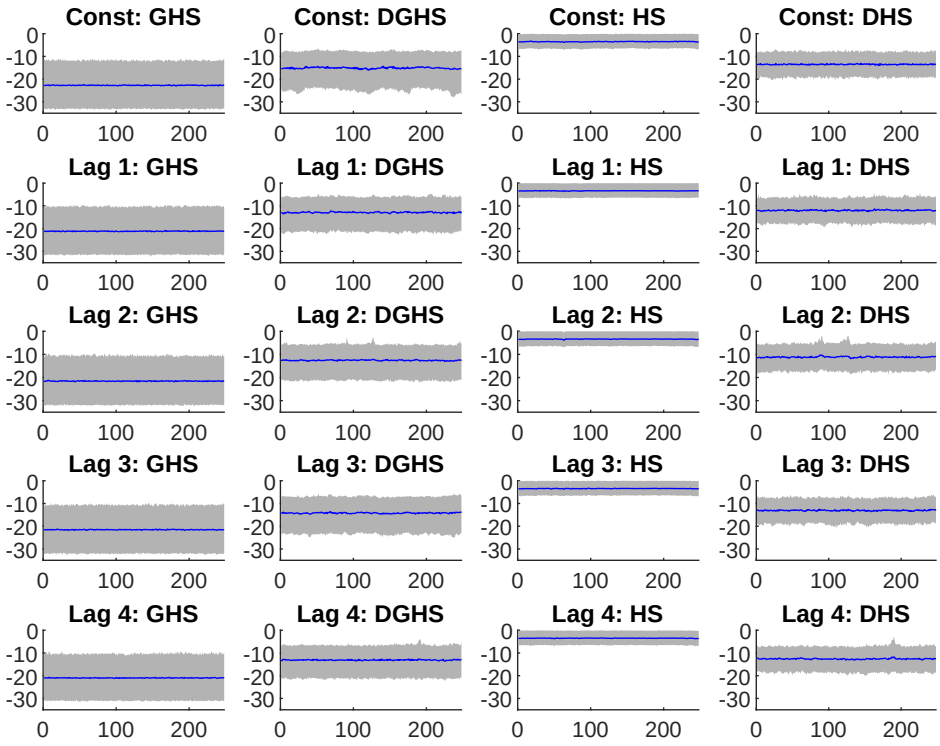

Note: The data source is the FRED database of the U.S. federal reserve bank of St. Louis. The data sample is from Q2 1957 to Q4 2020. The series "inflation rate" is the annualized log quarterly inflation rate based on the U.S. core consumer price index. For the interest rate and unemployment rate, their quarterly values are the average monthly values within each quarter.

Figure 6: Coefficient Estimates by the Gamma Horseshoe Prior: Inflation Rate



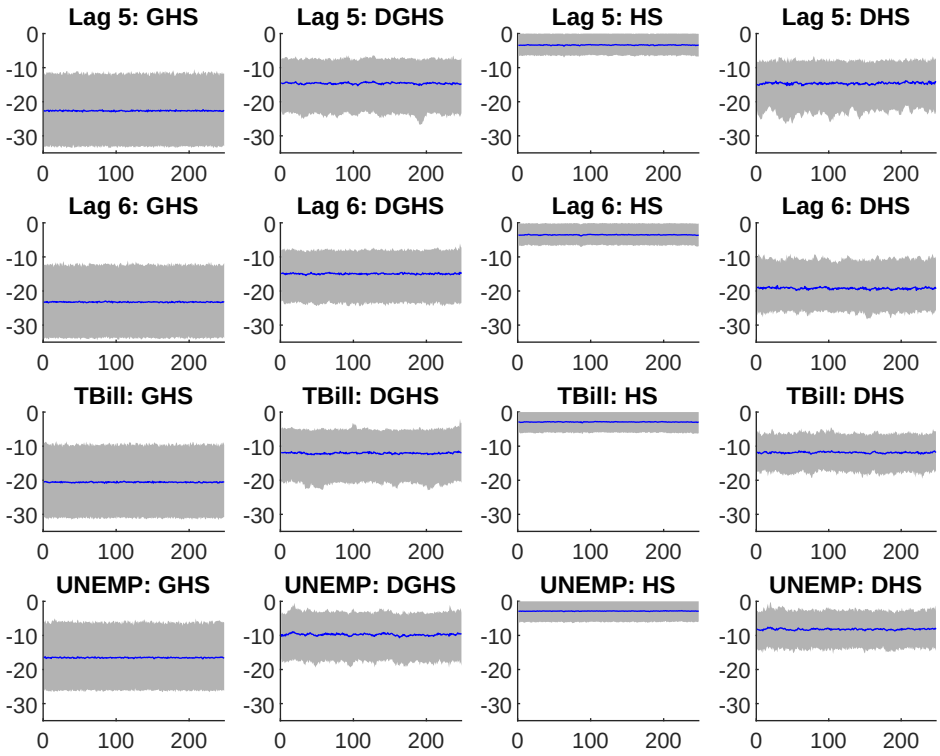

Note: The solid line is the point-wise posterior median of the coefficients with the grey shade being the point-wise 90% credible sets. The dash line is the OLS estimate.

