Dynamic Shrinkage Process in Volatility

Jason Cho, David S. Matteson

October 2023

Abstract

This paper presents a novel stochastic volatility model called Dynamic Shrinkage Process in Volatility (DSP-V). The key difference to the existing stochastic volatility model lies in the incorporation of global-local prior called Dynamic Shrinkage Process, allowing DSP-V to be dynamically adaptive to evolving processes. Unlike existing models, the strength of DSP-V is evident in its flexibility, as demonstrated by its superior or comparable performance when compared to established models such as Stochastic Volatility (SV), Markov Switching SV, Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), and Markov Switching GARCH, across diverse data generating scenarios. Notably, DSP-V exhibits resilience to model mispecification biases, making it a robust tool in financial modeling. In addition, DSP-V produces smoother and more interpretable estimate of log-volatility term h_t, facilitating a clearer understanding of underlying patterns and trends.

1 Introduction

Volatility modeling has become a pivotal topic in the realm of finance, demonstrating its significance through its diverse applications in asset pricing, portfolio management, and risk assessment. In its early days, the financial world assumed volatility to be constant. Pivotal work by Merton [1969] on portfolio selection and Black and Scholes [1973] on option pricing both assume volatility to be time invariant. It has now been widely acknowledged that volatility evolves over time, hence the name stochastic volatility. Several popular stochastic volatility models have emerged including the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle [1982] and later extended into the Generalized ARCH (GARCH) model by Bollerslev [1986]. Another notable model is the Stochastic Volatility (SV) model, which was explored by Hull and White [1987], Taylor [2008] and Melino and Turnbull [1990]. A key common assumption among these models is stationarity, or more specifically, constant unconditional volatility; while the conditional volatility is assumed to evolve over time, the unconditional volatility, on the other hand, is still assumed to be constant. This assumption becomes too restrictive when structural changes occur in the unconditional volatility. Many empirical (Dieobold [1986], French et al. [1987], Chou [1988], Poon and Taylor [1992], So et al. [1997], Su and Wang [2020]) as well as theoretical studies (Mikosch and Stărică [2004], Messow and Krämer [2013], Lamoureux and Lastrapes [1990]) point to the existence of structural changes in variety of financial and economic data sets. Therefore, the need for a more flexible stochastic volatility model has become evident.

A commonly employed method for addressing structural changes in the unconditional volatility is the incorporation of Markov switching into the existing models such as SV or (G)ARCH. Specifically, Markov Switching has been incorporated into the ARCH model by Hamilton and Susmel [1994] and Cai [1994], the GARCH model by Bauwens et al. [2010] and Gray [1996], and the SV model by So et al. [1998] and Hwang et al. [2004]. Such models allow parameters to come from one of several regimes, with transitions between regimes governed by a Markov process, directly modeling the structural breaks. However, a problem persists in this approach, as one needs to determine the number of regimes a priori, which can be a challenging task. Moreover, parameter estimations become even more challenging as we increase the number of regimes. Often, the number of regimes are set to 2 to 3 regimes denoting low to high volatility regimes.

We propose the model named Dynamic Shrinkage Process in Volatility (DSP-V). The model stands apart from existing models in its introduction of time-varying parameters, ω_t, with global-local priors called Dynamic Shrinkage Process (DSP) Kowal et al. [2019]. Rather than relying on a fixed number of parameters, the evolution of the process is assumed to be governed by time varying parameters ω_t. We then incorporate a suitable global-local shrinkage priors, on the variance term of ω_t to induce desirable locally adaptive shrinkage properties. The resulting model presents a flexible and adaptive solution for modeling stochastic volatility, capable of capturing both abrupt structural changes as well as gradual changes over time. In addition, the estimated volatility from DSP-V is less noisy when compared against the output from SV or GARCH type models. The absence of excessive noise in the estimates facilitates a clearer understanding of underlying patterns and trends. This smoother output not only aids in the identification of meaningful features but also contributes to a more transparent and interpretable model.

The paper is structured as follows. Section 2 provides an introduction to the model, its associated parameters, likelihood (Section 2.1), priors (Section 2.2), and brief overview of the Gibbs sampler (Section 2.3). The precise full conditional distributions for both in-sample and out-of-sample forecast are both detailed in Appendix A. Subsequently, an analysis of the estimated volatility by DSP-V is presented, comparing it with four commonly employed stochastic volatility models, namely the SV model, Markov Switching SV (MSSV), GARCH(1,1), and Markov Switching GARCH (MSGARCH), in Section 3. In Section 4, DSP-V is applied to three empirical data sets: weekly log-returns on S&P 500 index, weekly log-returns on EURO/USD exchange rate, and weekly changes in death tolls from COVID-19. The estimated log-volatility by DSP-V and SV are compared. Subsequent subsections delve into parameters from DSP-V: κ_t, μ, ϕ , which provide additional information about the underlying process.

2 Model

The Dynamic Shrinkage Process (DSP) by Kowal et al. [2019] introduces a locally adaptive Bayesian trend filter for modeling a time-varying process that exhibits both extended periods of stability and periods of dynamic changes via a conditionally Gaussian State Space Model. For a process $\{\beta_t\}$, they define the DSP of order d(usually equal to 1 or 2) as:

$$\Delta^d \beta_t = \begin{cases} \omega_t & \omega_t \sim N(0, \tau^2 \lambda_t^2), \\ & \textbf{or} \\ exp\left\{\frac{v_t}{2}\right\} \epsilon_t & \epsilon_t \sim N(0,1) \end{cases}$$

with

$$v_t = \mu + \psi_t + \eta_t \qquad \eta_t \overset{i.i.d}{\sim} Z(a, b, 0, 1),$$

where Z-distribution has the following density function:

$$f(z|a, b, \mu_z, \sigma_z) = (\sigma_z \beta(a, b))^{-1} \left(exp\left(\frac{z - \mu_z}{\sigma_z}\right)\right)^a \left(1 + exp\left(\frac{z - \mu_z}{\sigma_z}\right)\right)^{-(a+b)}.$$

Equivalently, $v_t = log(\tau^2 \lambda_t^2)$, thus may be thought of as the log-variance process of β_t. v_t consists of three parts. The global scale parameter μ represents the overall mean of the process. $\psi_t + \eta_t$, on the other hand, represents the local-scale parameter. ψ_t models the temporal dependence of the process. A few candidate models for ψ_t includes the Hidden Markov Models, linear regression, or spline. For computational efficiency, order 1 Autoregression (AR1) with lag is considered. Thus, $\psi_t := \phi(v_t - \mu)$. η_t corresponds to the i.i.d scale parameter, that dictates the shrinkage in the model. For illustration, let's consider the case $\psi_t = \mu = 0$. Then, $\omega_t \overset{i.i.d}{\sim} N(0, \tau\lambda_t^2)$, where $\lambda_t = exp(\frac{\eta_t}{2})$, $\eta_t \sim Z(a, b, 0, 1)$, $\tau = 1$. This is equivalent to imposing an inverted beta prior on λ_t. A few notable priors exist with various values on a and b. The specific prior we focus on in this paper is horseshoe prior Carvalho et al. [2010], by setting $a = b = \frac{1}{2}$. The benefit of having $Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$, or more broadly Z-distribution as prior on η_t lies in its shrinkage properties. Given the shrinkage parameter $\kappa_t := \frac{1}{1+var(\omega_t|\tau,\lambda_t)} = \frac{1}{1+exp(v_t)} = \frac{1}{1+\tau^2\lambda_t^2}$, an effective shrinkage prior should concentrate its values around 0 indicating minimal shrinkage and 1 indicating maximum shrinkage. As the name suggests, the expected value of the posterior shrinkage parameter $E(\kappa_t|y, \tau)$ for horseshoe prior induces 'horseshoe' or U shape on the shrinkage parameter, allocating most of its mass near 0 (minimal shrinkage) and 1 (maximal shrinkage). This advantage becomes particularly evident when comparing it to $\eta_t \sim N(0, \sigma_\eta^2)$, where the density for κ_t, unlike the Z-distribution, approaches 0 as $\kappa_t \to 0$ and $\kappa_t \to 1$. Thus, DSP is proposed in Kowal et al. [2019] where

$$v_{t+1} = \mu + \phi(v_t - \mu) + \eta_t \qquad \eta_t \overset{i.i.d}{\sim} Z(a, b, 0, 1).$$

We consider applying DSP model to explain changes in a stochastic volatility process. Let $\{y_t\}$ be the mean 0 observations. The Dynamic Shrinkage Process in Volatility Model (DSP-V),

$$y_t = \sigma_t \epsilon_t = exp\left\{\frac{h_t}{2}\right\} \qquad \epsilon_t \overset{i.i.d}{\sim} N(0, 1) \tag{1}$$

$$\Delta h_t = h_t - h_{t-1} = \omega_t \qquad \omega_t | v_t \sim N(0, e^{v_t}) \tag{2}$$

$$v_{t+1} = \mu + \phi(v_t - \mu) + \eta_t \qquad \eta_t \overset{i.i.d}{\sim} Z(a, b, 0, 1). \tag{3}$$

This is in context with the canonical first-order stochastic volatility SV(1) Model in which h_t is the first order autoregression and a constant innovation variance in volatility σ^2, that is

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t \qquad \eta_t \overset{i.i.d}{\sim} N(0, \sigma_\eta^2).$$

Compared to existing stochastic volatility models, DSP-V possesses two distinctive features. Primarily, it exhibits smooth yet locally adaptive estimate of h_t. Additionally, owing to locally adapted error variance on h_t, the credible region for h_t also exhibits locally adaptive property. Specifically, when substantial changes occur in h_t, it displays a wider credible region. Conversely, in the presence of little changes in h_t, the credible region shrinks. Characteristics of DSP-V are further explored and visualized using simulation data from several data generating processes as well as empirical data sets in Section 3 and 4.

2.1 Likelihood

Likelihood for DSP-V is analogous to the one specified in Kim et al. [1998]. Let $y_t^* := log(y_t^2)$ and define $y_{1:t}^*$ as $(y_1^*, y_2^*, \dots, y_t^*)'$, with T being the total number of observation. Instead of working directly with y, we will use the transformed data y^* and specify its likelihood. We also have four parameters $\mathbf{h} = (h_1, \dots, h_t)', \mathbf{v} = (v_1, \dots, v_t)', \mu$, and ϕ. Without any simplification, we have the following likelihood based on Bayes Theorem:

$$f(y^*|\mathbf{h}, \mathbf{v}, \mu, \phi) = f(y_1^*|h_1, v_1, \mu, \phi) \prod_{t=2}^{T} f(y_t^*|y_{1:t-1}^*, h_{1:t}, v_{1:t}, \mu, \phi)$$

As shown in equation 2, y_t^* only depends on the state variable h_t. Therefore,

$$f(y^*|\mathbf{h}, \mathbf{v}, \mu, \phi) = f(y^*|\mathbf{h}) = \prod_{t=1}^{T} f(y_t^*|h_t).$$

Instead of defining the likelihood on y directly, likelihood for y^* is defined:

$$y_t^* := log(y_t^2) = h_t + \epsilon_t^*,$$

where $\epsilon_t^* \sim log(\chi_1^2)$. We use 10-component Gaussian mixture distribution proposed in Omori et al. [2007] to approximate $log(\chi_1^2)$ distribution. The likelihood on y^* becomes:

$$f(y^*|\mathbf{h}, \mathbf{v}, \mu, \phi) = f(y^*|h) = \prod_{t=1}^{T} f(y_t^*|h_t) = \prod_{t=1}^{T} \sum_{i=1}^{10} p_i \mathcal{N}(y_t^*|h_t + \mu_i, \sigma_i).$$

