Dynamic Shrinkage Processes

Daniel R. Kowal
Rice University, Houston, TX, USA.
E-mail: daniel.kowal@rice.edu
David S. Matteson
Cornell University, Ithaca, NY, USA.
David Ruppert
Cornell University, Ithaca, NY, USA.

Summary. We propose a novel class of dynamic shrinkage processes for Bayesian time series and regression analysis. Building upon a global-local framework of prior construction, in which continuous scale mixtures of Gaussian distributions are employed for both desirable shrinkage properties and computational tractability, we model dependence among the local scale parameters. The resulting processes inherit the desirable shrinkage behavior of popular global-local priors, such as the horseshoe prior, but provide additional localized adaptivity, which is important for modeling time series data or regression functions with local features. We construct a computationally efficient Gibbs sampling algorithm based on a Pólya-Gamma scale mixture representation of the proposed process. Using dynamic shrinkage processes, we develop a Bayesian trend filtering model that produces more accurate estimates and tighter posterior credible intervals than competing methods, and apply the model for irregular curve-fitting of minute-by-minute Twitter CPU usage data. In addition, we develop an adaptive time-varying parameter regression model to assess the efficacy of the Fama-French five-factor asset pricing model with momentum added as a sixth factor. Our dynamic analysis of manufacturing and healthcare industry data shows that with the exception of the market risk, no other risk factors are significant except for brief periods.

Keywords: time series; trend filtering; dynamic linear model; stochastic volatility; asset pricing

1. Introduction

The global-local class of prior distributions is a popular and successful mechanism for providing shrinkage and regularization in a broad variety of models and applications. Global-local priors use continuous scale mixtures of Gaussian distributions to produce desirable shrinkage properties, such as (approximate) sparsity or smoothness, often leading to highly competitive and computationally tractable estimation procedures. For example, in the variable selection context, exact sparsity-inducing priors such as the spike-and-slab prior become intractable for even a moderate number of predictors. By comparison, global-local priors that shrink toward sparsity, such as the horseshoe prior (Carvalho et al., 2010), produce competitive estimators with greater scalability, and are validated by theoretical results, simulation studies, and a variety of applications (Datta and Ghosh, 2013; van der Pas et al., 2014). Unlike non-Bayesian counterparts such as the lasso (Tibshirani, 1996), shrinkage priors also provide adequate uncertainty quantification for parameters of interest (Kyung et al., 2010; van der Pas et al., 2014).

Global-local priors define a joint distribution for a set of variables $\{\omega_t\}_{t=1}^T$, and induce shrinkage via two key components: a *global* scale parameter τ, which controls the shrinkage common to all $\{\omega_t\}$, and *local* scale parameters $\{\lambda_t\}_{t=1}^T$, which control the shrinkage for each individual ω_t. Careful choice of priors for λ_t^2 and τ^2 provides both the flexibility to accommodate large signals and adequate shrinkage of noise (e.g., Carvalho et al., 2010). Most commonly, the local scale parameters $\{\lambda_t\}$ are assumed to be *a priori* independent and identically distributed (iid). However, it can be advantageous to forgo the independence assumption. In the dynamic setting, in which the observations are time-ordered, it is natural to allow the local scale parameter, λ_t, to depend on the history of the shrinkage process $\{\lambda_s\}_{s<t}$. Such model-based dependence may improve the ability of the model to adapt dynamically, which is important for time series estimation, forecasting, and inference.

For a motivating example, consider the minute-by-minute Twitter CPU usage data in Figure 1 (James et al., 2016). The data (top left) show an overall smooth trend interrupted by irregular jumps throughout the morning and early afternoon, with an increase in volatility from 16:00-18:00. It is important to identify both abrupt changes as well as slowly-varying intraday trends. To model these features, consider the following Gaussian dynamic linear model:

$$\begin{cases} y_t = \beta_t + \epsilon_t, & [\epsilon_t|\sigma_t] \overset{indep}{\sim} N(0, \sigma_t^2) \\ \Delta^2\beta_{t+1} = \omega_t, & [\omega_t|\tau, \lambda_t] \overset{indep}{\sim} N(0, \tau^2\lambda_t^2) \end{cases} \tag{1}$$

where $\Delta^2\beta_{t+1} = \Delta\beta_t - \Delta\beta_{t-1}$ for differencing operator Δ. Model (1) is a Bayesian adaption of the *trend filtering* model of Kim et al. (2009) and Tibshirani (2014), also proposed by Faulkner and Minin (2018), and includes stochastic volatility for the observation error variance, σ_t^2, with a global-local shrinkage prior for the second differences of the conditional mean, $\Delta^2\beta_{t+1} = \omega_t$. The shrinkage behavior of ω_t determines the path of β_t: when ω_t is pulled toward zero, β_t is locally linear, while large innovations $|\omega_t|$ correspond to large changes in the slope of β_t (Figure 1, top right). Naturally, the global and local scale parameters, τ and λ_t, play a vital role in determining the magnitude of each ω_t, and thus the smoothness and adaptability of the conditional mean β_t.

Figure 1 illustrates model (1) applied to the Twitter CPU usage data based on a *dynamic horseshoe prior* for $\{\omega_t\}$, introduced in Section 2. Unlike the classical horseshoe prior of Carvalho et al. (2010), the proposed dynamic extension incorporates dependence among the local scale parameters, λ_t, so that the shrinkage behavior at time t is informed by $\{\lambda_s\}_{s<t}$. Notably, the resulting posterior expectation of β_t and credible bands for the posterior predictive distribution of y_t adapt to both irregular jumps and smooth trends (Figure 1, top left). The horseshoe-like

shrinkage behavior of λ_t is evident: values of λ_t are either near zero, corresponding to aggressive shrinkage of $\omega_t = \Delta^2 \beta_t$ to zero, or large, corresponding to large absolute changes in the slope of β_t (Figure 1, bottom right). Importantly, Figure 1 also provides motivation for a *dynamic* shrinkage process: there is clear *volatility clustering* of $\{\lambda_t\}$, in which the shrinkage induced by λ_t persists for consecutive time points. The volatility clustering reflects—and motivates—the temporally adaptive shrinkage behavior of the dynamic shrinkage process.

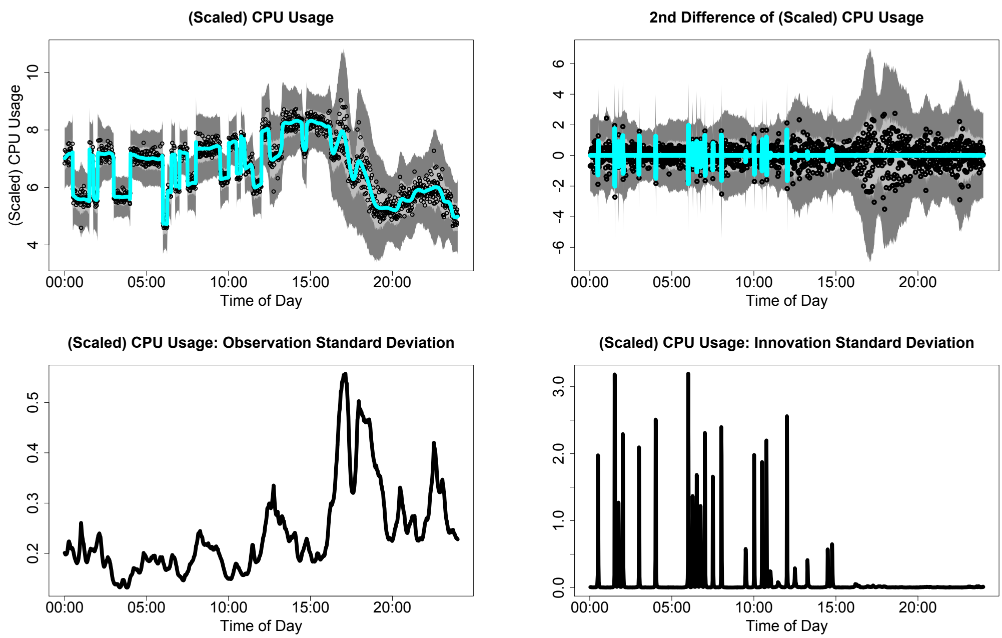

Fig. 1. Bayesian trend filtering with dynamic horseshoe process innovations of minute-by-minute Twitter CPU usage data. Top left: Observed data y_t (points), posterior expectation (cyan) of β_t, and 95% pointwise highest posterior density (HPD) credible intervals (light gray) and 95% simultaneous credible bands (dark gray) for the posterior predictive distribution of y_t. Top right: Second difference of observed data $\Delta^2 y_t$ (points), posterior expectation of $\omega_t = \Delta^2 \beta_t$ (cyan), and 95% pointwise HPD intervals (light gray) and simultaneous credible bands (dark gray) for the posterior predictive distribution of $\Delta^2 y_t$. Bottom left: Posterior expectation of time-dependent observation standard deviations, σ_t. Bottom right: Posterior expectation of time-dependent innovation (prior) standard deviations, $\tau\lambda_t$.

More broadly, we introduce a framework for modeling dependence among the local scale parameters $\{\lambda_t\}$. Using a novel log-scale representation of a broad class of shrinkage priors, we gain the ability to incorporate a variety of widely successful models for dependent data, such as (vector) autoregressions, linear regressions, Gaussian processes, and factor models, into the shrinkage process model. Focusing on the case of dynamic dependence, we demonstrate that the additional structure in the shrinkage process produces more accurate point estimates as well as substantially tighter and more adaptive credible intervals for both real and simulated data. To accompany the proposed log-scale representation of global-local shrinkage priors, we design new parameter expansion techniques using Pólya-Gamma random variables for efficient and scalable posterior inference.

Shrinkage priors have been applied successfully and broadly for time series modeling. Belmonte et al. (2014), Korobilis (2013), and Bitto and Frühwirth-Schnatter (2016) propose (global) shrinkage priors for dynamic linear models and time series regression, but do not allow for (local) time-specific shrinkage for each variable. Discrete mixture and spike-and-slab models offer an alternative approach, but suffer from inherent computational challenges. One option is to restrict the space of models under consideration: Chan et al. (2012) include or exclude a variable for all times, while Frühwirth-Schnatter and Wagner (2010) also consider whether each variable is globally static or dynamic. The extensions in Huber et al. (2018) and Uribe and Lopes (2017) allow for locally static or dynamic variables, while Nakajima and West (2013) provide a procedure for local thresholding of dynamic coefficients. Rockova and McAlinn (2017) develop an optimization approach for dynamic variable selection, which provides point estimates.