Figure 7: Posterior of Log State Variances: Inflation Rate, Part 1



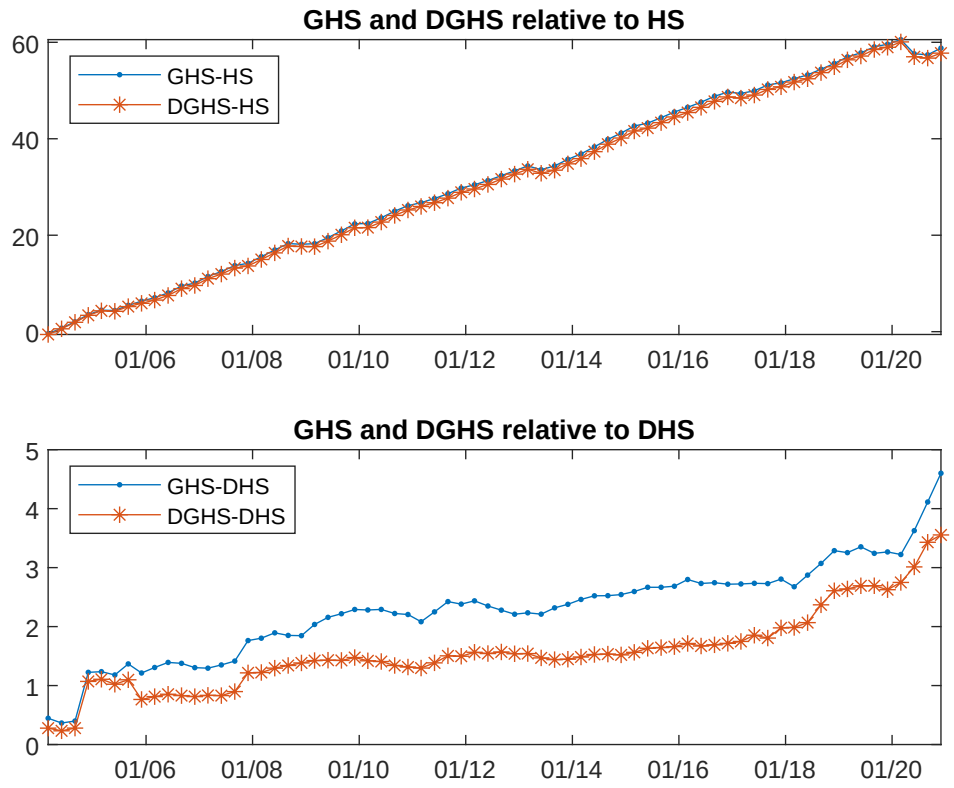

Note: The figure shows the point-wise posterior median (solid line) and 90% credible sets (grey shade) of the log state variance $\log(w_{j,t})$ over time $t$ for coefficients on the regressors "constant", "lag 1","lag 2","lag 3" and "lag 4" under the 4 TVP priors: gamma horseshoe (GHS), dynamic gamma horseshoe (DGHS), horseshoe (HS) and dynamic horseshoe (DHS).

Figure 8: Posterior of Log State Variances: Inflation Rate, Part 2




Note: The figure shows the point-wise posterior median (solid line) and 90% credible sets (grey shade) of the log state variance $\log(w_{j,t})$ over time $t$ for coefficients on the regressors "lag 5","lag 6","interest rate" (TBill) and "unemployment" (UNEMP) under the 4 TVP priors: gamma horseshoe (GHS), dynamic gamma horseshoe (DGHS), horseshoe (HS) and dynamic horseshoe (DHS).

Figure 9: Difference in Cumulative Log Predictive Likelihoods: Inflation Rate




Note: The line with the point marker (.) is the cumulative log predictive likelihood of the gamma horseshoe (GHS) prior relative to that of the horseshoe (HS) prior (the upper panel) and the dynamic horseshoe (DHS) prior (lower panel). The line with the asterisk marker (*) is the cumulative log predictive likelihood of the dynamic gamma horseshoe (DGHS) prior relative to that of the HS prior (the upper panel) and the DHS prior (lower panel).