ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A dynamic state-space HAR model☆

Mike Tsionas [a], Aya Ghalayini [b], Marwan Izzeldin [a], Lorenzo Trapani [c,*]

[a] Lancaster University Management School, Lancaster, UK
[b] College of Business and Social Sciences, Aston University, Birmingham, UK
[c] Universita' degli Studi di Pavia, Pavia, Italy, and University of Leicester, Leicester, UK

ARTICLE INFO

Keywords:
Time-varying coefficients
HAR models
Particle Gibbs sampling
State Space models
Volatility forecasting

ABSTRACT

The Heterogeneous AutoRegressive model for the logs of Realised Volatility (HARL) has established itself as the benchmark specification for modelling and forecasting return volatility, owing to its parsimony and ability to capture the strong persistence typically observed in RV. To address potential concerns such as measurement errors, nonlinearities, and non-spherical residuals, numerous variants of the baseline HARL model have been developed in the literature. This paper contributes to this body of work by proposing a new class of dynamic state-space models with time-varying parameters. The parameter dynamics are assumed to follow an autoregressive process, with or without stochastic volatility, giving rise to two specifications: SHARP and SHARP-SV. Both models are designed to capture the evolving nature of return volatility and are estimated via Bayesian inference using Particle Gibbs sampling. Empirical applications to high-frequency data on SPY, sector ETFs, representative NYSE stocks, and the VIX index demonstrate that our proposed models on average outperform alternative HARL-based specifications in forecasting volatility, particularly at medium- and long-term horizons. An extensive Monte Carlo analysis further illustrates the advantages of our approach in terms of both estimation accuracy and predictive performance.

1. Introduction

The presence of time-varying parameters and time-varying volatility in economic and financial data (see, *inter alia*, Buccheri and Corsi, 2021, and their literature review) is a very well-documented stylised fact. Models that fail to account for changes in the data generating process are liable to produce biased estimates and inaccurate predictions. Thus, it is essential to analyse and account for parameter behaviour in the statistical modelling of such data. The importance of taking such features into account has been emphasised in several studies, including those by Koop and Potter (2004), D'Agostino et al. (2013), Clark and Ravazzolo (2015), Bekierman and Manner (2018), Chen et al. (2018), and Buccheri and Corsi (2021).

As far as time variation is concerned, the use of Time Varying Parameters (TVP) models has been long proposed as a more flexible (and, possibly, more realistic) alternative to models with abrupt breaks. Several applications have shown the superior forecasting ability of TVP models, also in the context of high-dimensional Vector AutoRegressive models; whilst a comprehensive literature review goes well beyond the scope of this paper, we refer to the contributions by Doan et al. (1984), Sims (1993), Stock and Watson (1996), and Cogley and Sargent (2005), as seminal papers, and to the more recent works by Carriero et al. (2019) and Tsionas et al. (2022), also for an up-to-date list of references. Alongside this strand of the literature, modelling time variation in the volatility is

☆ This article is part of a Special issue entitled: 'Bayesian econometric approaches' published in Journal of Econometrics.

* Corresponding author.
E-mail addresses: m.tsionas@lancaster.ac.uk (M. Tsionas), a.ghalayini@aston.ac.uk (A. Ghalayini), m.izzeldin@lancaster.ac.uk (M. Izzeldin), lorenzo.trapani@unipv.it (L. Trapani).

https://doi.org/10.1016/j.jeconom.2025.106146
Received 6 August 2024; Received in revised form 30 September 2025; Accepted 7 November 2025

also a very important topic, e.g. in finance (owing to its significance in risk management, portfolio selection, and asset pricing), and also in macroeconomic forecasting (see, for example, the papers by Clark, 2011; Koop and Korobilis, 2013; Carriero et al., 2015; Clark and Ravazzolo, 2015; and Koop et al., 2019 - where accounting for heteroskedasticity is shown to yield dramatic improvements in forecasting). In particular, in the context of financial econometrics, nonparametric approaches based on the so-called Realised Volatility (RV) are often preferred to parametric approaches, based e.g. on GARCH-type models (Bollerslev, 1986).

Since the seminal contributions by Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), and Liu et al. (2015), modelling realised volatility has advanced markedly, with the Heterogeneous Autoregressive (HAR) model by Corsi (2009) becoming the empirical workhorse. By employing daily, weekly, and monthly RV averages, HAR parsimoniously captures persistence and heterogeneity across horizons. Corsi (2009) notes that modelling the logs of RV, as well as ensuring non-negativity, also improves forecast accuracy. We take this HAR-log (HARL) specification as the cornerstone of our contribution. Despite its success, the (linear) HARL can suffer from measurement error, nonlinearities, and autocorrelated, heteroskedastic residuals, motivating extensions such as regime-switching (McAleer and Medeiros, 2008), dynamic model averaging (Wang et al., 2016), and TVP variants (Bekierman and Manner, 2018; Chen et al., 2018; and Buccheri and Corsi, 2021).