Perhaps most comparable to the proposed methodology, Kalli and Griffin (2014) propose a class of priors which exhibit dynamic shrinkage using normal-gamma autoregressive processes. The Kalli and Griffin (2014) prior is a dynamic extension of the normal-gamma prior of Griffin and Brown (2010), and provides improvements in forecasting performance relative to non-dynamic shrinkage priors. However, the Kalli and Griffin (2014) model requires careful specification of several hyperparameters and hyperpriors, and the computation requires sophisticated adaptive MCMC techniques, which results in lengthy computation times. By comparison, our proposed class of dynamic shrinkage processes is far more general, and includes the dynamic horseshoe process as a special case—which notably does not require tuning of hyperparameters. Empirically, for time-varying parameter regression models with dynamic shrinkage, the Kalli and Griffin (2014) MCMC sampler requires several hours, while our proposed MCMC sampler runs in only a few minutes (see Section 4.1 for details and a comparison of these methods).

We introduce dynamic shrinkage processes in Section 2. In Section 3, we apply the processes to develop a more adaptive Bayesian trend filtering model for irregular curve-fitting. The proposed procedure is compared with state-of-the-art alternatives through simulations and the CPU usage data. In Section 4, we propose a time-varying parameter regression model with dynamic shrinkage processes for adaptive regularization and evaluate the model using simulations and an asset pricing example. In Section 5, we discuss the details of the Gibbs sampling algorithm, and conclude in Section 6. Proofs and additional details are in the supplement.

2. Dynamic Shrinkage Processes

Model (1) features a global-local scale mixture of Gaussian distributions for the innovations:

$$[\omega_t | \tau, \lambda_t] \overset{indep}{\sim} N(0, \tau^2 \lambda_t^2) \tag{2}$$

Table 1. Special cases of the inverted-Beta prior.

$\alpha = \beta = 1/2$	Horseshoe Prior	Carvalho et al. (2010)
$\alpha = 1/2, \beta = 1$	Strawderman-Berger Prior	Strawderman (1971)
$\alpha = 1, \beta = c - 2, c > 0$	Normal-Exponential-Gamma Prior	Griffin and Brown (2005)
$\alpha = \beta \to 0$	(Improper) Normal-Jeffreys' Prior	Figueiredo (2003)

Global-local priors are particularly well-suited for sparse data: τ determines the global level of sparsity for $\{\omega_t\}_{t=1}^T$, while large λ_t allows large absolute deviations of ω_t from its prior mean (zero) and small λ_t provides extreme shrinkage to zero. We propose to model dependence of the log-variance process $h_t = \log(\tau^2 \lambda_t^2)$ using the general dependent data model

$$h_t = \mu + \psi_t + \eta_t, \quad \eta_t \overset{iid}{\sim} Z(\alpha, \beta, 0, 1) \tag{3}$$

where $\mu = \log(\tau^2)$ corresponds to the global scale parameter, $(\psi_t + \eta_t) = \log(\lambda_t^2)$ corresponds to the local scale parameter, and $Z(\alpha, \beta, \mu_z, \sigma_z)$ denotes the *Z-distribution* with density function

$$[z] = \left[\sigma_z B(\alpha, \beta)\right]^{-1} \left\{ \exp\left[(z - \mu_z)/\sigma_z\right] \right\}^{\alpha} \left\{ 1 + \exp\left[(z - \mu_z)/\sigma_z\right] \right\}^{-(\alpha+\beta)}, z \in \mathbb{R}$$

where $B(\cdot, \cdot)$ is the Beta function. In model (3), the local scale parameter $\lambda_t = \exp[(\psi_t + \eta_t)/2]$ has two components: ψ_t, which models dependence (see below), and η_t, which corresponds to the usual iid (log-) local scale parameter. When $\psi_t = 0$, model (3) reduces to the static setting, and implies an *inverted-Beta* prior for λ_t^2 (see Section 2.1). Notably, the class of priors represented in (3) includes the important shrinkage distributions in Table 1, in each case extended to the dependent data setting.

The role of ψ_t in (3) is to provide locally adaptive shrinkage by modeling dependence. For dynamic dependence, we propose the *dynamic shrinkage process*

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad \eta_t \overset{iid}{\sim} Z(\alpha, \beta, 0, 1) \tag{4}$$

which is equivalent to (3) with $\psi_t = \phi(h_{t-1} - \mu)$. Relative to static shrinkage priors, model (4) only adds one parameter, ϕ, and reduces to the static setting when $\phi = 0$. Importantly, the proposed Gibbs sampler for the parameters in (4) is linear in the number of time points, T, and therefore scalable. Other examples of (3) include linear regression, $\psi_t = \boldsymbol{z}_t' \boldsymbol{\alpha}$ for a vector of predictors $\boldsymbol{z}_t$, Gaussian processes, and various multivariate models (see (7) in Section 4). We focus on dynamic dependence, but our modeling framework and computational techniques may be extended to incorporate more general dependence among shrinkage parameters.

2.1. Log-Scale Representations of Global-Local Priors

Models (3) and (4) do not automatically induce desirable locally adaptive shrinkage properties: we must consider appropriate distributions for μ and η_t. To illustrate this point, suppose $\eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2)$ in (4), which is a common assumption in stochastic volatility (SV) modeling (Kim et al., 1998). For the likelihood $[y_t|\omega_t] \stackrel{indep}{\sim} N(\omega_t, 1)$ and the prior (2), the posterior expectation of ω_t is $\mathbb{E}[\omega_t|\{y_s\}, \tau] = (1 - \mathbb{E}[\kappa_t|\{y_s\}, \tau])\, y_t$, where

$$\kappa_t \equiv \frac{1}{1 + \mathrm{Var}\left[\omega_t|\tau, \lambda_t\right]} = \frac{1}{1 + \tau^2\lambda_t^2} \tag{5}$$

is the *shrinkage parameter*. As noted by Carvalho et al. (2010), $\mathbb{E}[\kappa_t|\{y_s\}, \tau]$ is interpretable as the amount of shrinkage toward zero *a posteriori*: $\kappa_t \approx 0$ yields minimal shrinkage (for signals), while $\kappa_t \approx 1$ yields maximal shrinkage to zero (for noise). For the standard SV model and fixing $\phi = \mu = 0$ for simplicity, $\lambda_t = \exp(\eta_t/2)$ is log-normally distributed, and the shrinkage parameter has density $[\kappa_t] \propto \frac{1}{\kappa_t(1-\kappa_t)} \exp\left\{ -\frac{1}{2\sigma_\eta^2}\left[\log\left(\frac{1-\kappa_t}{\kappa_t}\right)\right]^2\right\}$. Notably, the density for κ_t approaches zero as $\kappa_t \to 0$ and as $\kappa_t \to 1$. As a result, direct application of the Gaussian SV model may overshrink true signals and undershrink noise.

By comparison, consider the horseshoe prior of Carvalho et al. (2010), which combines (2) with $[\lambda_t] \stackrel{iid}{\sim} C^+(0, 1)$, where C^+ denotes the half-Cauchy distribution. For fixed $\tau = 1$, the half-Cauchy prior on λ_t is equivalent to $\kappa_t \stackrel{iid}{\sim} \mathrm{Beta}(1/2, 1/2)$, which induces a "horseshoe" shape for the shrinkage parameter (see Figure 2). The horseshoe-like behavior is ideal in sparse settings, since the prior density allocates most of its mass near zero (minimal shrinkage of signals) and one (maximal shrinkage of noise).

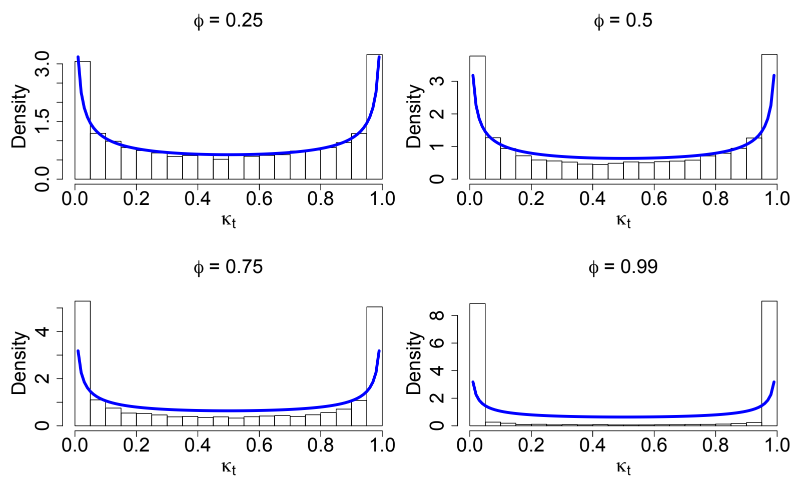

Fig. 2. Simulation-based estimate of the stationary distribution of κ_t for various AR(1) coefficients ϕ. The blue line indicates the density of κ_t in the static ($\phi = 0$) horseshoe, $[\kappa] \sim \mathrm{Beta}\,(1/2, 1/2)$.

To emulate the robustness and sparsity properties of the horseshoe and other shrinkage priors in the dynamic setting, we represent a general class of global-local shrinkage priors on the log-scale. As a motivating example, consider the special case of (2) and (4) with $\phi = 0$, so $\log(\lambda_t^2) = \eta_t$. This example is illuminating: we equivalently express the (static) horseshoe prior by letting $\eta_t \stackrel{D}{=} \log \lambda_t^2$, where $\stackrel{D}{=}$ denotes equality in distribution. In particular, $\lambda_t \sim C^+(0,1)$ implies $\left[\lambda_t^2\right] \propto \left(\lambda_t^2\right)^{-1/2} \left(1+\lambda_t^2\right)^{-1}$ which implies $[\eta_t] = \pi^{-1}\exp(\eta_t/2)\left[1+\exp(\eta_t)\right]^{-1}$ is Z-distributed with $\eta_t \sim Z(1/2, 1/2, 0, 1)$. Importantly, Z-distributions may be written as mean-variance scale mixtures of Gaussian distributions (Barndorff-Nielsen et al., 1982), which produces a useful framework for a parameter-expanded Gibbs sampler.