Clearly, the likelihood is Gaussian, conditional on the mixture component at time t, we call $\mathbf{j} = (j_1, j_2, \dots, j_T)'$:

$$f(y_t^*|h_t, j_t = k) = \mathcal{N}(y_t^*|h_t + \mu_k, \sigma_k^2)$$

$$j_t \overset{i.i.d}{\sim} Categorical(p_1, \dots, p_{10})$$

Exact distribution for $(j_t, \mu_{j_t}, \sigma_{j_t})$ may be found in Omori et al. [2007]. Thus the variable $\mathbf{j}$ is introduced for efficient sampling. The likelihood on y^* with the parameter expansion becomes:

$$f(y^*|\mathbf{h}, \mathbf{v}, \mu, \phi, \mathbf{j}) = f(y^*|\mathbf{h}, \mathbf{j}) = \mathcal{N}(y^*|\mathbf{h} + \mu_j, I\sigma_j).$$

2.2 Priors

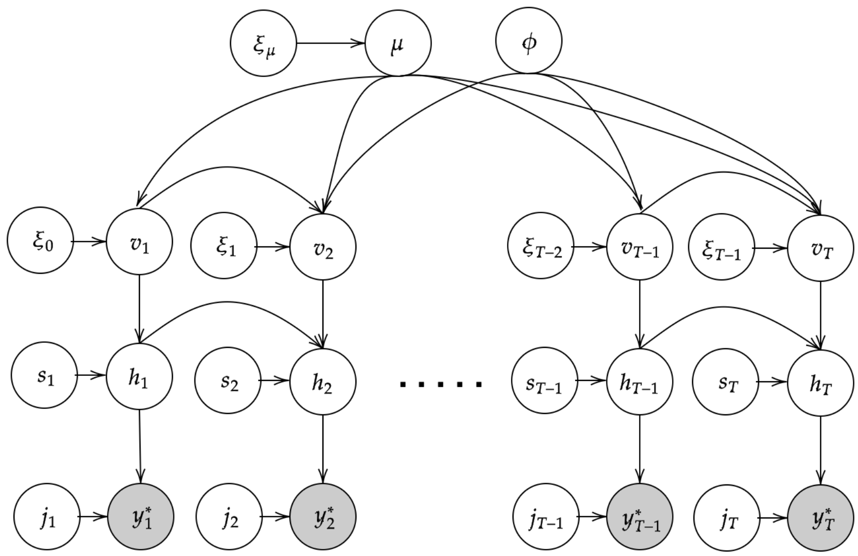


Figure 1: Graphical representations of Dynamic Shrinkage Process in Volatility (DSP-V) with parameter expansions. In addition to the existing parameters, $(\mathbf{h}, \mathbf{v}, \mu, \phi)$, parameters $(\mathbf{j}, \mathbf{s}, \boldsymbol{\xi}, \xi_\mu)$ are introduced for efficient Gibbs sampling.

In this section, priors on the parameters (h, v, μ, ϕ) are specified. For computational efficiency, we consider $\eta_t \sim Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$. As described in equation 2. The prior distribution for $\mathbf{h}$ depends on $\mathbf{v}$; $\mathbf{v}$ depends on μ and ϕ; and μ and ϕ are independent to all other variables. Thus, the priors on $(\mathbf{h}, \mathbf{v}, \mu, \phi)$ breaks up into four conditionally independent distributions:

$$\begin{aligned} f(\mathbf{h}, \mathbf{v}, \mu, \phi) &= f(\mathbf{h}|\mathbf{v}, \mu, \phi) f(\mathbf{v}|\mu, \phi) f(\mu|\phi) f(\phi) \\ &= f(\mathbf{h}|\mathbf{v}) f(\mathbf{v}|\mu, \phi) f(\mu) f(\phi) \end{aligned}$$

In addition to the exiting parameters, we also introduce four additional parameters $\mathbf{j} = (j_1, \ldots, j_T)'$, $\mathbf{s} = (s_1, \ldots, s_T)'$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)'$, and ξ_μ. The graphical representation of DSP-V with parameter expansions is found in Figure 1. Specifically, $\forall t$

$$\begin{aligned} f(y_t^*, j_t|h_t) =& f(y_t^*|h_t, j_t)f(j_t) \\ f(\omega_t^*, s_t|v_t) =& f(\omega_t^*|v_t, s_t)f(s_t) \qquad\qquad \omega_t^* = log(\omega_t^2) = \log((h_t - h_{t-1})^2) \\ f(v_t, \xi_t|\mu, \phi) =& f(v_t|\mu, \phi, \xi_t)f(\xi_t) \\ f(\mu, \xi_\mu) =& f(\mu|\xi_\mu)f(\xi_\mu) \end{aligned}$$

Thus, the parameter expanded priors maybe written as:

$$f(\mathbf{j}, \mathbf{h}, \mathbf{v}, \mathbf{s}, \boldsymbol{\xi}, \mu, \xi_\mu, \phi) = f(\mathbf{j})f(\mathbf{h}|\mathbf{v}, \mathbf{s})f(\mathbf{s})f(\mathbf{v}|\mu, \phi, \boldsymbol{\xi})f(\boldsymbol{\xi})f(\mu|\xi_\mu)f(\xi_\mu)f(\phi).$$

In the following, we specify the exact distribution for each conditional prior distributions.

1. $\boldsymbol{f(j)}$ **:** Priors on $\mathbf{j}$ is given in Omori et al. [2007], where 10-component Gaussian mixture distribution was used to approximate $log(\chi_1^2)$.

2. $\boldsymbol{f(h|v)f(\omega^*|v, s)}$ **:** Without any parameter expansion, h_t depends on its previous h_{t-1} and its log-variance term v_t: $h_{t+1}|h_t \sim N(h_t, e^{v_t})$. Also, we assume $f(h_1|v_1) = \mathcal{N}(h_1|0, e^{v_1})$. Then, we have:

$$f(\mathbf{h}|\mathbf{v}) = \mathcal{N}(h_1|0, e^{v_1}) \prod_{t=2}^{T} \mathcal{N}(h_t|h_{t-1}, e^{v_t})$$

   Remember that $\boldsymbol{\omega^*} := log(\boldsymbol{\omega}^2) = log(\Delta\mathbf{h}^2)$. Thus, $f(\boldsymbol{\omega}|\mathbf{v}) = \mathcal{N}(0, Ie^{\mathbf{v}})$. Then,

$$\omega_t^*|v_t = v_t + \lambda_t^* \qquad\qquad \lambda_t^* \overset{i.i.d}{\sim} log(\chi_1^2)$$

   Using the same logic used for the parameter expansion for y^*, we get

$$f(\boldsymbol{\omega}^*|\mathbf{v}, \mathbf{s}) = \mathcal{N}(\omega^*|\mathbf{v} + \mu_s, I\sigma_s^2),$$

   and $\mathbf{s}$ follows the identical distribution as $\mathbf{j}$ detailed in Omori et al. [2007].

3. $\boldsymbol{f(s)}$ **:** As explored in the previous item, $f(\mathbf{s})$ has the identical distribution to $f(\mathbf{j})$.

4. $\boldsymbol{f(\mathbf{v}|\mu, \phi, \xi)}$ **:** v_t depends on its previous value v_{t-1}, unconditional mean parameter μ, the autoregressive parameter ϕ and the variance term ξ_{t-1}. As shown in Kowal et al. [2019],

$\eta_t \sim Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$ is a mean-variance mixture of Gaussian distribution with

$$f(\eta_t|\xi_{t-1}) = \mathcal{N}(\eta_t|0, \frac{1}{\xi_{t-1}})$$
$$f(\xi_t) = \mathcal{PG}(\xi_t|1, 0)$$

where PG represents Polya-Gamma random variable. Then, the priors on v may be expressed as:

$$f(\mathbf{v}|\mu, \phi, \boldsymbol{\xi}) = \mathcal{N}(v_1|\mu, \frac{1}{\xi_0})\prod_{t=2}^{T}\mathcal{N}(v_t|\mu + \phi(v_{t-1} - \mu), \frac{1}{\xi_{t-1}}).$$

5. $\boldsymbol{f(\xi)}$ **:** As shown in item v, for $t \in \{0, \dots, T-1\}$

$$f(\xi_t) = \mathcal{PG}(\xi_t|1, 0)$$

6. $\boldsymbol{f(\mu|\xi_\mu)}$ **:** For μ we have $m = \sqrt{e^\mu} \sim C^+(\frac{1}{\sqrt{A}})$, where A = 1.

$$\begin{aligned} f(m) &= \frac{2\sqrt{A}}{\pi(1 + Am^2)} && m \geq 0 \\ f(\mu) &= \frac{2\sqrt{A}}{\pi(1 + Ae^\mu)}\left|\frac{d}{d\mu}\sqrt{e^\mu}\right| = \frac{\sqrt{Ae^\mu}}{\pi(1 + Ae^x)} = \frac{\sqrt{e^{\mu-(-\log(A))}}}{\pi\left(1 + e^{\mu-(-\log(A))}\right)} && \mu \in (-\infty, \infty) \\ &= \mathcal{Z}(\mu|\frac{1}{2}, \frac{1}{2}, -log(A), 1) = \mathcal{Z}(\mu|\frac{1}{2}, \frac{1}{2}, 0, 1). \end{aligned}$$

Using the same logic used in item $f(v|\mu, \phi, \xi)$, we have

$$f(\mu|\xi_\mu) = \mathcal{N}(\mu|0, \frac{1}{\xi_\mu})$$
$$f(\xi_\mu) = \mathcal{PG}(\xi_\mu|1, 0)$$

7. $\boldsymbol{f(\xi_\mu)}$ **:** The parameter ξ_μ was introduced to expand the parameter μ, which follows a half-Cauchy distribution. As shown in item $f(\mu|\xi_\mu)$,

$$f(\xi_\mu) = \mathcal{PG}(\xi_\mu|1, 0).$$

8. $\boldsymbol{f(\phi)}$ **:** Let $b = \frac{\phi+1}{2}$.

$$f\left(\frac{\phi + 1}{2}\right) = f(b) = Beta(b|10, 2).$$

2.3 Gibbs Sampling

Building upon the Gibbs sampler proposed in Kowal et al. [2019], we propose Gibbs sampler for DSP-V. It uses log variance sampling by Kastner and Frühwirth-Schnatter [2014] with Polya-Gamma sampler by Polson et al. [2013], and a Cholesky factorization algorithm by Rue [2001] for efficient sampling of the state variable h_t. In addition, 10-component Gaussian mixture distribution by Omori et al. [2007] for approximating $log(\chi_1^2)$ distribution was used. Graphical representation DSP-V with parameter expansion is illustrated in Figure 1. The prior distribution is explored in section 2.2. Define $\theta^{1:T} = (\mathbf{j}, \mathbf{h}, \mathbf{v}, \mathbf{s}, \boldsymbol{\xi}, \mu, \xi_\mu, \phi)$ and $\theta^{1:T}_{-j} = (\mathbf{h}, \mathbf{v}, \mathbf{s}, \boldsymbol{\xi}, \mu, \xi_\mu, \phi)$. Similarly, define $\theta^{1:T}_{-h}$ and so forth in a similar manner. The following is the list of conditional distributions for Gibbs sampling. Despite large number of parameters, many of the variables are conditionally independent.

$$f(\mathbf{j}|\theta^{1:T}_{-j}, y^*) = f(\mathbf{j}|\mathbf{h}, y^*) \tag{A.1}$$

$$f(\mathbf{h}|\theta^{1:T}_{-h}, y^*) = f(\mathbf{h}|\mathbf{j}, \mathbf{v}, \mathbf{s}, y^*) \tag{A.2}$$

$$f(\mathbf{v}|\theta^{1:T}_{-v}, y^*) = f(\mathbf{v}|\mathbf{s}, \mathbf{h}, \mu, \phi, \boldsymbol{\xi}) \tag{A.3}$$

$$f(\mathbf{s}|\theta^{1:T}_{-s}, y^*) = f(\mathbf{s}|\mathbf{h}, \mathbf{v}) \tag{A.4}$$

$$f(\boldsymbol{\xi}|\theta^{1:T}_{-\xi}, y^*) = f(\boldsymbol{\xi}|\mathbf{v}, \mu, \phi) \tag{A.5}$$

$$f(\mu|\theta^{1:T}_{-\mu}, y^*) = f(\mu|\mathbf{v}, \boldsymbol{\xi}, \xi_\mu, \phi) \tag{A.6}$$

$$f(\xi_\mu|\theta^{1:T}_{-\xi_\mu}, y^*) = f(\xi_\mu|\mu) \tag{A.7}$$

$$f(\phi|\theta^{1:T}_{-\phi}, y^*) = f(\phi|v, \xi, \mu) \tag{A.8}$$

In order to obtain samples from the posterior distribution $f(\theta^{1:T}|y^*)$, we first initialize each parameters appropriately. Then we sequentially sample each parameter from 8 conditional distributions above. Once the distributions are converged, samples generated from the distributions below are equivalent to the samples from the desired posterior distribution. Other than ϕ, the closed form full conditional distribution is derived in Appendix A. We use slice sampling by Neal [2003] for sampling ϕ.