We address the aforementioned limitations of the HARL framework by building on (and extending) the approaches mentioned above, accounting for nonlinearities, time variations in the parameters (and possible heteroskedasticity in the law of motion of the coefficients), and autocorrelation and heteroskedasticity in the residuals. Specifically, we present a novel dynamic state-space model with time-varying coefficients, assumed to follow an AutoRegressive (AR) process with or without Stochastic Volatility (SV) - henceforth, we refer to the former specification (that is, a model with TVP but no SV in the law of motion of the coefficients) as the SHARP model; and to the latter (i.e., a model with TVP and SV in the law of motion of the coefficients) as its SHARP-SV variant. An integral part of our methodology is the use, in both cases, of Bayesian inference with Particle Gibbs sampling, following the procedure by Creal and Tsay (2015), which allows for efficient computation of the latent variables (see also Andrieu et al., 2010). However, even though our main focus – as far as inference is concerned – is Bayesian, we also study in depth the dependence structure of our model. Indeed, the SHARP and SHARP-SV models produce observations – owing to the presence of time-varying parameters and stochastic volatility – whose dependence structure is highly complex. Still, we show that our proposed models generate observations which belong in a wide class of *weakly* dependent processes, which do not exhibit persistence or pseudo-long-memory behaviour.[1]

Empirically, we embed these new features into the HARL framework to improve RV forecasts. Our results show that this approach outperforms standard methods in predicting the volatility of both index and individual stock returns. In an extensive empirical analysis, we examine daily RV from the SPY-ETF – a tradable U.S. market index – alongside 10 sector ETFs (2006–2023), 20 representative NYSE stocks (2000–2016), and the VIX index (2003–2023). Across all datasets, both the SHARP and SHARP-SV models significantly outperform the HARL model and its time-varying extensions, as confirmed by the predictive ability test of Giacomini and White (2006). While SHARP-SV provides a moderate forecasting edge over SHARP, both consistently belong in the Model Confidence Set of Hansen et al. (2011). Results are broadly confirmed via a Monte Carlo simulation, where we study the performance of our estimation methodology, and the forecasting accuracy of our proposed models (and of Bayesian estimation *cum* Particle Gibbs) under possible misspecification, using several Data Generating Processes (DGPs henceforth) exhibiting the typical features of financial data.

The remainder of the paper is organised as follows. Section 2 provides an overview of the Realised Variance measure and existing extensions to the HARL model with dynamic and time-varying parameters. In Section 3, we introduce our proposed models and provide details on the estimation method. In Section 4, we apply and evaluate the proposed models in forecasting stock volatility using real data. A comprehensive Monte Carlo study is reported in Section 5, where we assess the performance of our Bayesian estimator under correct specification (Section 5.1), and predictive accuracy under possible mis-specification (Section 5.2). Section 6 concludes. Further results are in the Supplement where we: report the detailed forecasting results and some misspecification analysis (Section A), and further Monte Carlo evidence (Section B); present the posterior derivations (Section C); provide details on the Particle Filtering algorithm (Section D); study the dependence structure of the SHARP and SHARP-SV models (Section E), and present technical lemmas (Section F) and proofs (Section G).

2. Volatility measure and HARL family of models

Consider an asset whose log-price $\ln(P_s)$ follows the stochastic differential equation:

$$d\ln(P_s) = \mu_s ds + \sigma_s dW_s, \tag{2.1}$$

where μ_s denotes the drift, σ_s is the instantaneous volatility and W_t a standard Wiener process. The (latent) integrated variance for day t is defined as:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds, \tag{2.2}$$

and its nonparametric, ex-post estimate based on the realised variance, RV_t, is calculated by aggregating intra-daily squared returns over a one-day horizon, t, using M sub-intervals:

$$RV_t = \Sigma_{j=1}^{M} r_{j,t}^2, \tag{2.3}$$

[1] In order not to overshadow the main results in this paper, we relegate these results in Section E in the Supplement.

Table 2.1
Summary of competing models.