More generally, consider the inverted-Beta prior, denoted $IB(\beta,\alpha)$, for λ^2 with density $[\lambda^2] \propto \left(\lambda^2\right)^{\alpha-1}\left(1+\lambda^2\right)^{-(\alpha+\beta)}, \lambda > 0$ using the parameterization of Polson and Scott (2012a,b). Special cases of the inverted-Beta distribution are provided in Table 1. This broad class of priors may be equivalently constructed via the variances λ_t^2, the shrinkage parameters κ_t, or the log-variances η_t.

Proposition 1. *The following distributions are equivalent:*

(a) $\lambda^2 \sim IB(\beta,\alpha)$*;*
(b) $\kappa = 1/\left(1+\lambda^2\right) \sim Beta(\beta,\alpha)$*;*
(c) $\eta = \log(\lambda^2) = \log(\kappa^{-1}-1) \sim Z(\alpha,\beta,0,1)$*.*

Note that the ordering of the parameters α, β is identical for the inverted-Beta and Beta distributions, but reversed for the Z-distribution.

Now consider the dynamic setting in which $\phi \neq 0$. Model (4) implies that the conditional prior variance for ω_t is $\exp(h_t) = \exp(\mu + \phi(h_{t-1}-\mu) + \eta_t) = \tau^2 \lambda_{t-1}^{2\phi} \tilde{\lambda}_t^2$, where $\tau^2 = \exp(\mu)$, $\lambda_{t-1}^2 = \exp(h_{t-1}-\mu)$, and $\tilde{\lambda}_t^2 = \exp(\eta_t) \stackrel{iid}{\sim} IB(\beta,\alpha)$, as in the non-dynamic setting. This prior generalizes the $IB(\beta,\alpha)$ prior via the local variance term, $\lambda_{t-1}^{2\phi}$, which incorporates information about the shrinkage behavior at the previous time $t-1$ in the prior for ω_t. We formalize the role of this local adjustment term with the following result.

Proposition 2. *Suppose* $\eta \sim Z(\alpha,\beta,\mu_z,1)$ *for* $\mu_z \in \mathbb{R}$*. Then* $\kappa = 1/(1+\exp(\eta)) \sim TPB(\beta,\alpha,\exp(\mu_z))$*, where* $\kappa \sim TPB(\beta,\alpha,\gamma)$ *denote the* three-parameter Beta distribution *with density* $[\kappa] = [B(\alpha,\beta)]^{-1}\gamma^\beta \kappa^{\beta-1}(1-\kappa)^{\alpha-1}\left[1+(\gamma-1)\kappa\right]^{-(\alpha+\beta)}$ *for* $\kappa \in (0,1), \gamma > 0$*.*

The three-parameter Beta (TPB) distribution (Armagan et al., 2011) generalizes the Beta distribution: $\gamma = 1$ produces the Beta(β,α) distribution, while $\gamma > 1$ (respectively, $\gamma < 1$) allocates more mass near zero (respectively, one) relative to the Beta(β,α) distribution. In Section B of the supplementary material, we generalize for a Z-distribution with $\sigma_z \neq 1$, which produces a new class of shrinkage priors with additional flexibility in the distribution of κ, especially near zero and one.

For dynamic shrinkage processes, the TPB distribution arises as the conditional prior distribution of κ_{t+1} given $\{\kappa_s\}_{s\leq t}$.

Theorem 1. *For the dynamic shrinkage process* (4)*, the conditional prior distribution of the shrinkage parameter* $\kappa_{t+1} = 1/\left(1+\tau^2\lambda_{t+1}^2\right)$ *is* $[\kappa_{t+1} | \{\kappa_s\}_{s\leq t}, \phi, \tau] \sim TPB\left\{\beta, \alpha, \tau^{2(1-\phi)}\left[(1-\kappa_t)/\kappa_t\right]^\phi\right\}$ *or equivalently,* $[\kappa_{t+1} | \{\lambda_s\}_{s\leq t}, \phi, \tau] \sim TPB(\beta,\alpha,\tau^2\lambda_t^{2\phi})$*.*

The previous value of the shrinkage parameter, κ_t, together with the AR(1) coefficient ϕ, inform the magnitude and direction of the distributional shift of κ_{t+1}.

THEOREM 2. *For the dynamic horseshoe process of* (4) *with* $\alpha = \beta = 1/2$ *and fixed* $\tau = 1$, *the conditional prior distribution in Theorem 1 satisfies* $\mathbb{P}\big(\kappa_{t+1} < \varepsilon|\{\kappa_s\}_{s\le t}, \phi\big) \to 1$ *as* $\kappa_t \to 0$ *for any* $\varepsilon \in (0,1)$ *and fixed* $\phi \neq 0$.

The mass of the conditional prior distribution for κ_{t+1} concentrates near zero—corresponding to minimal shrinkage of signals—when κ_t is near zero, so the shrinkage behavior at time t informs the (prior) shrinkage behavior at time $t+1$.

We similarly characterize the posterior distribution of κ_{t+1} given $\{\kappa_s\}_{s\le t}$ in the following theorem, which extends the results of Datta and Ghosh (2013) to the dynamic setting.

THEOREM 3. *Under the likelihood* $[y_t|\omega_t] \overset{indep}{\sim} N(\omega_t, 1)$, *the prior* (2), *and the dynamic horseshoe process* (4) *with* $\alpha = \beta = 1/2$ *and fixed* $\phi \neq 0$, *the posterior distribution of* κ_{t+1} *given* $\{\kappa_s\}_{s\le t}$ *satisfies the following properties:*

(a) For any fixed $\varepsilon \in (0,1)$, $\mathbb{P}\big(\kappa_{t+1} > 1-\varepsilon\big|y_{t+1}, \{\kappa_s\}_{s\le t}, \phi, \tau\big) \to 1$ *as* $\gamma_t \to 0$ *uniformly in* $y_{t+1} \in \mathbb{R}$, *where* $\gamma_t = \tau^{2(1-\phi)}\left[(1-\kappa_t)/\kappa_t\right]^{\phi}$.

(b) For any fixed $\varepsilon \in (0,1)$ *and* $\gamma_t < 1$, $\mathbb{P}\big(\kappa_{t+1} < \varepsilon\big|y_{t+1}, \{\kappa_s\}_{s\le t}, \phi, \tau\big) \to 1$ *as* $|y_{t+1}| \to \infty$.

Theorem 3(a) demonstrates that the posterior mass of $[\kappa_{t+1}|\{\kappa_s\}_{s\le t}]$ concentrates near one as $\tau \to 0$, as in the non-dynamic horseshoe, but also as $\kappa_t \to 1$. Therefore, the dynamic horseshoe process provides an additional mechanism for shrinkage of noise, besides the global scale parameter τ, via the previous shrinkage parameter κ_t. Moreover, Theorem 3(b) shows that, despite the additional shrinkage capabilities, the posterior mass of $[\kappa_{t+1}|\{\kappa_s\}_{s\le t}]$ concentrates near zero for large absolute signals $|y_{t+1}|$, which indicates responsiveness of the dynamic horseshoe process to large signals analogous to the static horseshoe prior.

When $|\phi| < 1$, the log-variance process $\{h_t\}$ is stationary, which implies $\{\kappa_t\}$ is stationary. In Figure 2, we plot a simulation-based estimate of the stationary distribution of κ_t for various values of ϕ under the dynamic horseshoe process. The stationary distribution of κ_t is similar to the static horseshoe distribution ($\phi = 0$) for $\phi < 0.5$, while for large values of ϕ the distribution becomes more peaked at zero (less shrinkage of ω_t) and one (more shrinkage of ω_t). The result is intuitive: larger $|\phi|$ corresponds to greater persistence in shrinkage behavior, so marginally we expect states of aggressive shrinkage or little shrinkage.

2.2. *Scale Mixtures via Pólya-Gamma Processes*

For efficient computations, we develop a parameter expansion of model (3) using a conditionally Gaussian representation for η_t. In doing so, we may incorporate Gaussian models—and accompanying sampling algorithms—for dependent data in (3). Given a conditionally Gaussian parameter expansion, a Gibbs sampler for

(3) proceeds as follows: (i) draw the log-variances h_t, for which the conditional prior (3) is Gaussian, and (ii) draw the parameters in μ and ψ_t, for which the conditional likelihood (3) is Gaussian. For the log-variance sampler, we represent the likelihood for h_t on the log-scale and approximate the resulting distribution using a known discrete mixture of Gaussian distributions (see Section 5). This approach is popular in SV modeling (e.g., Kim et al., 1998), which is analogous to the dynamic shrinkage process in (4). Importantly, the proposed parameter expansion inherits the computational complexity of the samplers for h_t and ψ_t: for the dynamic shrinkage processes in (4), the proposed parameter expansion implies that the log-variance $\{h_t\}_{t=1}^T$ is a Gaussian dynamic linear model, and therefore $\{h_t\}_{t=1}^T$ may be sampled jointly in $\mathcal{O}(T)$ computations (see Section 5).

The proposed algorithm requires a parameter expansion of $\eta_t \sim Z(\alpha, \beta, 0, 1)$ in (4) as a scale mixture of Gaussian distributions. The representation of a Z-distribution as a mean-variance scale mixtures of Gaussian distributions is due to Barndorff-Nielsen et al. (1982). For implementation, we build on the framework of Polson et al. (2013), who propose a Pólya-Gamma scale mixture of Gaussians representation for Bayesian logistic regression. A *Pólya-Gamma* random variable ξ with parameters $b > 0$ and $c \in \mathbb{R}$, denoted $\xi \sim \mathrm{PG}(b, c)$, is an infinite convolution of Gamma random variables: $\xi \overset{D}{=} (2\pi^2)^{-1} \sum_{k=1}^{\infty} g_k \left[(k - 1/2)^2 - c^2/(4\pi^2)\right]^{-1}$ where $g_k \overset{iid}{\sim} \mathrm{Gamma}(b, 1)$. Properties of Pólya-Gamma random variables may be found in Barndorff-Nielsen et al. (1982) and Polson et al. (2013). Our interest in Pólya-Gamma random variables derives from their role in representing the Z-distribution as a mean-variance scale mixture of Gaussians.