2.4 Forecast

Only a couple minor changes on the Gibbs sampler in Section 2.3 are required to perform forecasting. Let's consider m-step ahead forecast. The subscript notation is used to denote time index. Our goal is to sample from the joint posterior predictive distribution:

$$\begin{aligned} f(y^*_{T+1:T+m}, \theta^{T+1:T+m}|y^*_{1:T}) &= \int f(y^*_{T+1:T+m}, \theta^{T+1:T+m}|\theta^{1:T}, y^*_{1:T}) f(\theta^{1:T}|y^*_{1:T}) d\theta^{1:T} \\ &= \int f(y^*_{T+1:T+m}, \theta^{\theta^{1:T+m}T+m}|y^*_{1:T}) d\theta^{1:T} \end{aligned}$$

which is equivalent to sampling from the joint distribution.

$$f(y^*_{T+1:T+m}, \theta^{1:T+m}|y^*_{1:T}).$$

Let's reminder ourselves that the distribution for y^*_t is completely determined by h_t and j_t $\forall t$ as shown in (Figure 1). Using Bayes Theorem, we get:

$$\begin{aligned} f(y^*_{T+1:T+m}, \theta^{1:T+m}|y^*_{1:T}) &= f(y^*_{T+1:T+m}|\theta^{1:T+m}, y^*_{1:T}) f(\theta^{1:T+m}|y^*_{1:T}) \\ &= f(y^*_{T+1:T+m}|j_{T+1:T+m}, h_{T+1:T+m}) f(\theta^{1:T+m}|y^*_{1:T}) \end{aligned}$$

Thus, we have two independent distributions to sample from: 1) $f(y^*_{T+1:T+m}|j_{T+1:T+m}, h_{T+1:T+m})$ and $f(\theta^{1:T+m}|y^*_{1:T})$. As explored in section 2.1, $\forall t$, $\forall k \in \{1, \dots, 10\}$, $f(y^*_t|j_t = k, h_t) = \mathcal{N}(y^*_t|h_t + \mu_k, \sigma^2_k)$, which can be easily sampled once j_t and h_t are sampled. Now let's direct our attention for Gibbs sampling, $f(\theta^{1:T+m}|y^*_{1:T})$. Let's remind ourselves that other than the variable **h** and **j**, variables don't depend on $y^*_{1:T}$. Therefore, the full conditional distribution for variables $h_{1:T+m}, v_{1:T+m}, s_{1:T+m}$ and $\xi_{1:T+m}$ are identical to the ones discussed in Section 2.3, which are further detailed in Appendix A. The full conditional distribution for $h_{1:T+m}$ and $j_{1:T+m}$ are also explored in Appendix B.

3 Simulation Study

3.1 Set-up

The goal of the simulation study was to compare proposed Dynamic Shrinkage Process in Volatility (DSP-V) against widely used volatility models including 1) Stochastic Volatility model (SV), 2) Markov-Switching Stochastic Volatility model (MSSV), 3) Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and 4) Markov-Switching GARCH (MSGARCH). 100 sample paths were generated under these four data generating schemes with each path consisting of 400 data points. Given $y_t = \sigma_t\epsilon_t$, $\epsilon_t \overset{i.i.d}{\sim} N(0,1)$,$\eta_t \overset{i.i.d}{\sim} N(0,1)$, $h_t := log(\sigma^2_t)$, we have the following data generating schemes.

1. **Scheme 1 (Stochastic Volatility Model):**

$$h_{t+1} = -2 + 0.9(h_t + 2) + 0.2\eta_t$$

2. **Scheme 2 (Markov Switching Stochastic Volatility Model):**

$$h_{t+1} = \mu_{s_{t+1}} + 0.5(h_{t-1} - \mu_{s_t}) + 2\eta_t$$
$$\mu_{s_t} = -10 + 13s_t, s_t \in \{0, 1\}$$
$$P(s_{t+1} = j|s_t = i) = p_{ij}$$

3. **Scheme 3 (GARCH model):**

$$\sigma^2_{t+1} = 1 + 0.1y^2_t + 0.5\sigma^2_t$$

4. **Scheme 4 (Markov Switching GARCH model):**

$$\sigma^2_{t+1} = \omega_{s_{t+1}} + 0.2y^2_t + 0.2\sigma^2_t$$
$$\omega_{s_t} = 1 + 99s_t, s_t \in \{0, 1\}$$
$$P(s_{t+1} = j|s_t = i) = p_{ij}$$

Transitions between states 0 and 1 for simulation schemes 2 and 4 were governed by the following transition matrix $P = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}$. Performances of each model were evaluated in terms of their in-sample mean absolute deviation in $h_t := log(\sigma^2_t)$ for Schemes 1 and 2 and σ_t for Schemes 3 and 4. Indeed, SV, MSSV, GARCH, and MSGARCH would perform the best respectively for simulation schemes 1 through 4, as they represent the true models in those scenarios. Our aim is to compare the performance of DSP-V against the true data generating process and other alternative models in a variety of data generating schemes. In order to have concise comparisons, DSP-V were compared against SV and MSSV model under scheme 1 and 2, and against GARCH and MSGARCH under scheme 3 and 4. All four models were implemented in R language (R Core Team [2013]). Specifically, Bayesian Stochastic Volatility model (SV model) was implemented via stochvol (Kastner [2016]) package, Markov Switching Stochastic Volatility model as described in Hamilton [1989] were implemented by the authors as no readily available packages in R exists,

	Scheme 1	Scheme 2		Scheme 3	Scheme 4
SV	0.2972	1.2894	GARCH	0.0693	3.5276
MSSV	0.4337	0.9706	MSGARCH	0.0755	2.6064
DSP-V	0.3078	1.0427	DSP-V	0.1191	3.0278

Table 1: Average Mean Absolute Errors (MAE) over 100 sample paths with either the volatility σ or log-squared volatility h as a target are compared between Dynamic Shrinkage Process in Volatility (DSP-V) and four other models, including Stochastic Volatility (SV), Markov Switching Stochastic Volatility (MSSV), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH(1,1) and Markov-Switching GARCH (MSGARCH).

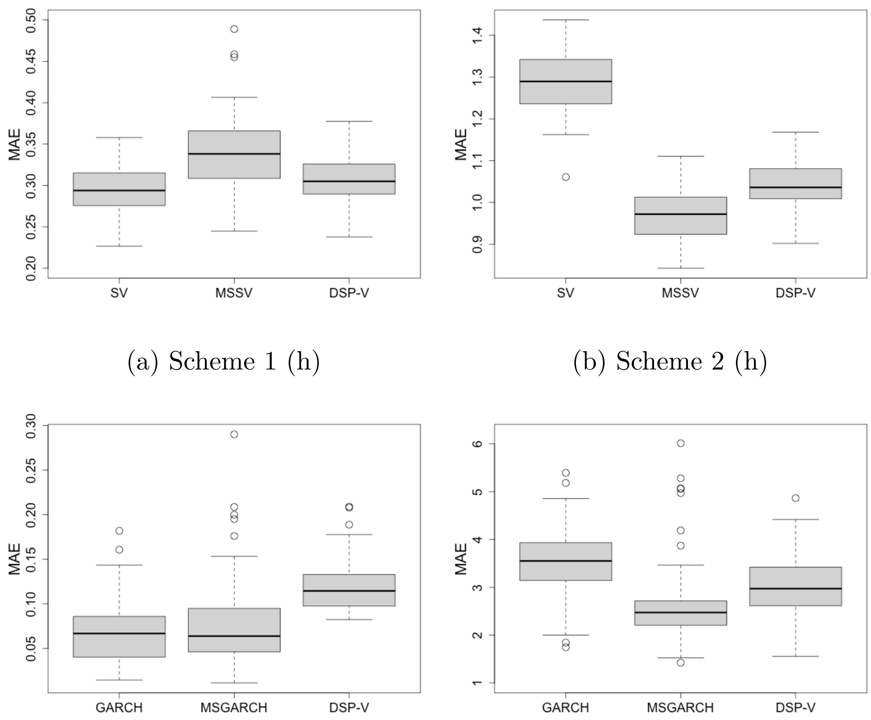
GARCH(1,1) models were fitted via fGarch (Wuertz et al. [2023]) package, and Markov Switching GARCH (MSGARCH) models were fitted via MSGARCH (Ardia et al. [2019]) package. For each model, 5000 samples were generated with 5000 burn-in samples.

3.2 Result

Figure 2 illustrates the Mean Absolute Error (MAE) distribution across four simulation schemes, and Table 1 provides the average MAE for the models under examination. Sample paths for simulation schemes 1 and 2 were generated from SV models, featuring 1 and 2 regimes, respectively. In contrast, sample paths for simulation schemes 3 and 4 were generated using GARCH models, each with 1 and 2 regimes. As anticipated, the true model, SV, MSSV, GARCH, and MSGARCH, had the lowest MAE in simulation scheme 1 through 4 respectively.

Our proposed method, DSP-V closely followed the true model with average MAE of 0.3078 for Scheme 1 and 1.0427 for Scheme 2. It is important to note that the performance of DSP-V is comparable to the true model, which had the average MAE of 0.2972 for Scheme 1 and 0.9706 for scheme 2. The alternative models, MSSV for Scheme 1 and SV for Scheme 2, had significantly higher average MAE of 0.4337 and 1.2894 respectively. The overall distributions of MAE are depicted in Figure 2a and 2b. The findings indicate that DSP-V is a versatile framework, adept at approximating SV models with varying numbers of regime changes, without the need for prior knowledge regarding the number of regimes. This stands in contrast to Markov-switching models, which demand that users specify the number of structural breaks in the sample before modeling. Given the common uncertainty surrounding the number of structural breaks in the sample in empirical analyses, DSP-V holds a clear advantage over MSSV in this regard.

DSP-V, nevertheless, performed less favorably than GARCH and MSGARCH (Figure 2c) in



(a) Scheme 1 (h)

(b) Scheme 2 (h)

(c) Scheme 3 (σ)

(d) Scheme 4 (σ)

Figure 2: Box plots of Mean Absolute Error (MAE) across 100 sample paths measured in h for scheme 1 and 2, and σ for scheme 3 and 4. Stochastic Volatility (SV), Markov-Switching Stochastic Volatility (MSSV), and Dynamic Shrinkage Process in Volatility (DSP-V) were compared in scheme 1 and 2. Generalized AutoRegressive Conditional Heteroskedasticity (GARCH(1,1)), Markov-Switching GARCH (MSGARCH), and DSP-V were compared in scheme 3 and 4.