Model name	Author(s)	Model	Estimation method
HARL (benchmark model)	Corsi (2009)	$RV_t^l = \beta_1 + \beta_2 RV_{t-1}^l + \beta_3 RV_{w,t-1}^l + \beta_4 RV_{m,t-1}^l + \nu_t; \quad \nu_t \sim N(0, \sigma_v^2)$	OLS
HARLQ	inspired by (Bollerslev et al., 2016) & Buccheri and Corsi (2021)	$RV_t^l = \beta_1 + \beta_{2,t} RV_{t-1}^l + \beta_3 RV_{w,t-1}^l + \beta_4 RV_{m,t-1}^l + \nu_t; \quad \nu_t \sim N(0, \sigma_v^2)$ where $\beta_{2,t} = \beta_2 + \gamma \frac{\sqrt{RQ_{t-1}}}{RV_{t-1}}$	OLS
HARSL	Bekierman and Manner (2018)	$RV_t^l = \beta_1 + (\beta_2 + \lambda_t) RV_{t-1}^l + \beta_3 RV_{w,t-1}^l + \beta_4 RV_{m,t-1}^l + \nu_t; \quad \nu_t \sim N(0, \sigma_v^2)$ where $\lambda_t = \phi \lambda_{t-1} + \eta_t; \quad \eta_t \sim N(0, \sigma_\eta^2)$	Maximum Likelihood using Kalman filter
TVCHAR	Chen et al. (2018)	$RV_t^l = \beta_1(\tau_t) + \beta_2(\tau_t) RV_{t-1}^l + \beta_3(\tau_t) RV_{w,t-1}^l + \beta_4(\tau_t) RV_{m,t-1}^l + \nu_t; \quad \nu_t \sim N(0, \sigma_v^2)$ Using first-order Taylor expansion: $\beta_j(\tau_t) \approx \beta_j(\tau) + \beta_j'(\tau)(\tau_t - \tau)$ where $\beta_j'(\tau)$ is the first-order derivative of $\beta_j(\tau)$	Local Linear Method
SHARK	Buccheri and Corsi (2021)	$RV_t^l = \alpha_t + \nu_t; \quad \nu_t \sim N(0, h_t)$ $\alpha_t = \beta_{1,t} + \beta_{2,t} RV_{t-1}^l + \beta_{3,t} RV_{w,t-1}^l + \beta_{4,t} RV_{m,t-1}^l + \eta_t; \quad \eta_t \sim N(0, q_t)$ Let $f_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \beta_{4,t}, log q_t)'$ The update rule is: $f_{t+1} = f_t + C s_t$ where s_t is a function of the Kalman filter prediction error and its covariance matrix. $V_t = \frac{\Sigma_{j=1}^M r_{j,t}^4}{(\Sigma_{j=1}^M r_{j,t}^2)^2}$, a consistent estimate of the measurement error variance of RV^l, is selected as a proxy for h_t.	Maximum Likelihood using Kalman filter

where $r_{j,t} = \ln(P_{(t-1)M+j}) - \ln(P_{(t-1)M+(j-1)})$ is the intra-day return of the jth sub-interval within the tth day, and $P_{(t-1)M+j}$ is the asset price at the start of the jth interval computed as the average of the closing and opening prices of intervals $j-1$ and j, respectively.[2] Several models can be used in order to forecast the daily RV; our benchmark model is the Heterogeneous Autoregressive (log) Realised Variance (HARL) of Corsi (2009) defined as:

$$RV_t^l = \beta_1 + \beta_2 RV_{t-1}^l + \beta_3 RV_{w,t-1}^l + \beta_4 RV_{m,t-1}^l + \nu_t, \quad \text{with} \quad \nu_t \sim N(0, \sigma_v^2), \tag{2.4}$$