THEOREM 4. *The random variable $\eta \sim Z(\alpha, \beta, 0, 1)$, or equivalently $\eta = \log(\lambda^2)$ with $\lambda^2 \sim IB(\beta, \alpha)$, is a mean-variance scale mixture of Gaussian distributions with $[\eta|\xi] \sim N\left(\xi^{-1}[\alpha - \beta]/2, \xi^{-1}\right)$ and $[\xi] \sim PG(\alpha + \beta, 0)$. Moreover, the conditional distribution of ξ is $[\xi|\eta] \sim PG(\alpha + \beta, \eta)$.*

When $\alpha = \beta$, the Z-distribution is symmetric and $\mathbb{E}[\eta|\xi] = 0$. Polson et al. (2013) propose a sampling algorithm for Pólya-Gamma random variables, which is extremely efficient when $b = 1$. In our setting, this corresponds to $\alpha + \beta = 1$, for which the horseshoe prior is the prime example. Importantly, this representation allows us to construct an efficient sampling algorithm that combines an $\mathcal{O}(T)$ sampling algorithm for the log-volatilities $\{h_t\}_{t=1}^T$ with a Pólya-Gamma sampler for the mixing parameters.

3. Bayesian Trend Filtering with Dynamic Shrinkage Processes

Dynamic shrinkage processes are particularly appropriate for dynamic linear models (DLMs). DLMs, such as model (1), combine an observation equation, which relates the observed data to latent state variables, and an evolution equation, which allows the state variables—and therefore the conditional mean of the data—to be dynamic. By construction, DLMs contain many parameters, and therefore may benefit from structured regularization. The proposed dynamic shrinkage processes offer such regularization, and unlike existing methods, do so adaptively.

The DLM (1) may be generalized to a Dth order random walk model: $\Delta^D \beta_{t+1} = \omega_t$, for $D \in \mathbb{Z}^+$ and $\beta_{t+1} = \omega_t \sim N(0, \tau^2 \lambda_t^2)$ for $t = 0, \ldots, D-1$, where $D \in \mathbb{Z}^+$ is the degree of differencing. We refer to model (1) as a *Bayesian trend filtering* (BTF) model, with various choices available for the distribution of the innovation standard deviations, $(\tau\lambda_t)$. The aggressive shrinkage of the horseshoe prior forces small values of $|\omega_t| = |\Delta^D \beta_{t+1}|$ toward zero, while the responsiveness of the horseshoe prior permits large values of $|\Delta^D \beta_{t+1}|$. When $D = 2$, model (1) will shrink the conditional mean β_t toward a piecewise linear function with breakpoints determined adaptively, while allowing large absolute changes in the slopes. Further, using the *dynamic* horseshoe process, the shrinkage effects induced by λ_t are time-dependent, which provides localized adaptability to regions with rapidly- or slowly-changing features. Following Carvalho et al. (2010) and Polson and Scott (2012b), we assume a half-Cauchy prior for the global scale parameter $\tau \sim C^+(0, \sigma_\epsilon/\sqrt{T})$, in which we scale by the observation error variance and the sample size (Piironen and Vehtari, 2016). Using Pólya-Gamma mixtures, the implied conditional prior on $\mu = \log(\tau^2)$ is $[\mu|\sigma_\epsilon, \xi_\mu] \sim N(\log \sigma_\epsilon^2 - \log T, \xi_\mu^{-1})$ with $\xi_\mu \sim \text{PG}(1, 0)$. We include the details of the Gibbs sampling algorithm for model (1) in Section 5, which is notably *linear* in the number of time points, T: the full conditional posterior precision matrices for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_T)'$ and $\boldsymbol{h} = (h_1, \ldots, h_T)'$ are D-banded and tridiagonal, respectively, which admit highly efficient $\mathcal{O}(T)$ back-band substitution sampling algorithms (see Section C of the supplement for empirical evidence).

3.1. *Bayesian Trend Filtering: Simulations*

To assess the performance of the Bayesian trend filtering (BTF) model (1) with dynamic horseshoe innovations (**BTF-DHS**), we compared the proposed methods to several competitive alternatives using simulated data. We considered the following variations on BTF model (1): normal-inverse-Gamma (**BTF-NIG**) innovations via $\tau^{-2} \sim \text{Gamma}(0.001, 0.001)$ with $\lambda_t = 1$; and (static) horseshoe priors for the innovations (**BTF-HS**) via $\tau, \lambda_t \overset{iid}{\sim} C^+(0, 1)$. In addition, we include the (non-Bayesian) trend filtering model of Tibshirani (2014) implemented using the `R` package `genlasso` (Arnold and Tibshirani, 2014), for which the regularization tuning parameter is chosen using cross-validation (**Trend Filtering**). For all trend filtering models, we select $D = 2$, but the relative performance is similar for $D = 1$. Among non-trend filtering models, we include a smoothing spline estimator implemented via `smooth.spline()` in `R` (**Smoothing Spline**); the wavelet-based estimator of Abramovich et al. (1998) (**BayesThresh**) implemented in the `wavethresh` package (Nason, 2016); and the nested Gaussian Process (**nGP**) model of Zhu and Dunson (2013), which relies on a state space model framework for efficient computations, comparable to—but empirically less efficient than—the BTF model (1).

We simulated 100 data sets from the model $y_t = y_t^* + \epsilon_t$, where y_t^* is the true function and $\epsilon_t \overset{indep}{\sim} N(0, \sigma_*^2)$. We use the following true functions y_t^* from Donoho and Johnstone (1994): *Doppler*, *Bumps*, *Blocks*, and *Heavisine*, implemented in the `R` package `wmtsa` (Constantine and Percival, 2016). The noise variance σ_*^2 is determined by selecting a root-signal-to-noise ratio (RSNR) and computing $\sigma_* =$

$\mathrm{sd}(y_t^*)/\mathrm{RSNR}$, where $\mathrm{sd}(y_t^*)$ is the sample standard deviation of $\{y_t^*\}_{t=1}^T$. As in Zhu and Dunson (2013), we select RSNR $= 7$ and use a moderate length time series, $T = 128$.

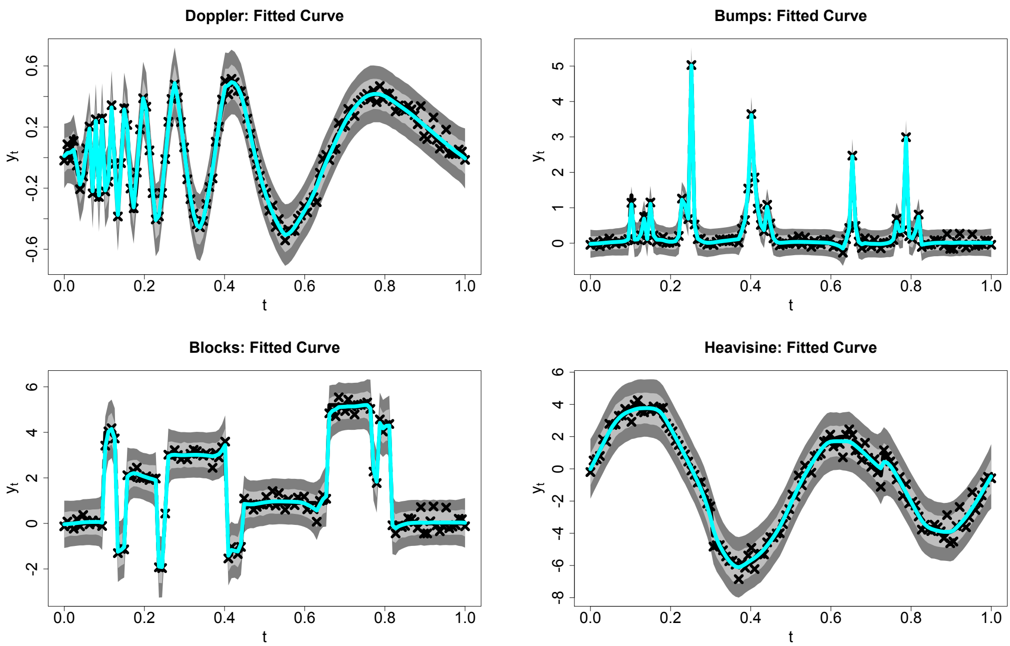

Fig. 3. Fitted curves for simulated data with $T = 128$ and RSNR $= 7$. Each panel includes the simulated observations (x-marks), the posterior expectations of β_t (cyan), and the 95% pointwise HPD credible intervals (light gray) and 95% simultaneous credible bands (dark gray) for the posterior predictive distribution of $\{y_t\}$ under BTF-DHS model (1). The proposed estimator, as well as the uncertainty bands, accurately capture both slowly- and rapidly-changing behavior in the underlying functions.

In Figure 3, we provide an example of each true curve y_t^*, together with the proposed BTF-DHS posterior expectations and credible bands. Notably, the Bayesian trend filtering model (1) with dynamic horseshoe innovations provides an exceptionally accurate fit to each data set. Importantly, the posterior expectations and the posterior credible bands adapt to both slowly- and rapidly-changing behavior in the underlying curves. The implementation is also efficient: the computation time for 10,000 iterations of the Gibbs sampling algorithm, implemented in `R` (on a MacBook Pro, 2.7 GHz Intel Core i5), is about 45 seconds.

To compare the aforementioned procedures, we compute the root mean squared errors $\mathrm{RMSE}(\hat{y}) = \sqrt{\frac{1}{T}\sum_{t=1}^T (y_t^* - \hat{y}_t)^2}$ for all estimators $\hat{y}$ of the true function, y^*. The results are displayed in Figure 4. The proposed BTF-DHS implementation substantially outperforms all competitors, especially for rapidly-changing curves (Doppler and Bumps). The exceptional performance of BTF-DHS is paired with comparably small variability of RMSE, especially relative to non-dynamic horseshoe model (BTF-HS). Interestingly, the magnitude and variability of the RMSEs for BTF-DHS are related to the AR(1) coefficient, ϕ: the 95% HPD intervals (corresponding to Figure 3) are $(0.77, 0.97)$ (Doppler), $(0.81, 0.97)$ (Bumps), $(0.76, 0.96)$ (Blocks), and $(-0.04, 0.74)$ (Heavisine). For the smoothest function, Heavisine, there is less separation among the estimators. Nonetheless, BTF-DHS performs the best, even though the HPD interval for ϕ is wider and contains zero.