Scheme 3, in which the data were generated from GARCH(1,1) model. As explored in Section 2, DSP-V closely resembles the SV model, as it directly models the log-volatility term h_t instead of σ_t. Thus, this discrepancy may introduce model-misspecification bias resulting in high MAE for DSP-V when compared against GARCH and MSGARCH. Despite the model misspecification bias, DSP-V was still able to perform comparably in simulation scheme 4 (Figure 2d). The average MAE for the true model MSSV, is 2.6064, followed by DSP-V and GARCH with the average MAE of 3.0278 and 3.5276, respectively. Despite model misspecification bias, DSP-V overall outperforms GARCH(1,1) when large regime changes occur in the sample path.

The simulation result underscores the versatility of DSP-V as a robust framework for modeling stochastic volatility, especially in the presence of extreme regime switches. The significance of DSP-V lies in its ability to produce reliable results across diverse data generating processes, a crucial aspect considering that, in empirical analysis, the true data generating process is often unknown. Consequently, DSP-V stands out by providing results that, while not optimal, are still comparable, all without requiring knowledge of the actual data generating process. Regime-switching models such as MSSV and MSGARCH model both require that we know the number of regimes prior to the analysis. And when the number of regimes are misspecified, bias is introduced. On the other hand, DSP-V, with its locally adaptive DSP, can learn regime switching from the data. Through a series of simulation studies, we demonstrate that DSP-V is a flexible framework that outperforms widely used models such as SV, MSSV, GARCH, and MSGARCH across diverse data generating schemes.

4 Empirical Study

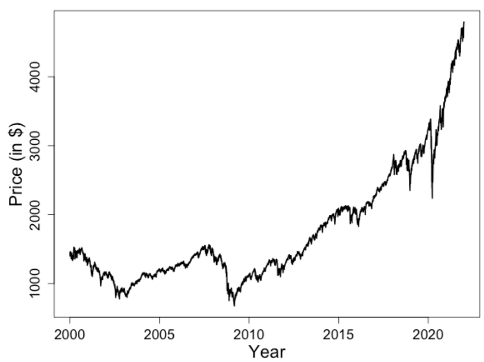
4.1 Set-up

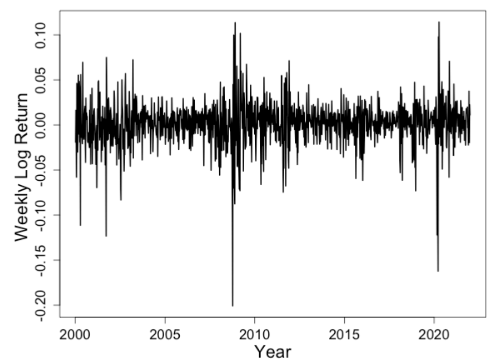
In this section, we applied DSP-V to analyze empirical data. Specifically, we looked at the price of S&P500 from 2012-01-01 to 2021-12-31 (n = 1148, Figure 3a), EURO to US Dollar exchange rate between 2000-01-03 to 2012-04-04 (n = 639, Figure 3c), and new deaths from COVID-19 from 2020-01-03 to 2023-11-13 (n = 202, Figure 3e). DSP-V, were then, applied to the weekly log return series for S&P 500 and EURO/USD exchange (Figure 3b and Figure 3d), and the weekly first difference series for global COVID-19 death tolls (Figure 3f). We first compared the estimated log-volatility (h) derived from DSP-V against the one generated from the SV model. Subsequently, a more detailed examination of DSP parameters such as κ_t (the shrinkage parameter for h_t), μ, and ϕ is performed.

4.2 Results

4.2.1 Log-Volatility (h)

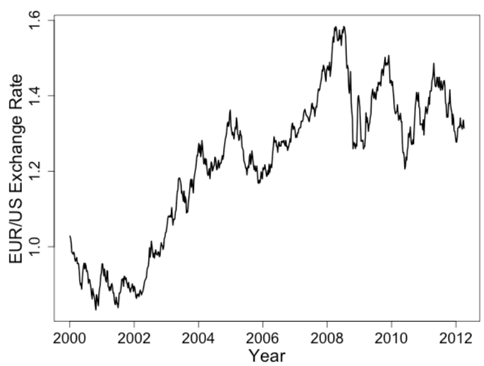
In this analysis, we assess the log-volatility (h) estimates generated by DSP-V and SV models across three data sets. Figure 4 illustrates the estimated h based on DSP-V and SV. Focusing on the S&P 500 dataset, the most prominent feature of the series in both SV and SVSDP are the large spikes in 2009 and 2020 due to financial crisis and COVID-19 outbreaks. In terms of

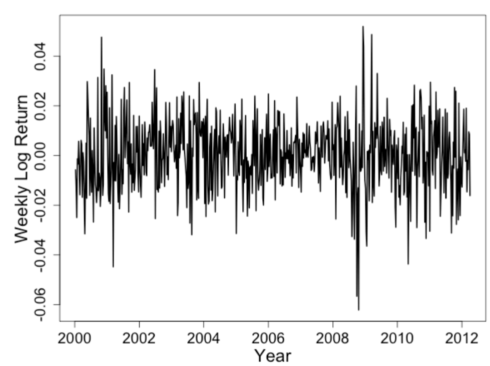




(a) S&P500 Index

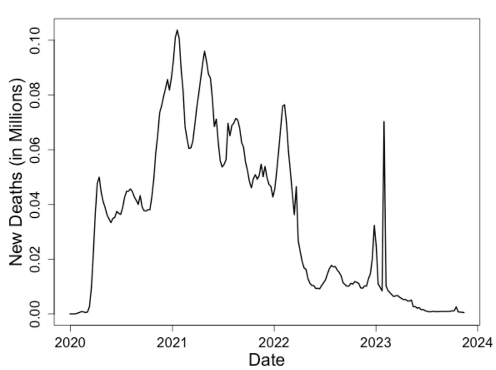
(b) Weekly Log Return on S&P500 Index

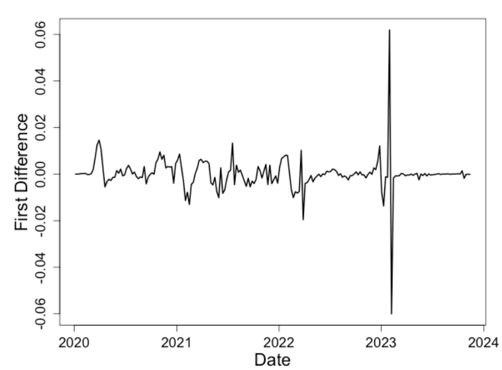




(c) EURO/USD Exchange Rate

(d) Weekly Log Return on Exchange Rate



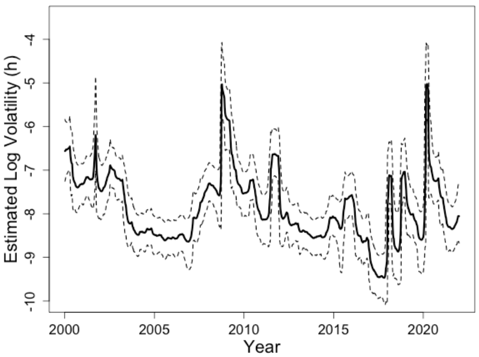


(e) Global COVID-19 Deaths

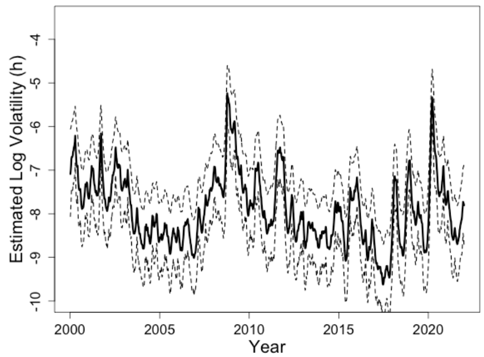
(f) First Difference COVID Deaths

Figure 3: Price and log weekly return of S&P500 between 2012-01-01 and 2021-12-31, EURO/USD exchange rate and its log weekly return between 2000-01-03 and 2012-04-04, and weekly global COVID-19 death tolls and its weekly changes between 2020-01-03 to 2023-11-13 are illustrated from (a) to (f), respectively.

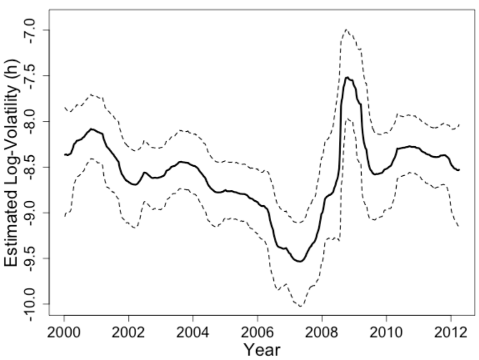
moderate changes, the one around 2001 coincides with 9/11 attack and the 2011 surge aligns with the European Debt crisis, both of which caused a stock market decline. The stock market in



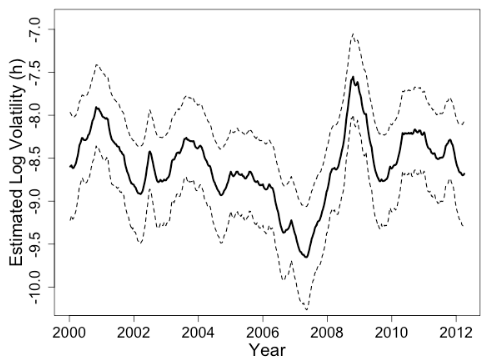
(a) S&P 500 Index (DSP-V)



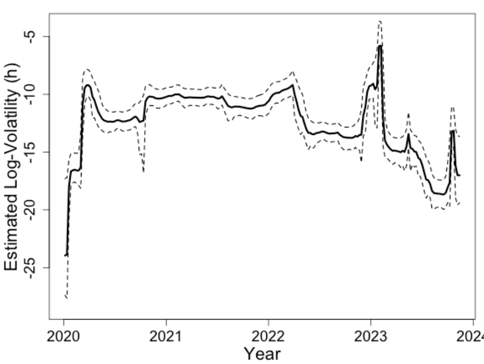
(b) S&P 500 Index (SV)



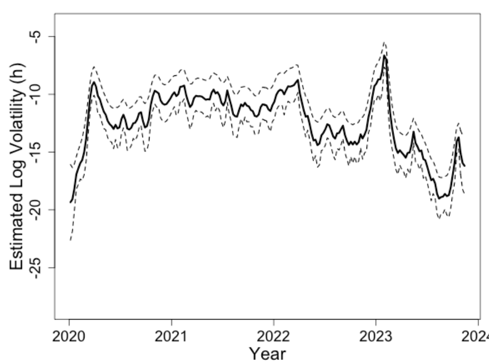
(c) EURO/USD Exchange Rate (DSP-V)



(d) EURO/USD Exchange Rate (SV)



(e) COVID-19 Deaths (DSP-V)



(f) COVID-19 Deaths (SV)

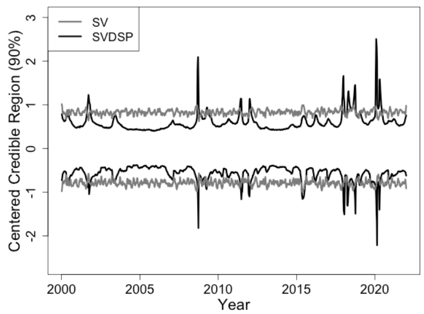
Figure 4: Estimated h with 90% quantile-based credible regions on weekly returns on S&P 500 between 2000-01-01 and 2021-12-31, weekly returns on EURO/USD exchange rate between 2000-01-03 and 2012-04-04, and weekly changes in Global COVID-19 death tolls between 2020-01-03 and 2023-11-13 based on Dynamic Shrinkage Process in Volatility (DSP-V) shown in (a),(c), and (e) and based on Stochastic Volatility (SV) model shown in (b),(e), and (f).

2018 followed the strongest 2017 market since the financial crisis but was overall exhibited extreme volatility. The S&P 500 index closed at approximately 2930 in September but plummeting 19.73% to 2351 by Christmas Eve in 2018. While discernible in DSP-V, small to moderate changes in volatility is hard to distinguish from noise in SV's noisier estimates.