We are also interested in uncertainty quantification, and in particular how the dynamic horseshoe prior compares to the horseshoe prior. We compute the mean credible intervals widths $\text{MCIW} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\beta}_t^{(97.5)} - \hat{\beta}_t^{(2.5)})$ where $\hat{\beta}_t^{(97.5)}$ and $\hat{\beta}_t^{(2.5)}$ are the 97.5% and 2.5% quantiles, respectively, of the posterior distribution of β_t in (1) for the BTF-DHS and BTF-HS. The results are in Figure 5. The dynamic horseshoe provides massive reductions in MCIW, again in all cases except for Heavisine, for which the methods perform similarly. Therefore, in addition to more accurate point estimation (Figure 4), the BTF-DHS model produces significantly tighter credible intervals—while maintaining the correct nominal (frequentist) coverage.

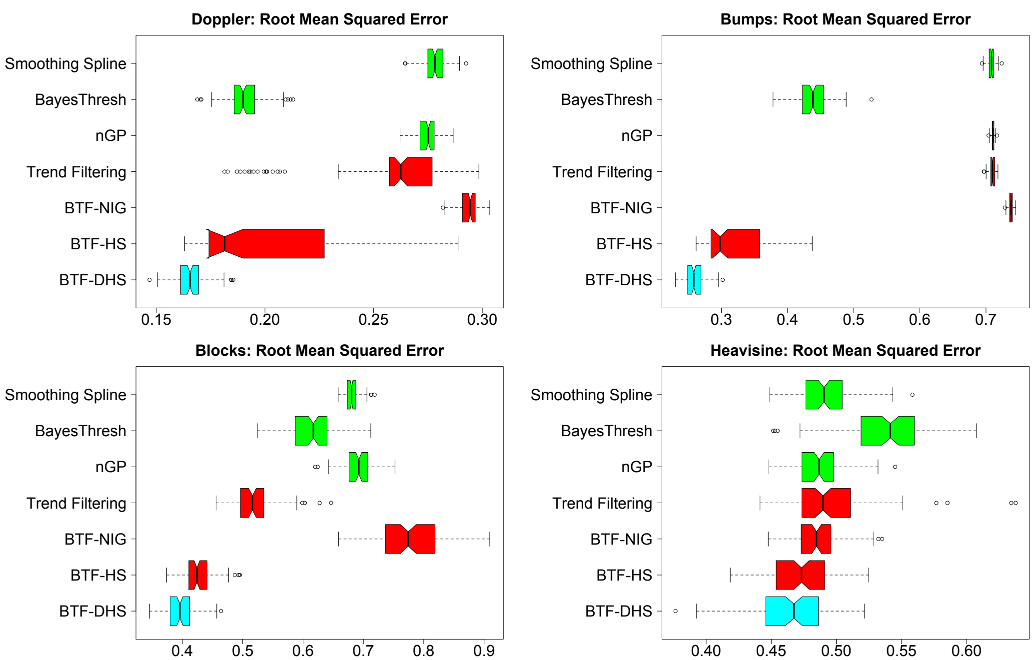

Fig. 4. Root mean squared errors for simulated data with $T = 128$ and RSNR $= 7$. Non-overlapping notches indicate significant differences between medians. The Bayesian trend filtering (BTF) estimators differ in their innovation distributions, which determines the shrinkage behavior of the second order differences: normal-inverse-Gamma (NIG), horseshoe (HS), and dynamic horseshoe (DHS).

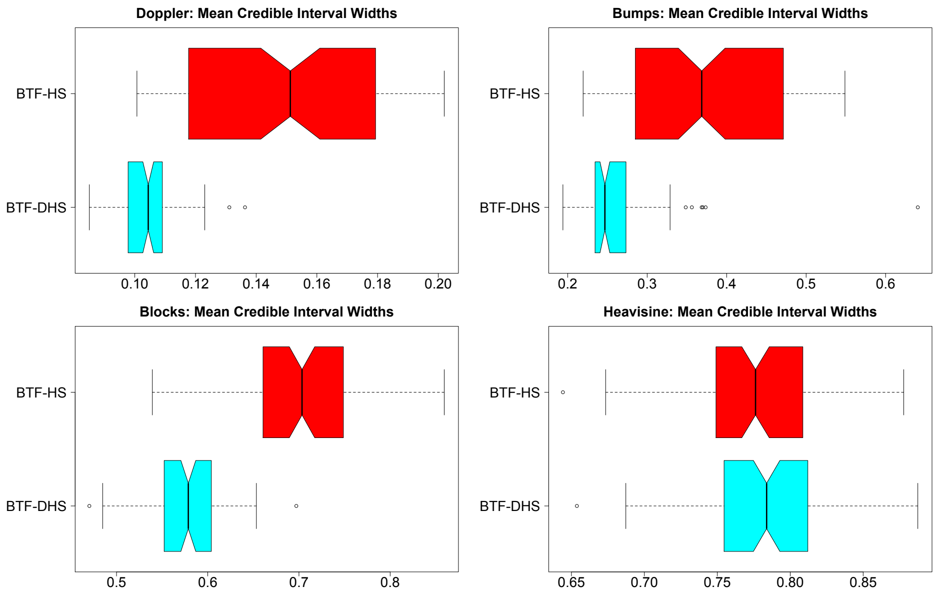

Fig. 5. Mean credible interval widths for simulated data with $T = 128$ and RSNR $= 7$. Non-overlapping notches indicate significant differences between medians. The Bayesian trend filtering (BTF) estimators differ in their innovation distributions, which determines the shrinkage behavior of the second order differences: normal-inverse-Gamma (NIG), horseshoe (HS), and dynamic horseshoe (DHS).

3.2. Bayesian Trend Filtering: Application to CPU Usage Data

To demonstrate the adaptability of the dynamic horseshoe process for model (1), we consider the CPU usage data in Figure 1. The data exhibit substantial complexity: an overall smooth intraday trend but with multiple irregularly-spaced jumps, and an increase in volatility from 16:00-18:00. Our goal is to provide an accurate measure of the trend, including jumps, with appropriate uncertainty quantification. For this purpose, we employ the BTF-DHS model (1), with a Gaussian AR(1) model on $\log \sigma_t^2$. For the additional sampling of the stochastic volatility parameters, we use the algorithm of Kastner and Frühwirth-Schnatter (2014) implemented in the `R` package `stochvol` (Kastner, 2016).

We augment the simulation study of Section 3.1 with an out-of-sample comparison for the CPU usage data. We fit each model using 90% ($T = 1296$) of the data selected randomly for training and the remaining 10% ($T = 144$) for testing, which was repeated 100 times. Models were compared using RMSE and MCIW.

Unlike the simulation study in Section 3.1, the subsampled data are *not* equally spaced. We employ a model-based imputation scheme for the unequally-spaced data $y_{t_i}, i = 1, \ldots, T$, which is similar to Elerian et al. (2001) and valid for missing observations. We expand the operative data set to include missing observations along an equally-spaced grid, $t^* = 1, \ldots, T^*$, such that for each observation point i, $y_{t_i} = y_{t^*}$ for some t^*. Although $T^* \geq T$, possibly with $T^* \gg T$, all computations within the sampling algorithm, including the imputation sampling scheme for $\{y_{t^*} : t^* \ne t_i\}$, are linear in the number of (equally-spaced) time points, T^*. Therefore, we may apply the same Gibbs sampling algorithm as before, with the additional step of drawing $y_{t^*} \overset{indep}{\sim} N(\beta_{t^*}, \sigma^2_{t^*})$ for each unobserved $t^* \ne t_i$. Implicitly, this procedure assumes that the unobserved points are missing at random, which is satisfied by the aforementioned subsampling scheme.

The results of the out-of-sample estimation study are displayed in Figure 6. The BTF procedures are notably superior to the non-Bayesian trend filtering and smoothing spline estimators, and, as with the simulations of Section 3.1, the proposed BTF-DHS model substantially outperforms all competitors. Importantly, the significant reduction in MCIW by BTF-DHS indicates that the posterior credible intervals for the out-of-sample points y_{t^*} are substantially tighter for our method. By reducing uncertainty—while maintaining the approximately correct nominal (frequentist) coverage—the proposed BTF-DHS model provides greater power to detect local features. In addition, the MCMC for the BTF-DHS is fast, despite the imputation procedure: 10,000 iterations runs in about 80 seconds (in `R` on a MacBook Pro, 2.7 GHz Intel Core i5).

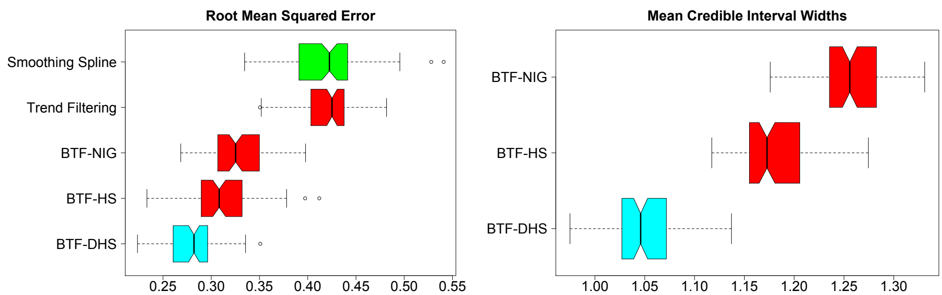

Fig. 6. Root mean squared error (**left**) and mean credible interval widths (**right**) for out-of-sample CPU usage data. Non-overlapping notches indicate significant differences between medians. The Bayesian trend filtering (BTF) estimators differ in their innovation distributions, which determines the shrinkage behavior of the second order differences: normal-inverse-Gamma (NIG), horseshoe (HS), and dynamic horseshoe (DHS).