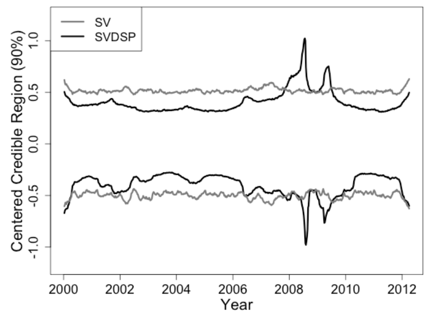
The EURO/USD exchange rate's most notable volatility pattern includes a steady decline from 2001 to 2008, a substantial spike in 2008 and 2009 due to the financial crisis, and a minor bump in 2010 and 2011, likely tied to the European Debt crisis as well. Figure 3c suggests a steady increase in the EURO/USD exchange rate between 2001 and 2008, resulting in low volatility. Both DSP-V and SV outputs exhibit a very similar pattern, with DSP-V displaying a smoother output. Because of this, The steady decline in volatility from 2001 to 2007 is easier to notice from DSP-V than from SV.

A discernible pattern from both SV and DSP-V for COVID-19 death tolls is the significant volatility spike occurring around the onset of the pandemic in early 2020 and another one in early 2023. Subsequently, seasonal effect appears to be present; there are small to large increases in volatility at the end and beginning of each year, followed by reductions in the summer months, particularly noticeable in 2022 and 2023. In mid-2022, a substantial dip in volatility is observed, succeeded by a notable spike toward the end of 2022 and the beginning of 2023. By the mid-2023, volatility plateaus, and experience dramatic decline, only to exhibit another spike by the end of 2023. This pattern is also observed to a lesser extent in 2020 and 2021, according to DSP-V estimates. Due to noisier output from SV model, on the other hand, such pattern becomes less noticeable. The seasonal cycle aligns with the winter wave of COVID-19, suggesting a higher likelihood of contracting the virus in winter months in the Northern Hemisphere. Despite the data being global death tolls, the observed seasonal pattern corresponds to that of the Northern Hemisphere, where approximately 90% of the global population resides. The seasonal effect of the Northern Hemisphere's population outweighs that of the Southern Hemisphere.

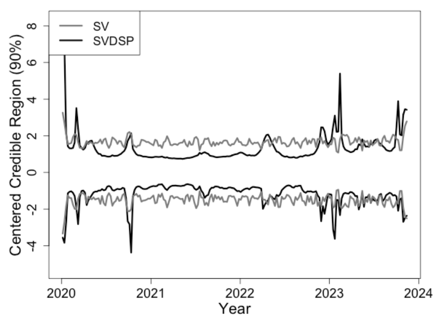
Thus, both DSP-V (Figure 4a, 4c, and 4e) and SV (Figure 4b, 4d, and 4f) produce similar overall pattern of h. The primary distinction between the two models, however, lies in the degree of smoothness; the proposed model, DSP-V, produces a notably smooth output when compared against the output from SV model. A key advantage of this enhanced smoothness is improved interpretability. As discussed earlier, DSP-V excels, particularly in discerning small to medium changes in volatility, offering better clarity in distinguishing signal from noise compared to SV, whose outputs tend to be noisier, rendering signals often indistinguishable from the noise. The reduced noise in DSP-V's output enables users to gain a clearer understanding of underlying patterns and trends.



(a) S&P 500 Index



(b) EURO/USD Exchange Rate



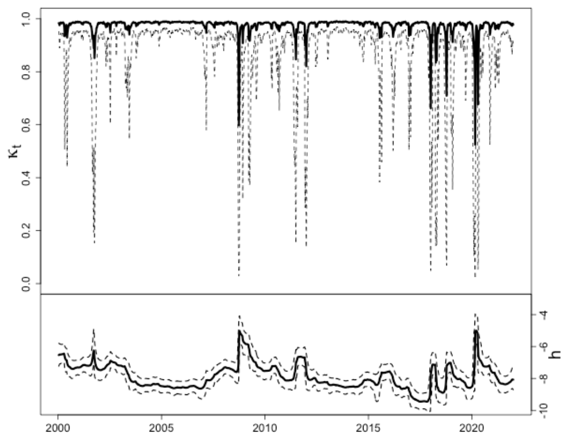
(c) COVID-19 Deaths

Figure 5: From the 90% quantile-based credible regions the expected h_t were subtracted to allow comparisons of the credible regions in terms of their width. The three datasets include weekly returns on S&P 500 between 2000-01-01 and 2021-12-31, weekly returns on EURO/USD exchange rate between 2000-01-03 and 2012-04-04, and weekly changes in Global COVID-19 death tolls between 2020-01-03 and 2023-11-13. The centered credible regions for Dynamic Shrinkage Process in Volatility (DSP-V) are drawn in black and the ones based on Stochastic Volatility (SV) model are in grey.

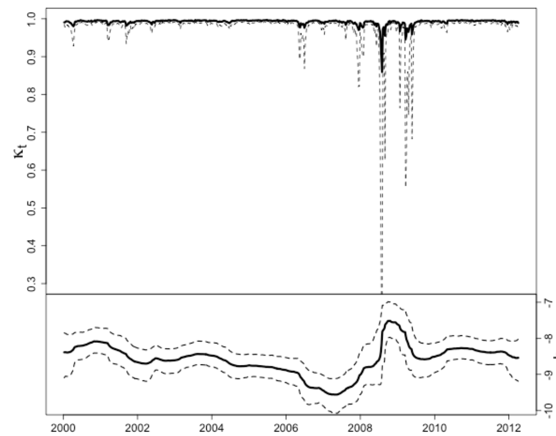
Another distinguishing feature of DSP-V is its credible region. Due to time-varying variance term v_t, credible region around h_t produced by DSP-V also changes over time. The mean-centered credible regions generated by SV and DSP-V on the three data sets are presented in Figure 5. Notably, the credible regions maintain a constant width for the SV model across the sample path. The credible region around h for DSP-V, on the other hand, exhibits local adaptability. Specifically, the patterns on the centered credible regions match that of h_t explored in previous paragraphs.

When changes in h_t are slow moving, the width of the credible region is narrow; while abrupt changes in h_t induce large credible regions.

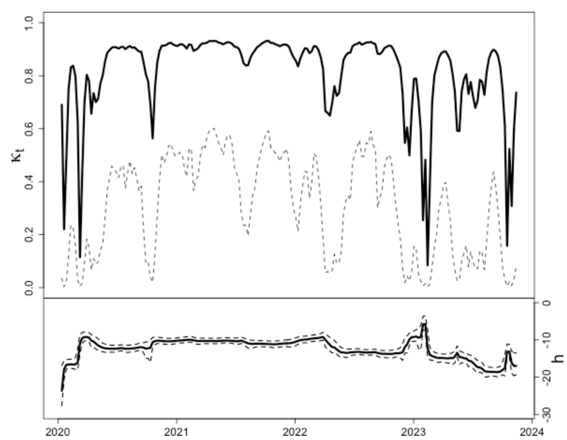
4.2.2 Shrinkage Parameter κ_t



(a) S&P 500 Index



(b) EURO/USD Exchange Rate



(c) COVID-19 Deaths

Figure 6: Comparison between the expected shrinkage parameter $\kappa_t := \frac{1}{1+var(\omega_t|\tau,\lambda_t)} = \frac{1}{1+exp(v_t)} = \frac{1}{1+\tau^2\lambda_t^2}$ and expected h_t based on Dynamic Shrinkage Process in Volatility (DSP-V) estimated on weekly returns on S&P 500 between 2000-01-01 and 2021-12-31, weekly returns on EURO/USD exchange rate between 2000-01-03 and 2012-04-04, and weekly changes in Global COVID-19 death tolls between 2020-01-03 and 2023-11-13, respectively. The dotted lines represent the one-sided 95th and centered 90th percentile credible regions for κ_t and h_t respectively.e

As specified in section 2, the parameter v_t is the log-variance term for the first difference of h_t. Thus, a small v_t corresponds to minor changes in h_t, while a large v_t signifies substantial variations in h_t. Alternatively, one can conceptualize v_t as influencing the degree of shrinkage applied to h_t,

	S&P 500		EURO/USD		COVID Deaths	
	Estimate	90% CR	Estimate	90% CR	Estimate	90% CR
μ	-7.508	(-8.413,-6.783)	-10.03	(-11.85,-8.357)	-3.186	(-4.615,-1.916)
ϕ	0.352	(-0.093,0.5907	0.450	(0.125, 0.680)	0.580	(0.246,0.769)

Table 2: Expected value and 90% quantile based credible region for parameters μ and ϕ parameters from Dynamic Shrinkage Process in Volatility (DSP-V) estimated for weekly returns on S&P 500 between 2000-01-01 and 2021-12-31, weekly returns on EURO/USD exchange rate between 2000-01-03 and 2012-04-04, and weekly changes in Global COVID19 death tolls between 2020-01-03 and 2023-11-13.

denoted as $\kappa_t := \frac{1}{1+\mathrm{var}(\omega_t|\tau,\lambda_t)} = \frac{1}{1+\exp(v_t)} = \frac{1}{1+\tau^2\lambda_t^2}$. Analyzing κ_t rather than v_t is preferred for clarity, as $\kappa_t \in (0,1)$, unlike $v_t \in (-\infty,+\infty)$. A high κ_t (near 1) signals strong model certainty, while a low value (near 0) indicates less certainty in the estimate. The comparison between the estimated log-volatility h_t and the shrinkage parameter κ_t are depicted in Figure 6. In line with our understanding of DSP-V, prominent peaks in h_t align with values near 0 in κ_t. During periods of gradual evolution, on the other hand, κ_t is close to 1. In general, the κ_t for COVID-19 deaths data is comparatively more volatile than that observed for the S&P 500 and exchange rate data, indicating a lower level of certainty in the model's estimates. This discrepancy could be attributed to the limited sample size of the COVID-19 dataset, which comprises only of 203 data points. In contrast, the S&P 500 and exchange rate datasets contain 1148 and 639 data points, respectively.

4.2.3 Dynamic Shrinkage Process Parameters

The dynamics of v_t, the time-varying variance term for h_t, are governed by an AR(1) process with a slightly different error term, $\eta_t \sim Z(a,b,0,1)$, to induce adequate shrinkage. For all three data sets, $Z(\frac{1}{2},\frac{1}{2},0,1)$ was used. The term μ represents the global parameter for v_t; it represents the overall level of changes occurring in h_t with the large value signifying large changes h_t in sample and vice versa. The expected value and the 90% credible regions for μ and ϕ across the three data sets are summarized in Table 2. The data set with largest μ were COVID-19 data set with the expected value of -3.186, followed by S&P 500 and the exchange rate data. As illustrated in Figure 4e, h_t ranges between -25 to -5, whereas the range of h_t for the other two data sets are much smaller.

The parameter ϕ represents the autoregressive strength, indicating the dependence between successive v_t values. In all three cases, $\phi > 0$. This observation suggests that large changes in h_t

are likely followed by subsequent large changes. Notably, COVID-19 data exhibited the highest expected ϕ at 0.58, followed by 0.434 for the exchange rate and 0.352 for S&P 500 data. The higher ϕ parameter in COVID-19 data is in accordance with the inherent dynamics of infectious diseases, where notable surges or declines often signal underlying regime changes. For example, a substantial surge in cases can catalyze additional infections leading to an exponential growth pattern in COVID-19 cases. Conversely, a significant decrease in COVID-19 deaths might be associated with seasonal effects or vaccination efforts, influencing the overall trajectory of the disease in the opposite direction. In contrast, economic downturns such as financial crisis in 2009 or COVID-19 outbreak induce structural changes in which large volatility is generally observed. However, not all substantial changes in financial market volatility may be solely ascribed to such regime changes. Some fluctuations may stem from the intrinsic unpredictability inherent in financial markets.

5 Conclusion

In this paper, we introduced a novel Dynamic Shrinkage Process in Volatility (DSP-V) as a flexible and adaptive framework for modeling stochastic volatility in financial data. The key innovation lies in the incorporation of a global-local prior called Dynamic Shrinkage Process, allowing DSP-V to dynamically adapt to evolving processes. Through an extensive simulation study, we demonstrated that DSP-V consistently outperforms or yields comparable results to established models such as Stochastic Volatility (SV), Markov Switching Stochastic Volatility, GARCH(1,1), and Markov Switching GARCH, across diverse data generating scenarios. Notably, DSP-V exhibited resilience to model mispecification biases, showcasing its robustness in financial modeling. When compared against estimated volatility from other existing models, DSP-V distinguishes itself from others by producing smooth, yet locally adaptive output. To this end, DSP-V emerges as a powerful tool for capturing the complexities of stochastic volatility in financial data. Future research could explore additional applications of DSP-V in various financial contexts and extend its capabilities to handle high-frequency data and other complex financial phenomena.

References

David Ardia, Keven Bluteau, Kris Boudt, Leopoldo Catania, and Denis-Alexandre Trottier. Markov-switching garch models in r: The msgarch package. *Journal of Statistical Software*, 91(4):1–38, 2019. doi: 10.18637/jss.v091.i04.

Luc Bauwens, Arie Preminger, and Jeroen V. K. Rombouts. Theory and inference for a markov switching garch model. *The Econometrics Journal*, 13(2):218–244, 2010. ISSN 13684221, 1368423X. URL http://www.jstor.org/stable/23117467.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1831029.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(86)90063-1. URL https://www.sciencedirect.com/science/article/pii/0304407686900631.

Jun Cai. A markov model of switching-regime arch. *Journal of Business & Economic Statistics*, 12(3):309–316, 1994. ISSN 07350015. URL http://www.jstor.org/stable/1392087.

CARLOS M. Carvalho, NICHOLAS G. POLSON, and JAMES G. SCOTT. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. ISSN 00063444, 14643510. URL http://www.jstor.org/stable/25734098.

Ray Yeutien Chou. Volatility persistence and stock valuations: Some empirical evidence using garch. *Journal of Applied Econometrics*, 3(4):279–294, 1988. ISSN 08837252, 10991255. URL http://www.jstor.org/stable/2096644.

Francis X. Dieobold. Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):51–56, 1986. doi: 10.1080/07474938608800096. URL https://doi.org/10.1080/07474938608800096.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912773.

Kenneth French, G. Schwert, and Robert Stambaugh. Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29, 1987. URL https://EconPapers.repec.org/RePEc:eee:jfinec:v:19:y:1987:i:1:p:3-29.

Stephen F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62, September 1996. URL https://ideas.repec.org/a/eee/jfinec/v42y1996i1p27-62.html.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912559.

James D Hamilton and Raul Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1):307–333, 1994. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(94)90067-1. URL https://www.sciencedirect.com/science/article/pii/0304407694900671.

JOHN Hull and ALAN White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987. doi: https://doi.org/10.1111/j.1540-6261.1987.tb02568.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1987.tb02568.x.

Soosung Hwang, Steve E. Satchell, and Pedro L. Valls Pereira. How Persistent is Volatility? An Answer with Stochastic Volatility Models with Markov Regime Switching State Equations. Econometric Society 2004 Latin American Meetings 198, Econometric Society, August 2004. URL https://ideas.repec.org/p/ecm/latm04/198.html.

Gregor Kastner. Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(5):1–30, 2016. doi: 10.18637/jss.v069.i05.

Gregor Kastner and Sylvia Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423, aug 2014. doi: 10.1016/j.csda.2013.01.002. URL https://doi.org/10.1016%2Fj.csda.2013.01.002.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393, 1998. ISSN 00346527, 1467937X. URL http://www.jstor.org/stable/2566931.

Daniel R. Kowal, David S. Matteson, and David Ruppert. Dynamic Shrinkage Processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(4):781–804, 05 2019. ISSN 1369-7412. doi: 10.1111/rssb.12325. URL https://doi.org/10.1111/rssb.12325.

Christopher G Lamoureux and William Lastrapes. Heteroskedasticity in stock return data: Volume versus garch effects. *Journal of Finance*, 45(1):221–29, 1990. URL https://EconPapers.repec.org/RePEc:bla:jfinan:v:45:y:1990:i:1:p:221-29.

Angelo Melino and Stuart M. Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2):239–265, 1990. URL https://EconPapers.repec.org/RePEc:eee:econom:v:45:y:1990:i:1-2:p:239-265.

Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/1926560.

Philip Messow and Walter Krämer. Spurious persistence in stochastic volatility. *Economics Letters*, 121(2):221–223, 2013. ISSN 0165-1765. doi: https://doi.org/10.1016/j.econlet.2013.08.008. URL https://www.sciencedirect.com/science/article/pii/S0165176513003704.

Thomas Mikosch and Cătălin Stărică. Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *The Review of Economics and Statistics*, 86(1):378–390, 2004. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/3211680.

Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705 – 767, 2003. doi: 10.1214/aos/1056562461. URL https://doi.org/10.1214/aos/1056562461.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2006.07.008. URL https://www.sciencedirect.com/science/article/pii/S0304407606001436.

Nicholas G. Polson, James G. Scott, and Jesse Windle. Bayesian inference for logistic models using polya-gamma latent variables, 2013.

Ser-Huang Poon and Stephen J. Taylor. Stock returns and volatility: An empirical study of the uk stock market. *Journal of Banking & Finance*, 16(1):37–59, 1992. ISSN 0378-4266. doi: https://doi.org/10.1016/0378-4266(92)90077-D. URL https://www.sciencedirect.com/science/article/pii/037842669290077D. Special Issue on European Capital Markets.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

Havard Rue. Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2):325–338, 2001. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/2680602.

Mike K. P. So, K. Lam, and W. K. Li. A stochastic volatility model with markov switching. *Journal of Business & Economic Statistics*, 16(2):244–253, 1998. ISSN 07350015. URL http://www.jstor.org/stable/1392580.

Mike K.P. So, K. Lam, and W.K. Li. An Empirical Study of Volatility in Seven Southeast Asian Stock Markets Using ARV Models. *Journal of Business Finance & Accounting*, 24(2):261–276, March 1997. doi: 10.1111/1468-5957.00104. URL https://ideas.repec.org/a/bla/jbfnac/v24y1997i2p261-276.html.

Fei Su and Lei Wang. Conditional volatility persistence and realized volatility asymmetry: Evidence from the chinese stock markets. *Emerging Markets Finance and Trade*, 56(14):3252–3269, 2020. doi: 10.1080/1540496X.2019.1574566. URL https://doi.org/10.1080/1540496X.2019.1574566.

S.J. Taylor. *Modelling Financial Time Series*. G - Reference,Information and Interdisciplinary Subjects Series. World Scientific, 2008. ISBN 9789812770844. URL https://books.google.com/books?id=KQ5pDQAAQBAJ.

Diethelm Wuertz, Yohan Chalabi, Tobias Setz, Martin Maechler, and Georgi N. Boshnakov. *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*, 2023. URL https://www.rmetrics.org. R package version 4031.90.

A Full Conditional Distributions for Gibbs Sampling

In the following subsections, we will derive the conditional posterior for each parameter: $(\mathbf{j}, \mathbf{h}, \mathbf{v}, \mathbf{s}, \boldsymbol{\xi}, \mu, \xi_\mu, \phi)$.

A.1 j

$\mathbf{j} = (j_1, \ldots, j_T)$ was introduced to expand the likelihood on y^*. In this section, we show that $\forall k \in \{1, \ldots, 10\}$:

$$\begin{aligned} p(j_t = k|h_t, v_t, s_t, \xi_t, \mu_t, \xi_\mu, \phi, y_t^*) &= p(j_t = k|h_t, y_t^*) \\ &= \frac{\mathcal{N}(y_t^*|h_t + \mu_k, \sigma_k^2)p_k}{\sum_{i=1}^{10} p_i \mathcal{N}(y_t^*|h_t + \mu_i, \sigma_i^2)}, \qquad \forall t \in \{1, \ldots, T\} \end{aligned}$$

$\mathbf{j}$ is only associated with y^*, which is only associated with $\mathbf{h}$ and $\mathbf{j}$:

$$p(j_t = k|h_t, v_t, s_t, \xi_t, \mu_t, \xi_\mu, \phi, y_t^*) = p(j_t = k|h_t, y_t^*) = \frac{f(y_t^*|h_t, j_t)p(j_t = k)}{\int f(y_t^*|h, j_t) f(j_t) dj_t}$$

As described in Section 2.1, parameter $j = (j_1, \ldots, j_T)'$ is introduced, where $\forall t$

$$f(y_t^*|j_t, h_t) = \mathcal{N}(y_t^*|h_t + \mu_{j_t}, \sigma_{j_t}^2)$$

Naturally, $\forall k \in \{1, \ldots, 10\}$

$$\begin{aligned} p(j_t = k|h_t, y_t^*) &= \frac{f(y_t^*|h_t, j_t = k)p(j_t = k)}{f(y_t^*|h_t)} \\ &= \frac{f(y_t^*|h_t, j_t = k)p(j_t = k)}{\sum_{i=1}^{10} p(j_t = i) f(y_t^*|h_t, j_t = i)} \\ &= \frac{\mathcal{N}(y_t^*|h_t + \mu_k, \sigma_k^2)p_k}{\sum_{i=1}^{10} p_i \mathcal{N}(y_t^*|h_t + \mu_i, \sigma_i^2)}, \qquad \forall t \in \{1, \ldots, T\}, \end{aligned}$$

which is what we wanted to show. Exact distribution on $j_t \overset{i.i.d}{\sim} Categorical(p_1, \ldots, p_{10})$ as well as corresponding mean and the variance parameter of each component is described in Section 2.2.

A.2 h

In this section, we derive the conditional distribution the state variable $\mathbf{h}$:

$$f(\mathbf{h}|j, v, y^*) = \mathcal{N}\left(\mathbf{h} \middle| \left(Q_v + I\frac{1}{\sigma_j^2}\right)^{-1} \frac{y - \mu_j}{\sigma_j^2}, \left(Q_v + I\frac{1}{\sigma_j^2}\right)^{-1}\right),$$

where Q_v is a symmetric tridiagonal matrix defined below. Based on section 2.2,

$$f(y^*|\mathbf{j},\mathbf{h}) = \prod_{t=1}^{T} \mathcal{N}(y_t^*|h_t+\mu_{jt},\sigma_{jt}^2) = \mathcal{N}(y^*|h+\mu_j, I\sigma_j^2)$$
$$f(\mathbf{h}|\mathbf{v}) = \mathcal{N}(h_1|0,e^{v_1})\prod_{t=1}^{T}\mathcal{N}(h_t|h_{t-1},e^{v_t})$$

We derive conditional joint prior on $h|v$ like we did for $f(y^*|h,j)$. Based on section 2:

$$\boldsymbol{\Delta}\mathbf{h}|\mathbf{v} = \begin{pmatrix} h_1|v_1 \\ h_2-h_1|v_2 \\ \vdots \\ h_T-h_{T-1}|v_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{bmatrix} e^{v_1} & & & \\ & e^{v_2} & & \\ & & \ddots & \\ & & & e^{v_t} \end{bmatrix}\right)$$

Indeed,

$$\mathbf{h}|\text{v} \sim \begin{bmatrix} 1 & & & & 0 \\ 1 & 1 & & & \\ 1 & 1 & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ 1 & 1 & \dots & \dots & 1 \end{bmatrix} \Delta h|v \sim N(0, Q_v^{-1})$$

where

$$Q_v^{-1} = \begin{bmatrix} 1 & & & & 0 \\ 1 & 1 & & & \\ 1 & 1 & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ 1 & 1 & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} e^{v_1} & \dots & \dots & 0 \\ \vdots & e^{v_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & e^{v_t} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ & 1 & 1 & \dots & 1 \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ 0 & & & & 1 \end{bmatrix}$$