4. Joint Shrinkage for Time-Varying Parameter Models

Dynamic shrinkage processes are appropriate for multivariate time series and functional data models that may benefit from locally adaptive shrinkage properties. As outlined in Dangl and Halling (2012), models with *time-varying parameters* are particularly important in financial and economic applications, due to the inherent structural changes in regulations, monetary policy, market sentiments, and macroeconomic interrelations that occur over time. Consider the following time-varying parameter regression model with multiple dynamic predictors $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{p,t})'$:

$$\begin{cases} y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + \epsilon_t, & [\epsilon_t|\sigma_\epsilon] \overset{indep}{\sim} N(0, \sigma_\epsilon^2) \\ \Delta^D \boldsymbol{\beta}_{t+1} = \boldsymbol{\omega}_t, & [\omega_{j,t}|\tau_0, \{\tau_k\}, \{\lambda_{k,s}\}] \overset{indep}{\sim} N(0, \tau_0^2\tau_j^2\lambda_{j,t}^2) \end{cases} \tag{6}$$

where $\boldsymbol{\beta}_t = (\beta_{1,t}, \ldots, \beta_{p,t})'$ is the vector of dynamic regression coefficients and $D \in \mathbb{Z}^+$ is the degree of differencing. Model (6) is also a (discretized) *concurrent functional linear model* (e.g., Ramsay and Silverman, 2005) and a *varying-coefficient model* (Hastie and Tibshirani, 1993) in the index t, and therefore is broadly applicable. The prior for the innovations $\omega_{j,t}$ incorporates three levels of global-local shrinkage: a global shrinkage parameter τ_0, a predictor-specific shrinkage parameter τ_j, and a predictor- and time-specific local shrinkage parameter $\lambda_{j,t}$. Relative to existing time-varying parameter regression models, our approach incorporates an additional layer of dynamic dependence: not only are the parameters time-varying, but also the *relative influence* of the parameters is time-varying via the shrinkage parameters—which are dynamically dependent themselves.

To provide jointly localized shrinkage of the dynamic regression coefficients $\{\beta_{j,t}\}$ analogous to the Bayesian trend filtering model of Section 3, we extend (4) to allow for multivariate time dependence via a vector autoregression (VAR) on the log-variance:

$$\begin{cases} [\omega_{j,t}|\tau_0, \{\tau_k\}, \{\lambda_{k,s}\}] \overset{indep}{\sim} N(0, \tau_0^2\tau_j^2\lambda_{j,t}^2) & \\ h_{j,t} = \log(\tau_0^2\tau_j^2\lambda_{j,t}^2), & j = 1, \ldots, p, t = 1, \ldots, T \\ \boldsymbol{h}_{t+1} = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, & \eta_{j,t} \overset{iid}{\sim} Z(\alpha, \beta, 0, 1) \end{cases} \tag{7}$$

where $\boldsymbol{h}_t = (h_{1,t}, \ldots, h_{p,t})'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$, $\boldsymbol{\eta}_t = (\eta_{1,t}, \ldots, \eta_{p,t})'$, and $\boldsymbol{\Phi}$ is the $p \times p$ VAR coefficient matrix. We assume $\boldsymbol{\Phi} = \mathrm{diag}\,(\phi_1, \ldots, \phi_p)$ for simplicity, but non-diagonal extensions are available. Contemporaneous dependence may be introduced via a copula model for $\boldsymbol{\eta}_t$, but may reduce computational and MCMC efficiency. As in the univariate setting, we use Pólya-Gamma mixtures for the log-variance evolution errors, $[\eta_{j,t}|\xi_{j,t}] \overset{indep}{\sim} N(\xi_{j,t}^{-1}[\alpha - \beta]/2, \xi_{j,t}^{-1})$ with $\xi_{j,t} \overset{iid}{\sim} \mathrm{PG}(\alpha + \beta, 0)$ and $\alpha = \beta = 1/2$. We augment model (7) with half-Cauchy priors for the predictor-specific and global parameters, $\tau_j \overset{indep}{\sim} C^+(0, 1)$ and $\tau_0 \sim C^+(0, \sigma_\epsilon/\sqrt{Tp})$.

4.1. Time-Varying Parameter Models: Simulations

We conducted a simulation study to evaluate competing variations of the time-varying parameter regression model (6), in particular relative to the proposed dynamic shrinkage process (**DHS**) in (7). Similar to the simulations of Section 3.1, we focus on the distribution of the innovations, $\omega_{j,t}$, and again include the normal-inverse-Gamma (**NIG**) and the (static) horseshoe (**HS**) as competitors, in each case selecting $D = 1$. We also include Belmonte et al. (2014), which uses the Bayesian Lasso as a prior on the innovations (**BL**). Lastly, we include Kalli and Griffin (2014), which offers an alternative approach for dynamic shrinkage (**KG**). Among models with non-dynamic regression coefficients, we include a lasso regression (Tibshirani, 1996) and an ordinary linear regression. These non-dynamic methods were non-competitive and are excluded from the figures.

We simulated 100 data sets of length $T = 200$ from the model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t^* + \epsilon_t$, where the $p = 20$ predictors are $x_{1,t} = 1$ and $x_{j,t} \sim N(0,1)$ for $j > 2$, and $\epsilon_t \sim N(0, \sigma_*^2)$ independently. We also consider autocorrelated predictors $x_{j,t}$ in Section D.2 of the supplement with similar results. The true regression coefficients $\boldsymbol{\beta}_t^* = (\beta_{1,t}^*, \ldots, \beta_{p,t}^*)'$ are the following: $\beta_{1,t}^* = 2$ is constant; $\beta_{2,t}^*$ is piecewise constant with $\beta_{2,t}^* = 0$ everywhere except $\beta_{2,t}^* = 2$ for $t = 41, \ldots, 80$ and $\beta_{2,t}^* = -2$ for $t = 121, \ldots, 160$; $\beta_{3,t}^* = \frac{1}{\sqrt{100}} \sum_{s=1}^t Z_s$ with $Z_s \stackrel{iid}{\sim} N(0,1)$ is a scaled random walk for $t \leq 100$ and $\beta_{3,t}^* = 0$ for $t > 100$; and $\beta_{j,t}^* = 0$ for $j = 4, \ldots, p = 20$. The predictor set contains a variety of functions: a constant nonzero function, a locally constant function, a slowly-varying function that thresholds to zero for $t > 100$, and 17 true zeros. The noise variance σ_*^2 is determined by selecting a root-signal-to-noise ratio (RSNR) and computing $\sigma_* = \mathrm{sd}(y_t^*)/\mathrm{RSNR}$, where $y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta}_t^*$ and $\mathrm{sd}(y_t^*)$ is the sample standard deviation of $\{y_t^*\}_{t=1}^T$. We select $\mathrm{RSNR} = 3$.

We evaluate competing methods using RMSEs for both y_t^* and $\boldsymbol{\beta}_t^*$ defined by $\mathrm{RMSE}(\hat{y}) = \sqrt{\frac{1}{T}\sum_{t=1}^T (y_t^* - \hat{y}_t)^2}$ and $\mathrm{RMSE}(\hat{\boldsymbol{\beta}}) = \sqrt{\frac{1}{Tp}\sum_{t=1}^T \sum_{j=1}^p (\beta_{j,t}^* - \hat{\beta}_{j,t})^2}$ for all estimators $\hat{\boldsymbol{\beta}}_t$ of the true regression functions, $\boldsymbol{\beta}_t^*$ with $\hat{y}_t = \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t$. The results are displayed in Figure 7. The proposed BTF-DHS model substantially outperforms the competitors in both recovery of the true regression functions, $\beta_{j,t}^*$ and estimation of the true curves, y_t^*. Our closest competitor is Kalli and Griffin (2014), which also uses dynamic shrinkage, yet is less accurate in estimating the regression coefficients $\beta_{j,t}^*$ and the fitted values y_t^*. In addition, our MCMC algorithm is vastly more efficient: for 10,000 MCMC iterations, the Kalli and Griffin (2014) sampler ran for 3 hours and 40 minutes (using Matlab code from Professor Griffin's website), while our algorithm completed in 6 minutes (on a MacBook Pro, 2.7 GHz Intel Core i5).

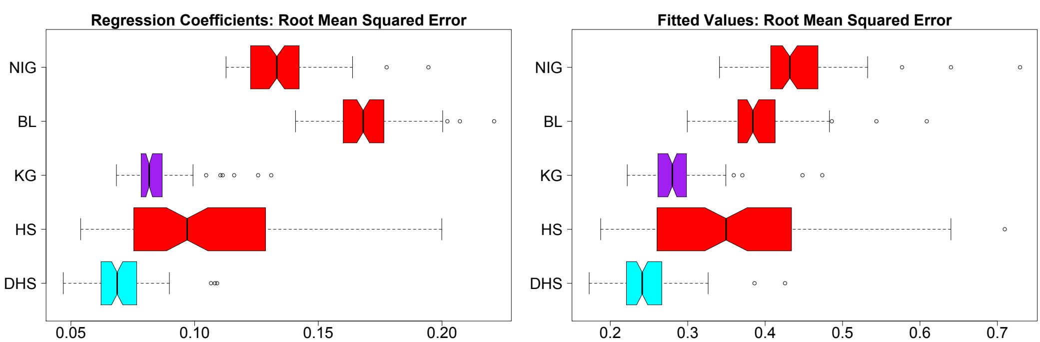

Fig. 7. Root mean squared errors for the regression coefficients, $\beta_{j,t}^*$ (**left**) and the true curves, $y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta}_t^*$ (**right**) for simulated data. Non-overlapping notches indicate significant differences between medians.

4.2. Time-Varying Parameter Models: The Fama-French Asset Pricing Model

Asset pricing models commonly feature highly structured factor models to parsimoniously model the co-movement of stock returns. Such fundamental factor models identify common risk factors among assets, which may be treated as exogenous predictors in a time series regression. Popular approaches include the one-factor Capital Asset Pricing Model (CAPM, Sharpe, 1964) and the three-factor Fama-French model (FF-3, Fama and French, 1993). Recently, the five-factor Fama-French model (FF-5, Fama and French, 2015) was proposed as an extension of FF-3 to incorporate additional common risk factors. However, outstanding questions remain regarding which, and how many, factors are necessary. Importantly, an attempt to address these questions must consider the dynamic component: the relevance of individual factors may change over time, particularly for different assets.

We apply model (6) to extend these fundamental factor models to the dynamic setting, in which the factor loadings are permitted to vary—perhaps rapidly—over time. For further generality, we append the momentum factor of Carhart (1997) to FF-5 to produce a fundamental factor model with six factors and dynamic factor loadings. Importantly, the shrinkage towards sparsity induced by the dynamic horseshoe process allows the model to effectively select out unimportant factors, which also may change over time. As in Section 3.2, we modify model (6) to include stochastic volatility for the observation error.

To study various market sectors, we use weekly industry portfolio data from the website of Kenneth R. French, which provide the value-weighted return of stocks in the given industry. We focus on manufacturing (Manuf) and healthcare (Hlth). For a given industry portfolio, the response variable is the returns in excess of the risk free rate, $y_t = R_t - R_{F,t}$, with predictors $\boldsymbol{x}_t = (1, R_{M,t} - R_{F,t}, SMB_t, HML_t, RMW_t, CMA_t, MOM_t)'$, defined as follows: the *market risk factor*, $R_{M,t} - R_{F,t}$ is the return on the market portfolio $R_{M,t}$ in excess of the risk free rate $R_{F,t}$; the *size factor*, SMB_t (small minus big) is the difference in returns between portfolios of small and large market value stocks; the *value factor*, HML_t (high minus low) is the difference in returns between portfolios of high and low book-to-market value stocks; the *profitability factor*, RMW_t is the difference in returns between portfolios of robust and weak profitability stocks; the *investment factor*, CMA_t is the difference in returns between portfolios of stocks of low and high investment firms; and the *momentum factor*, MOM_t is the difference in returns between portfolios of stocks with high and low prior returns. These data are publicly available on Kenneth R. French's website, which provides additional details on the portfolios. We standardize all predictors and the response to have unit variance.

We conduct inference on the time-varying regression coefficients $\beta_{j,t}$ using simultaneous credible bands. Unlike pointwise credible intervals, simultaneous credible bands control for multiple testing, and may be computed as in Ruppert et al. (2003). Letting $B_{j,t}(\alpha)$ denote the $(1-\alpha)\%$ simultaneous credible band for predictor j at time t, we compute *Simultaneous Band Scores* (SimBaS; Meyer et al., 2015), $P_{j,t} = \min\{\alpha : 0 \notin B_{j,t}(\alpha)\}$. The SimBaS $P_{j,t}$ indicate the minimum level for which the simultaneous bands do not include zero, while controlling for multiple testing, and therefore may be used to detect which predictors j are important at time t. Globally, we compute *global Bayesian p-values* (GBPVs; Meyer et al., 2015), $P_j = \min_t\{P_{j,t}\}$ for each predictor j, which indicate whether or not a predictor is important at *any* time t. SimBaS and GBPVs have proven useful in functional regression models, but also are suitable for time-varying parameter regression models to identify important predictors. We validate the selection ability of SimBaS in Section D.1 of the supplementary material for the simulated data of Section 4.1.

In Figure 8, we plot the posterior expectation and credible bands for the time-varying regression coefficients and observation error stochastic volatility for the weekly manufacturing industry data, from 4/1/2007 - 4/1/2017 ($T = 522$); an analogous plot for the healthcare industry data is in the supplementary material (Section E). For the manufacturing industry, the important factors identified by the GBPVs at the 5% level are the market risk ($R_{M,t} - R_{F,t}$, GBPV = 0.000), investment (CMA_t, GBPV = 0.024), and momentum (MOM_t, GBPV = 0.019). However, the SimBaS $P_{j,t}$ for CMA_t and MOM_t are below 5% only for brief periods (red lines), which suggests that these important effects are intermittent. For the healthcare industry, the GBPVs identify market risk (GBPV = 0.001) and value (HML_t, GBPV = 0.023) as the only important factors. Notably, the only common factor flagged by GBPVs in both the manufacturing and healthcare industries under model (6) over this time period is the market risk. This result suggests that the aggressive shrinkage behavior of the dynamic shrinkage process is important in this setting, since several factors may be effectively irrelevant for some or all time points.

5. MCMC Sampling Algorithm and Computational Details

We design a Gibbs sampling algorithm for the dynamic shrinkage process. The sampling algorithm is both computationally and MCMC efficient, and builds upon two main components: (i) a log-variance sampling algorithm (Kastner and Frühwirth-Schnatter, 2014) augmented with a Pólya-Gamma sampler (Polson et al., 2013); and (ii) a Cholesky Factor Algorithm (CFA, Rue, 2001) for sampling the state variables in a dynamic linear model. Importantly, computations for each component are linear in the number of time points, which produces an efficient sampling algorithm.

The general sampling algorithm is as follows: (i) sample the dynamic shrinkage components (the log-volatilities $\{h_t\}$, the Pólya-Gamma mixing parameters $\{\xi_t\}$, the unconditional mean of log-variance μ, the AR(1) coefficient of log-variance ϕ, and the discrete mixture component indicators $\{s_t\}$); (ii) sample the state variables $\{\boldsymbol{\beta}_t\}$; and (iii) sample the observation error variance σ_ϵ^2. We provide details of the dynamic shrinkage process sampling algorithm in Section 5.1 and include the details for sampling steps (ii) and (iii) in the supplement (Section C).

5.1. *Efficient Sampling for the Dynamic Shrinkage Process*

Consider the (univariate) dynamic shrinkage process in (4) with the Pólya-Gamma parameter expansion of Theorem 4. We provide implementation details for the dynamic horseshoe process with $\alpha = \beta = 1/2$, but extensions to other cases are straightforward. The sampling framework of Kastner and Frühwirth-Schnatter

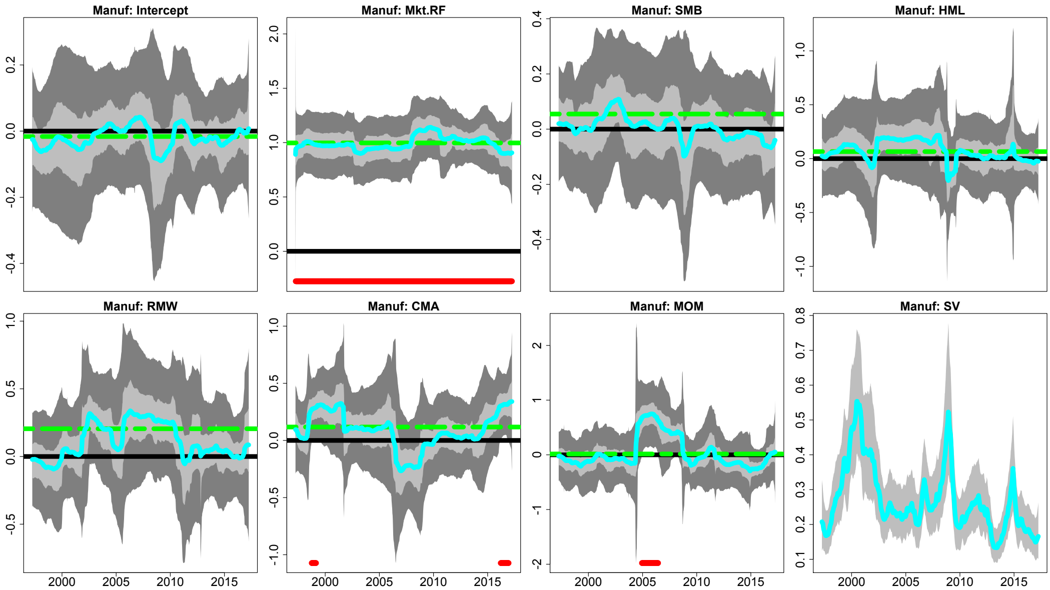

Fig. 8. Posterior expectations (cyan), 95% pointwise HPD credible intervals (light gray) and 95% simultaneous credible bands (dark gray) for $\beta_{j,t}$ and σ_t (bottom right) under the BTF-DHS model given by (6) and (7) for value-weighted manufacturing industry returns. The solid black line is zero, the dashed green line is the ordinary linear regression estimate, and the solid red line indicates periods for which the 95% simultaneous credible bands do not contain zero, or, equivalently, $P_{j,t}$ (SimBaS) is less than 0.05.

(2014) represents the likelihood for h_t in (2) on the log-scale, and approximates the ensuing $\log \chi^2_1$ distribution for the errors via a known discrete mixture of Gaussian distributions. In particular, let $\tilde{y}_t = \log(\omega_t^2 + c)$, where c is a small offset to avoid numerical issues. Conditional on the mixture component indicators s_t, the likelihood is $\tilde{y}_t \overset{indep}{\sim} N(h_t + m_{s_t}, v_{s_t})$ where m_i and $v_i, i = 1, \ldots, 10$ are the pre-specified mean and variance components of the 10-component Gaussian mixture provided in Omori et al. (2007). Under model (4), the evolution equation is $h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t$ with initialization $h_1 = \mu + \eta_0$ and innovations $[\eta_t|\xi_t] \overset{indep}{\sim} N(0, \xi_t^{-1})$ for $[\xi_t] \overset{iid}{\sim} \mathrm{PG}(1, 0)$. Note that model (3) provides a more general setting, which similarly may be combined with the Gaussian likelihood for $\tilde{y}_t$ above.

To sample $\boldsymbol{h} = (h_1, \ldots, h_T)$ jointly, we directly compute the posterior distribution of $\boldsymbol{h}$ and exploit the tridiagonal structure of the resulting posterior precision matrix. In particular, we equivalently have $\tilde{\boldsymbol{y}} \sim N(\boldsymbol{m} + \tilde{\boldsymbol{h}} + \tilde{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_v)$ and $\boldsymbol{D}_\phi \tilde{\boldsymbol{h}} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\xi)$, where $\boldsymbol{m} = (m_{s_1}, \ldots, m_{s_T})'$, $\tilde{\boldsymbol{h}} = (h_1 - \mu, \ldots, h_T - \mu)'$, $\tilde{\boldsymbol{\mu}} = (\mu, (1-\phi)\mu, \ldots, (1-\phi)\mu)'$, $\boldsymbol{\Sigma}_v = \mathrm{diag}\left(\{v_{s_t}\}_{t=1}^T\right)$, $\boldsymbol{\Sigma}_\xi = \mathrm{diag}\left(\{\xi_t^{-1}\}_{t=1}^T\right)$, and $\boldsymbol{D}_\phi$ is a lower triangular matrix with ones on the diagonal, $-\phi$ on the first off-diagonal, and zeros elsewhere. We sample from the posterior distribution of $\boldsymbol{h}$ by sampling from the posterior distribution of $\tilde{\boldsymbol{h}}$ and setting $\boldsymbol{h} = \tilde{\boldsymbol{h}} + \mu\boldsymbol{1}$ for $\boldsymbol{1}$ a T-dimensional vector of ones. The required posterior distribution is $\tilde{\boldsymbol{h}} \sim N\big(\boldsymbol{Q}_{\tilde{h}}^{-1}\boldsymbol{\ell}_{\tilde{h}}, \boldsymbol{Q}_{\tilde{h}}^{-1}\big)$, where $\boldsymbol{Q}_{\tilde{h}} = \boldsymbol{\Sigma}_v^{-1} + \boldsymbol{D}_\phi'\boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{D}_\phi$ is a tridiagonal symmetric matrix with diagonal elements $\boldsymbol{d}_0(\boldsymbol{Q}_{\tilde{h}})$ and first off-diagonal elements $\boldsymbol{d}_1(\boldsymbol{Q}_{\tilde{h}})$ defined as