And Q_v is a band matrix

$$Q_v = \begin{bmatrix} \left(\frac{1}{e^{v_2}} - \frac{1}{e^{v_1}}\right) & -\frac{1}{e^{v_2}} & 0 & \dots & \dots & 0 \\ -\frac{1}{e^{v_2}} & \left(\frac{1}{e^{v_3}} - \frac{1}{e^{v_2}}\right) & -\frac{1}{e^{v_3}} & \ddots & \ddots & \vdots \\ 0 & -\frac{1}{e^{v_3}} & \left(\frac{1}{e^{v_4}} - \frac{1}{e^{v_3}}\right) & -\frac{1}{e^{v_4}} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -\frac{1}{e^{v_{T-1}}} & \left(\frac{1}{e^{v_{T-1}}} - \frac{1}{e^{v_T}}\right) & -\frac{1}{e^{v_T}} \\ 0 & \dots & \dots & 0 & -\frac{1}{e^{v_T}} & \frac{1}{e^{v_T}} \end{bmatrix}$$

We have $f(\mathbf{h}|\mathbf{v}) = \mathcal{N}(\mathbf{h}|0, Q_v^{-1})$. Thus, we may conclude that.

$$\begin{aligned} f(\mathbf{h}|\mathbf{j}, \mathbf{v}, y^*) &\propto \mathcal{N}(y^*|\mathbf{h} + \mu_j, I\sigma_j^2)\mathcal{N}(\mathbf{h}|0, Q_v^{-1}) \\ &\propto \mathcal{N}(y^* - \mu_j|\ \mathbf{h}, I\sigma_j^2)\mathcal{N}(\mathbf{h}|0, Q_v^{-1}). \end{aligned}$$

Posterior on normal likelihood with a known variance and the normal prior on the mean, also known as normal-normal model, has a closed form posterior distribution, which is also Gaussian.

$$f(\mathbf{h}|j, v, y^*) = \mathcal{N}\left(\mathbf{h}\middle|\left(Q_v + I\frac{1}{\sigma_j^2}\right)^{-1}\frac{y-\mu_j}{\sigma_j^2}, \left(Q_v + I\frac{1}{\sigma_j^2}\right)^{-1}\right),$$

which is what we wanted to show.

A.3 v

In this section, we show that

$$\begin{aligned} f(\mathbf{v}|\theta_{-v}^{1:T}, y^*) &= f(\mathbf{v}|\mathbf{s}, \mathbf{h}, \mu, \phi, \boldsymbol{\xi}) = f(\mathbf{v}|\mathbf{s}, \boldsymbol{\omega}^*, \mu, \phi, \boldsymbol{\xi}) \\ &= \mathcal{N}\left(\mathbf{v}\middle|\left(Q_{\xi,\phi} + I\frac{1}{\sigma_s^2}\right)^{-1}\left(\frac{\omega^* - \mu_s}{\sigma_s^2} + Q_{\xi,\phi}\mathbf{1}\mu\right), \left(Q_{\xi,\phi} + I\frac{1}{\sigma_s^2}\right)^{-1}\right) \end{aligned}$$

Let $\omega_1^* = log(h_1^2)$ and $\omega_t^* = log((h_t - h_{t-1})^2), \forall t \geq 2$. Since $h_t - h_{t-1} \sim N(0, e^{v_t})$, $\omega_t^*|v_t = v_t + \lambda_t$, where $\lambda_t \overset{i.i.d}{\sim} log(\chi_1^2)$. Then, without parameter expansion we have:

$$f(v|h, \mu, \phi) = f(v|\omega^*, \mu, \phi) \propto f(\omega^*|v)f(v|\mu, \phi).$$

Both $\boldsymbol{\omega}^*|\mathbf{v}$ and $\mathbf{v}|\mu, \phi$ will be further expanded with variables $\mathbf{s} = (s_1, \ldots, s_T)'$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)'$, respectively for efficient sampling. In the following sections (A.3.1 and A.3.2) we show that:

$$\begin{aligned} f(\boldsymbol{\omega}^*|\mathbf{v}, \mathbf{s}) &= \mathcal{N}(\boldsymbol{\omega}^* - \mu_s|v, I\sigma_s^2) \\ f(\mathbf{v}|\boldsymbol{\xi}, \mu, \phi) &= \mathcal{N}(\mathbf{v}|\mathbf{1}\mu, Q_{\xi,\phi}^{-1}), \end{aligned}$$

which is indeed, the normal-normal model, with the following posterior distribution:

$$f(\mathbf{v}|\mathbf{s}, \boldsymbol{\omega}^*, \mu, \phi, \boldsymbol{\xi}) = \mathcal{N}\left(\mathbf{v}\middle|\left(Q_{\xi,\phi} + I\frac{1}{\sigma_s^2}\right)^{-1}\left(\frac{\omega^* - \mu_s}{\sigma_s^2} + Q_{\xi,\phi}\mathbf{1}\mu\right), \left(Q_{\xi,\phi} + I\frac{1}{\sigma_s^2}\right)^{-1}\right),$$

where $Q_{\xi,\phi}$ is a tridiagonal matrix derived below.

A.3.1 Parameter expansion: $f(\omega^*|v, s)$

Similar to the derivation on the likelihood y^*, we may approximate $log(\chi_1^2)$ with 10-component Gaussian Mixture proposed by Omori et al. [2007]. Parameter $\mathbf{s} = (s_1, \ldots, s_T)'$ is introduced to expand $\omega_t^* = \log(\lambda_t^2)$. (Please refer to section 2.2 for more detail.) Then, we get the following parameter expanded conditional prior $\boldsymbol{\omega}^*|\mathbf{v}, \mathbf{s}$.

$$
\begin{aligned}
f(\omega_t^*|v_t) &= \sum_{i=1}^{10} p_i \mathcal{N}(\omega_t^*|\mu_i + v_t, \sigma_i^2) \\
f(\omega_t^*|v_t, s_t = k) &= \mathcal{N}(\omega_t^*|\mu_k + v_t, \sigma_k^2) = \mathcal{N}(\omega_t^* - \mu_k|v_t, \sigma_k^2) \\
f(\boldsymbol{\omega}^*|\mathbf{v}, \mathbf{s}) &= \mathcal{N}(\omega^* - \mu_s|v, I\sigma_s^2)
\end{aligned}
$$

A.3.2 Parameter expansion: $f(v|\xi, \mu, \phi)$

Let's now consider the conditional prior on $\mathbf{v}|\mu, \phi$. Note that we had $v_1 = \mu + \eta_0$ and $\forall t \geq 2$, $v_t = \mu + \phi(v_{t-1} - \mu) + \eta_{t-1}$, where $\eta_t \sim Z(\frac{1}{2}, \frac{1}{2}, 0, 1)$, with

$$
f(\mathbf{v}|\mu, \phi) = \mathcal{Z}(v_1|\frac{1}{2}, \frac{1}{2}, \mu, 1) \prod_{t=2}^{T} \mathcal{Z}(v_t|\frac{1}{2}, \frac{1}{2}, \mu + \phi(v_{t-1} - \mu), 1).
$$

$\boldsymbol{\xi} = (\xi_0, \ldots, \xi_{T-1})'$ is introduced to expand $\mathbf{v}|\mu, \phi$, particularly the error term η_t such that

$$
\begin{aligned}
f(v|\xi, \mu, \phi) &= \mathcal{N}(v_1|\mu, \frac{1}{\xi_0}) \prod_{t=2}^{T} \mathcal{N}(v_t|\mu + \phi(v_{t-1} - \mu), \frac{1}{\xi_{t-1}}) \\
\xi_t &\overset{i.i.d}{\sim} PG(1, 0),
\end{aligned}
$$

where PG represents the Polya-Gamma random variable. As an intermediary step, we introduce $\mathbf{v}^* = (v_1^*, \ldots, v_T^*)'$, where $v_1^* = v_1$ and $v_{t+1}^* = v_{t+1} - \phi v_t, t \geq 2$. This makes $\mathbf{v}^*$ conditionally independent Gaussian random variables:

$$
\begin{aligned}
f(v_1^*|\xi_1, \mu) &= f(v_1|\xi_0, \mu) = \mathcal{N}(v_t^*|\mu, \frac{1}{\xi_0}) \\
f(v_t^*|\xi_{t-1}, \mu, \phi) &= f(v_t - \phi v_{t-1}|\xi_{t-1}, \mu, \phi) = \mathcal{N}(v_t^*|\mu(1-\phi), \frac{1}{\xi_{t-1}}) \qquad t \geq 2,
\end{aligned}
$$

which can be expressed as multivariate Gaussian:

$$\mathbf{v}^*|\boldsymbol{\xi}, \mu, \phi \sim N\Bigg(\begin{pmatrix} \mu \\ \mu(1-\phi) \\ \vdots \\ \mu(1-\phi) \end{pmatrix}, \begin{bmatrix} \frac{1}{\xi_0} & & & \\ & \frac{1}{\xi_1} & & \\ & & \ddots & \\ & & & \frac{1}{\xi_{T-1}} \end{bmatrix} \Bigg).$$

Similar to the derivation for $h|v$ in section A.2, we have:

$$v|\xi, \mu, \phi = \begin{bmatrix} 1 & & & & 0 \\ \phi & 1 & & & \\ \phi^2 & \phi & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ \phi^{T-1} & \phi^{T-2} & \dots & \dots & 1 \end{bmatrix} v^*|\xi, \mu, \phi$$

And $Q_{\xi,\phi}$ is also a band matrix

$$Q_{\xi,\phi} = \begin{bmatrix} \xi_1 + \phi^2\xi_2 & -\phi\xi_2 & 0 & \dots & \dots & 0 \\ -\phi\xi_2 & \xi_2 + \phi^2\xi_3 & -\phi\xi_3 & \ddots & \ddots & \vdots \\ 0 & -\phi\xi_3 & \xi_3 + \phi^2\xi_4 & -\phi\xi_4 & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -\phi\xi_{T-1} & \xi_{T-1} + \phi^2\xi_T & -\phi\xi_T \\ 0 & \dots & \dots & 0 & -\phi\xi_T & \xi_T \end{bmatrix}$$

Thus, we get

$$f(\mathbf{v}|\boldsymbol{\xi}, \mu, \phi) = \mathcal{N}(v|\mathbf{1}\mu, Q_{\xi,\phi}^{-1}).$$

A.4 s

In section A.2, $\mathbf{s} = (s_1, \ldots, s_T)'$ was introduced to expand $\boldsymbol{\omega}^*|\mathbf{v}$. Based on the same argument used in section A.1, $\forall k \in \{1, \ldots, 10\}$

$$\begin{aligned} p(s_t = k|\omega_t^*, v_t) &= \frac{f(\omega_t^*|v_t, s_t = k)p(s_t = k)}{\sum_{i=1}^{10} f(\omega_t^*|v_t, s_t = i)p(s_t = i)} \\ &= \frac{\mathcal{N}(\omega_t^*|v_t + \mu_k, \sigma_k^2)p_k}{\sum_{i=1}^{10} p_i \mathcal{N}(\omega_t^*|v_t + \mu_i, \sigma_i^2)}, \qquad \forall t \in \{1, \ldots, T\}, \end{aligned}$$

$\forall k \in \{1, \ldots, 10\}, (p_k, \mu_k, \sigma_k^2)$ can be found in Omori et al. [2007].