$$
\begin{aligned}
\boldsymbol{d}_0(\boldsymbol{Q}_{\tilde{h}}) &= \Big[(v_{s_1}^{-1} + \xi_1 + \phi^2\xi_2), \ldots, (v_{s_{T-1}}^{-1} + \xi_{T-1} + \phi^2\xi_T), (v_{s_T}^{-1} + \xi_T)\Big], \\
\boldsymbol{d}_1(\boldsymbol{Q}_{\tilde{h}}) &= \big[(-\phi\xi_2), \ldots, (-\phi\xi_{T-1})\big], \text{ and} \\
\boldsymbol{\ell}_{\tilde{h}} &= \boldsymbol{\Sigma}_v^{-1}\big(\tilde{\boldsymbol{y}} - \boldsymbol{m} - \tilde{\boldsymbol{\mu}}\big) \\
&= \Big[\frac{\tilde{y}_1 - m_{s_1} - \mu}{v_{s_1}}, \frac{\tilde{y}_2 - m_{s_2} - (1-\phi)\mu}{v_{s_2}}, \ldots, \frac{\tilde{y}_T - m_{s_T} - (1-\phi)\mu}{v_{s_T}}\Big]'.
\end{aligned}
$$

Drawing from this posterior distribution is straightforward and efficient, using band back-substitution described in Kastner and Frühwirth-Schnatter (2014): (i) compute the Cholesky decomposition $\boldsymbol{Q}_{\tilde{h}} = \boldsymbol{L}\boldsymbol{L}'$, where $\boldsymbol{L}$ is lower triangle; (ii) solve $\boldsymbol{L}\boldsymbol{a} = \boldsymbol{\ell}_{\tilde{h}}$ for $\boldsymbol{a}$; and (iii) solve $\boldsymbol{L}'\tilde{\boldsymbol{h}} = \boldsymbol{a} + \boldsymbol{e}$ for $\tilde{\boldsymbol{h}}$, where $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}_T)$.

Conditional on the log-volatilities $\{h_t\}$, we sample the AR(1) evolution parameters: the log-innovation precisions $\{\xi_t\}$, the autoregressive coefficient ϕ, and the unconditional mean μ. The precisions are distributed $[\xi_t|\eta_t] \sim \mathrm{PG}(1, \eta_t)$ for $\eta_t = h_{t+1} - \mu - \phi(h_t - \mu)$, which we sample using Polson et al. (2013). The Pólya-Gamma sampler is efficient: using only exponential and inverse-Gaussian draws, Polson et al. (2013) construct an accept-reject sampler for which the probability of acceptance is uniformly bounded below at 0.99919, which does not require any tuning. Next, we assume the prior $[(\phi + 1)/2] \sim \mathrm{Beta}(a_\phi, b_\phi)$, which restricts $|\phi| < 1$ for stationarity, and sample from the full conditional distribution of ϕ using the slice sampler of Neal (2003). We select $a_\phi = 10$ and $b_\phi = 2$, which places most of the mass for the density of ϕ in $(0, 1)$ with a prior mean of $2/3$ and a prior mode of $4/5$ to reflect the likely presence of persistent volatility clustering. The prior for the global scale parameter is $\tau \sim C^+(0, \sigma_\epsilon/\sqrt{T})$, which implies $\mu = \log(\tau^2)$ is $[\mu|\sigma_\epsilon, \xi_\mu] \sim N(\log(\sigma_\epsilon^2/T), \xi_\mu^{-1})$ with $\xi_\mu \sim \mathrm{PG}(1, 0)$. Including the initialization $h_1 \sim N(\mu, \xi_0^{-1})$ with $\xi_0 \sim \mathrm{PG}(1, 0)$, the posterior distribution for μ is $\mu \sim N(Q_\mu^{-1}\ell_\mu, Q_\mu^{-1})$ with $Q_\mu = \xi_\mu + \xi_0 + (1-\phi)^2\sum_{t=1}^{T-1}\xi_t$ and $\ell_\mu = \xi_\mu\log(\sigma_\epsilon^2/T) + \xi_0 h_1 + (1-\phi)\sum_{t=1}^{T-1}\xi_t(h_{t+1} - \phi h_t)$. Sampling ξ_μ and ξ_0 follows the Pólya-Gamma sampling scheme above.

Finally, we sample the discrete mixture component indicators s_t. The discrete mixture probabilities are straightforward to compute: the prior mixture probabilities are the mixing proportions given by Omori et al. (2007) and the likelihood is $\tilde{y}_t \overset{indep}{\sim} N(h_t + m_{s_t}, v_{s_t})$; see Kastner and Frühwirth-Schnatter (2014) for details.

6. Discussion and Future Work

Dynamic shrinkage processes provide a computationally convenient and widely applicable mechanism for incorporating adaptive shrinkage and regularization into existing models. By extending a broad class of global-local shrinkage priors to the dynamic setting, the resulting processes inherit the desirable shrinkage behavior, but with greater time-localization. The success of dynamic shrinkage processes suggests that other priors may benefit from log-scale or other appropriate representations, with or without additional dependence modeling.

As demonstrated in Sections 3 and 4, dynamic shrinkage processes are particularly appropriate for DLMs, including trend filtering and time-varying parameter regression. In both settings, the DLMs with dynamic horseshoe innovations outperform all competitors in simulated data, and produce reasonable and interpretable results for real data applications. Dynamic shrinkage processes may be useful in other DLMs, such as for modeling seasonality or change points. Given the exceptional curve-fitting capabilities of the Bayesian trend filtering model (1) with dynamic horseshoe innovations (BTF-DHS), a natural extension would be to incorporate the BTF-DHS into more general additive, functional, or longitudinal data models in order to capture irregular or local curve features. Similarly, models (1) and (6), as well as the dependent shrinkage of (3), may be extended for multivariate responses to provide both contemporaneous and dynamic shrinkage.

Another promising area for applications of the proposed methodology is compressive sensing and signal processing, which commonly rely on approximations for estimation and prediction (e.g., Ziniel and Schniter, 2013; Wang et al., 2016). The linear time complexity of our MCMC algorithm for BTF-DHS may offer the computational scalability to provide full Bayesian inference, and perhaps improved prediction and uncertainty quantification, which is notably absent from the papers cited above.

References

Abramovich, F., Sapatinas, T., and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):725–749.

Armagan, A., Clyde, M., and Dunson, D. B. (2011). Generalized Beta mixtures of Gaussians. In *Advances in neural information processing systems*, pages 523–531.

Arnold, T. B. and Tibshirani, R. J. (2014). *genlasso: Path algorithm for generalized lasso problems.* R package version 1.3.

Barndorff-Nielsen, O., Kent, J., and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 145–159.

Belmonte, M. A., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94.

Bitto, A. and Frühwirth-Schnatter, S. (2016). Achieving shrinkage in a time-varying parameter model framework. *arXiv preprint arXiv:1611.01310.*

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

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

Chan, J. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012). Time varying dimension models. *Journal of Business & Economic Statistics*, 30(3):358–367.

Constantine, W. and Percival, D. (2016). *wmtsa: Wavelet Methods for Time Series Analysis*. R package version 2.0-2.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Datta, J. and Ghosh, J. K. (2013). Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Analysis*, 8(1):111–132.

Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Faulkner, J. R. and Minin, V. N. (2018). Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian analysis*, 13(1):225.

Figueiredo, M. A. (2003). Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1150–1159.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Griffin, J. E. and Brown, P. J. (2005). Alternative prior distributions for variable selection with very many more variables than observations. Technical report, University of Warwick, Centre for Research in Statistical Methodology.

Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 757–796.

Huber, F., Kastner, G., and Feldkircher, M. (2018). Should i stay or should i go? A latent threshold approach to large-scale mixture innovation models. *Journal of Applied Econometrics*.

James, N. A., Kejariwal, A., and Matteson, D. S. (2016). Leveraging cloud data to mitigate user experience from 'Breaking Bad'. In *2016 IEEE International Conference on Big Data*, pages 3499–3508. IEEE.

Kalli, M. and Griffin, J. E. (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the `R` package `stochvol`. *Journal of Statistical Software*, 69(5):1–30.

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

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009). ℓ_1 Trend Filtering. *SIAM review*, 51(2):339–360.

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, 29(1):43–59.

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5(2):369–411.

Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., and Morris, J. S. (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics*, 71(3):563–574.

Nakajima, J. and West, M. (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164.

Nason, G. (2016). *wavethresh: Wavelets Statistics and Transforms.* `R` package version 4.6.8.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, pages 705–741.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Piironen, J. and Vehtari, A. (2016). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. *arXiv preprint arXiv:1610.05559*.

Polson, N. G. and Scott, J. G. (2012a). Local shrinkage rules, Lévy processes and regularized regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):287–311.

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

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

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis.* Springer.

Rockova, V. and McAlinn, K. (2017). Dynamic variable selection with spike-and-slab process priors. *arXiv preprint arXiv:1708.00085.*

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):325–338.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression.* Number 12. Cambridge University Press.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Strawderman, W. E. (1971). Proper bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, 42(1):385–388.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323.

Uribe, P. V. and Lopes, H. F. (2017). Dynamic sparsity on dynamic regression models. *Manuscript, available at http://hedibert.org/wpcontent/uploads/2018/06/uribe-lopes-Sep2017. pdf.*

van der Pas, S., Kleijn, B., and van der Vaart, A. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8(2):2585–2618.

Wang, H., Yu, H., Hoy, M., Dauwels, J., and Wang, H. (2016). Variational Bayesian dynamic compressive sensing. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1421–1425. IEEE.

Zhu, B. and Dunson, D. B. (2013). Locally adaptive Bayes nonparametric regression via nested Gaussian processes. *Journal of the American Statistical Association*, 108(504):1445–1456.

Ziniel, J. and Schniter, P. (2013). Dynamic compressive sensing of time-varying signals via approximate message passing. *IEEE transactions on signal processing*, 61(21):5270–5284.