A.5 ξ

In section A.3, $\boldsymbol{\xi} = (\xi_0, \ldots, \xi_{T-1})'$ were introduced to expand $\mathbf{v}|\mu, \phi$. As shown in Kowal et al. [2019], given $\boldsymbol{\eta}|\boldsymbol{\xi} \sim N(0, \frac{1}{\boldsymbol{\eta}})$ and $\boldsymbol{\xi} \sim PG(1, 0)$, we have $\boldsymbol{\xi}|\boldsymbol{\eta} \sim PG(1, \boldsymbol{\eta})$. As explored in Section 2.2:

$$\begin{aligned} v_1 &= \mu + \eta_0 \\ v_t &= \mu + \phi(v_{t-1} - \mu) + \eta_{t-1}, \qquad \forall t \geq 2. \end{aligned}$$

Naturally, $\eta_1 = v_1 - \mu$ and $\eta_t = v_{t+1} - \phi v_t - \mu(1 - \phi), \forall t \geq 2$.

$$\begin{aligned} f(\xi_0|\mathbf{v}, \mu, \phi) &= \mathcal{PG}(\xi_0|1, v_1 - \mu) \\ f(\xi_{t-1}|\mathbf{v}, \mu, \phi) &= \mathcal{PG}(\xi_{t-1}|1, v_t - \phi v_{t-1} - \mu(1 - \phi)) \qquad \forall t \geq 2, \end{aligned}$$

where $\mathcal{PG}$ represents the density function for Polya-Gamma random variable.

A.6 μ

In this section, we show that:

$$\begin{aligned} f(\mu|\mathbf{v}, \boldsymbol{\xi}, \xi_\mu, \phi) = f(\mu|\hat{v}_\phi^*, \boldsymbol{\xi}, \xi_\mu, \phi) &= f(\hat{v}_\phi^*|\mu, \phi, \boldsymbol{\xi})f(\mu|\xi_\mu) \\ &= \mathcal{N}\left(\mu \middle| \left(\frac{1}{\sigma_{\xi,\phi}^2} + \xi_\mu\right)^{-1} \frac{\hat{v}_\phi^*}{\sigma_{\xi,\phi}^2}, \left(\frac{1}{\sigma_{\xi,\phi}^2} + \xi_\mu\right)^{-1}\right), \end{aligned}$$

where $\hat{v}_\phi^*$ is defined below.

In section A.3.2, we defined v^* and showed that

$$\mathbf{v}^*|\boldsymbol{\xi},\mu,\phi \sim N\left(\begin{pmatrix}\mu\\ \mu(1-\phi)\\ \vdots\\ \mu(1-\phi)\end{pmatrix}, \begin{bmatrix}\frac{1}{\xi_1} & & & \\ & \frac{1}{\xi_2} & & \\ & & \ddots & \\ & & & \frac{1}{\xi_T}\end{bmatrix}\right).$$

Let's further transform the parameter $\mathbf{v}^*$ and define v^*_ϕ

$$v^*_\phi|\boldsymbol{\xi},\mu,\phi = \begin{pmatrix}v^*_1\\ \frac{v^*_2}{(1-\phi)}\\ \vdots\\ \frac{v^*_T}{(1-\phi)}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu\\ \mu\\ \vdots\\ \mu\end{pmatrix}, \begin{bmatrix}\frac{1}{\xi_0} & & & \\ & \frac{1}{(1-\phi)^2\xi_1} & & \\ & & \ddots & \\ & & & \frac{1}{(1-\phi)^2\xi_{T-1}}\end{bmatrix}\right),$$

Then, we can see that

$$\hat{v}^*_\phi|\boldsymbol{\xi},\mu,\phi := \frac{1}{T}\sum_{t=1}^{T} v^*_{\phi,t} \sim N(\mu,\sigma^2_{\xi,\phi})$$

$$\sigma^2_{\xi,\phi} := \frac{1}{T^2}\left(\frac{1}{\xi_0} + \frac{1}{(1-\phi)^2}\sum_{t=1}^{T-1}\frac{1}{\xi_t}\right)$$

Let's then consider $f(\mu|\xi_\mu)$. Based on section 2.2, we showed that $\mu \sim Z(\frac{1}{2},\frac{1}{2},0,1)$. Similar to the parameter $\boldsymbol{\xi}$ for $\mathbf{h}$ in section A.3.2, we introduce a parameter ξ_μ, so that

$$f(\mu|\xi_\mu) = \mathcal{N}(\mu|0,\frac{1}{\xi_\mu})$$

$$f(\xi_\mu) = \mathcal{PG}(\xi_\mu|1,0).$$

Clearly, we have the normal-normal model. We conclude

$$
\begin{aligned}
f(\hat{v}_\phi^*|\boldsymbol{\xi}, \mu, \phi) &= \mathcal{N}(\hat{v}_\phi^*|\mu, \sigma_{\xi,\phi}^2) \\
f(\mu|\xi_\mu) &= \mathcal{N}(\mu|0, \frac{1}{\xi_\mu}) \\
f(\mu|\hat{v}_\phi^*, \boldsymbol{\xi}, \xi_\mu, \phi) &= \mathcal{N}\left(\mu\middle|\left(\frac{1}{\sigma_{\xi,\phi}^2} + \xi_\mu\right)^{-1} \frac{\hat{v}_\phi^*}{\sigma_{\xi,\phi}^2}, \left(\frac{1}{\sigma_{\xi,\phi}^2} + \xi_\mu\right)^{-1}\right).
\end{aligned}
$$

A.7 ξ_μ

Note that ξ_μ only depend on μ, where prior on $\mu|\xi_\mu \sim N(0, \frac{1}{\xi_\mu})$. Using the same derivation for $\xi|v, \mu, \phi$ in section A.5, we have:

$$
f(\xi_\mu|\mu) = \mathcal{PG}(\xi_\mu|1, \mu)
$$

A.8 ϕ

For the parameter ϕ, we have

$$
f(\phi|v, \boldsymbol{\xi}, \mu) = f(\phi|\hat{v}_\mu, \boldsymbol{\xi}, \mu) \propto f(\hat{v}_\mu|\boldsymbol{\xi}, \mu, \phi) f(\phi)
$$

Let's remind ourselves that $t \geq 2$, $v_t = \mu + \phi(v_{t-1} - \mu) + \eta_{t-1}$, and with parameter expansion on η_t, we have

$$
f(v_t|v_{t-1}, \mu, \xi_t) = \mathcal{N}(v_t|\mu + \phi(v_{t-1} - \mu), \frac{1}{\xi_{t-1}})
$$

$t \geq 2$, let's define:

$$
\begin{aligned}
v_{\mu,t} &= \frac{1}{2}\left(\frac{v_t - \mu}{v_{t-1} - \mu} + 1\right) \\
\hat{v}_\mu &= \frac{1}{T-1} \sum_{t=2}^{T} v_{\mu,t}
\end{aligned}
$$

Then, we have:

$$
f(v_{\mu,t}|\mu, \xi_{t-1}) = \mathcal{N}\left(\frac{1}{2}\left(\frac{v_t - \mu}{v_{t-1} - \mu} + 1\right)\middle|\frac{\phi + 1}{2}, \frac{1}{4\xi_{t-1}(v_{t-1} - \mu)^2}\right) \qquad t \geq 2
$$

$$
f(\hat{v}_\mu|\mu, \boldsymbol{\xi}) = \mathcal{N}\left(\hat{v}_\mu\middle|\frac{\phi + 1}{2}, \frac{1}{(T-1)^2} \sum_{t=2}^{T} \frac{1}{4\xi_{t-1}(v_{t-1} - \mu)^2}\right)
$$

And we have:

$$\frac{\phi+1}{2} \sim Beta(10,2)$$

We may think of this as a normal likelihood with the known variance and an unknown mean parameter that follows a Beta distribution. Slice sampling by Neal [2003] was used to sample from the conditional distribution.

B Conditional Distribution for Forecast

Continuing our discussion in section 2.4, we further explore Gibbs sampling for the following joint distribution, $f(\theta^{1:T+m}|y^*_{1:T})$:

$$\begin{aligned}
f(j_{1:T+m}|\theta^{1:T+m}_{-j_{1:T+m}} y^*_{1:T}) &= f(j_{1:T+m}|h_{1:T+m}, y^*_{1:T}) \\
f(h_{1:T+m}|\theta^{1:T+m}_{-h_{1:T+m}}, y^*_{1:T}) &= f(h_{1:T+m}|j_{1:T+m}, v_{1:T+m}, s_{1:T+m}, y^*_{1:T}) \\
f(v_{1:T+m}|\theta^{1:T+m}_{-v_{1:T+m}}, y^*_{1:T}) &= f(v_{1:T+m}|h_{1:T+m}, s_{1:T+m}, \mu, \xi_\mu, \phi) \\
f(s_{1:T+m}|\theta^{1:T+m}_{-s_{1:T+m}}, y^*_{1:T}) &= f(s_{1:T+m}|h_{1:T+m}, v_{1:T+m}) \\
f(\xi_{1:T+m}|\theta^{1:T+m}_{-\xi_{1:T+m}}, y^*_{1:T}) &= f(\xi_{1:T+m}|v_{1:T+m}, \mu, \phi) \\
f(\mu|\theta^{1:T+m}_{-\mu}, y^*_{1:T}) &= f(\mu|v_{1:T+m}, \xi_{1:T+m}, \xi_\mu, \phi) \\
f(\xi_\mu|\theta^{1:T+m}_{-\xi_\mu}, y^*_{1:T}) &= f(\xi_\mu|\mu) \\
f(\phi|\theta^{1:T+m}_{-\phi}, y^*_{1:T}) &= f(\phi|v_{1:T+m}, \xi_{1:T+m}, \mu)
\end{aligned}$$

Note that other than the distributions for $j_{1:T+m}$ and $h_{1:T+m}$, the conditional distributions are identical to the ones derived in section 2.3, and further explored in detail in Appendix A.

B.1 h

By conditional independence,

$$\begin{aligned}
f(h_{1:T+m}|\theta^{1:T+m}_{-h_{1:T+m}}, y^*) &= f(h_{1:T+m}|j_{1:T+m}, v_{1:T+m}, y^*_{1:T}) \\
&= f(h_{T+1:T+m}|v_{T+1:T+m}) f(h_{1:T}|j_{1:T}, v_{1:T}, y^*_{1:T}) \\
&= \prod_{i=1}^{m} \mathcal{N}(h_{T+i}|h_{T+i-1}, e^{v_{T+i}}) f(h_{1:T}|j_{1:T}, v_{1:T}, y^*_{1:T})
\end{aligned}$$

$f(h_{1:T}|j_{1:T}, v_{1:T}, y^*_{1:T})$ is the joint conditional explored in A.2, and is indeed sampled regardless of whether prediction is made or not. Based on the sample from h_T, we may sequentially or jointly sample from $h_{1:T+m}$. The joint distribution for $\mathbf{h}$ conditional on v were explored in section A.2, with the precision matrix Q_v. The same recipe maybe used.

B.2 j

Another variable that depends on y^* were $\mathbf{j}$. The goal is to sample from

$$
\begin{aligned}
f(j_{1:T+m}|h_{1:T+m}, v_{1:T+m}, s_{1:T+m}, \xi_{1:T+m}, \mu, \xi_\mu, \phi, y^*)
&= f(j_{T+1:T+m}|j_{1:T}, j_{T+1}, h_{1:T+m}, v_{1:T+m}, s_{1:T+m}, \xi_{1:T+m}, \mu, \xi_\mu, \phi, y^*_{1:T}) \\
&\quad f(j_{1:T}|h_{1:T+m}, v_{1:T+m}, s_{1:T+m}, \xi_{1:T+m}, \mu, \xi_\mu, \phi, y^*_{1:T}) \\
&= \prod_{i=1}^{m} f(j_{T+i}) f(j_{1:T}|s_{1:T}, y_{1:T}).
\end{aligned}
$$

Thus, $j_{T+1}, \ldots, j_{T+m}$ and $j_1, \ldots, j_T$ are sampled from independent distributions. For $j_{T+1}, \ldots, j_{T+m}$ are sampled from the Omori et al. [2007] 10-component mixture. The conditional distribution, $f(j_{1:T}|s_{1:T}, y_{1:T})$ is explored in section A.1 and is sampled whether we make forecast or not.