where RV_t^l denotes the log-transformation of daily RV_t, and $RV_{w,t}^l$ and $RV_{m,t}^l$ denote the weekly and monthly log-transformations of the RV realised at time t, respectively, computed over a recursive rolling window of fixed length (week or month) as $RV_{w,t}^l = \Sigma_{i=1}^5 RV_{t-i}^l/5$ and $RV_{m,t}^l = \Sigma_{i=1}^{22} RV_{t-i}^l/22$ respectively. As common wisdom would suggest, by using the log-transformation, the HARL model of Eq. (2.4), and indeed all the other variants estimated on the log series, are less affected by the huge peaks of RV. Nevertheless, the HARL model is liable to suffer, albeit less severely, from several issues such as measurement error, time variation in the parameters, etc... Hence, the literature has developed several extensions of the basic HARL, which we also use in our paper by way of comparison and which we briefly review here. Bollerslev et al. (2016, 2018) propose an extension (called the HARQ model) in order to deal with the estimation error of the RV, using the "realised quarticity" defined as $RQ_t = \frac{M}{3} \Sigma_{j=1}^M r_{j,t}^4$. This approach has the major advantage of being estimable using OLS, and it can be naturally extended to the HARL set-up (resulting in a model which we call the HARLQ specification). Further, again in order to address the measurement error, Bekierman and Manner (2018) - building on the original works by Barndorff-Nielsen and Shephard (2002) and Bollerslev et al. (2016) - propose the HARSL model, which, in essence, is a state-space HARL model which assumes a latent Gaussian AR(1) process for the daily coefficient β_2. The model is estimated using maximum likelihood with a standard Kalman filter, and it on average outperforms the HARL model in forecasting the RV. The success of the HARSL model is due to its ability to capture other sources of temporal variation in addition to the variance of the measurement error. However, Bekierman and Manner (2018) note that the maximum likelihood estimator of the model is inefficient, and allowing all coefficients to follow an AR process is computationally challenging to perform using their employed estimation method. Thus, in a related contribution, Chen et al. (2018) introduce a novel approach – based on a local linear smoothing method – to estimate the HARL model (called the TVCHAR model), which assumes time-varying coefficients of an unknown functional form. The TVCHAR specification is found to also outperform the benchmark HARL model, especially over longer forecasting horizons. In a similar vein, Buccheri and Corsi (2021) propose the SHARK model, which accounts for time-varying coefficients and heteroskedastic error terms while also handling measurement errors. They find that their model produces moderate improvements in one-day-ahead forecasts but is more effective for longer-term forecasting. In Table 2.1, we summarise the set of HARL versions that we consider in this paper to compare alongside our proposed models, described in the next section.

[2] We point out that Zhang et al. (2005) provide a discussion on optimising the sampling frequency for the estimation of RV_t. However, Buccheri and Corsi (2021) show that the relative forecast performance of models with time-varying coefficients, such as the ones discussed in this paper, is independent of the sampling frequency. Therefore, for conciseness, we use sub-intervals of length 300 s in constructing the daily RV series. The latter defines 78 intraday sub-intervals and combines balanced information from high-frequency data and microstructure effects (Andersen et al., 2001).

3. Methodology

As the (short) literature review in the previous section demonstrates, research within the HARL family of models has been very active. In this section, we: propose two variants of the benchmark HARL model of Eq. (2.4), which account for the possible mis-specifications in the HARL set-up by considering a quite general specification for the law of motion of the time-varying coefficients (Section 3.1); briefly discuss (Bayesian) estimation (Section 3.2) and the priors we have chosen for our empirical exercise; and report the forecasting algorithm using both variants of the basic HARL model (Section 3.3).

3.1. The SHARP and SHARP-SV models

We describe the two variations of the proposed model. In order for the presentation not to be overshadowed by model complexity, here we present a simplified version of our model, where innovations are *i.i.d.* and Gaussian. However, the measurement Eq. (3.1) below is quite flexible, and the model can be extended to include autoregressive structure, deterministics such as a constant, (linear or nonlinear) trends, or seasonal dummies. Also, in principle, other specifications than Eq. (3.5) below for time-varying heteroskedasticity are possible. For extensions, we refer to Section 5, where we show that using Bayesian estimation with the Particle Gibbs sampler within this simplified class of dynamic state-space models is sufficient to capture several features of financial time series, such as: correlation between innovations, presence of diffusive leverage and multiple regimes, and heavy tails. Further, as mentioned in the Introduction, in Section E in the Supplement, we derive a characterisation of the dependence structure of the observations y_t under a more complex version of the model below, which also nests (3.1)–(3.5).

Consider the following state-space specification,

$$y_t = x_t'\beta_t + \nu_t, \tag{3.1}$$

where $1 \leq t \leq n$, and $\nu_t \sim i.i.d.\mathcal{N}(0, \sigma_v^2)$; we model the law of motion of β_t as

$$\beta_{j,t} = \alpha_j + \rho_j \beta_{j,t-1} + \varepsilon_{j,t}, \tag{3.2}$$

where $\left|\rho_j\right| < 1$ for $1 \leq j \leq k$. Whilst spelt out for a general model in terms of $\left\{y_t, x_t'\right\}$, in our context the measurement Eq. (3.1) describes the dependent variable $y_t = RV_t^l$ in terms of the covariates $x_t = (1, RV_{t-1}^l, RV_{w,t-1}^l, RV_{m,t-1}^l)'$ in (2.4), of the state vector, β_t, and of the disturbances, ν_t. The transition Eq. (3.2) describes the evolution of the coefficients over time, with the (stationary) AR process of the intercept also capturing the residuals autocorrelation in the measurement equation.

In the state Eqs. (3.2), we use two different specifications, which characterise the SHARP and the SHARP-SV model respectively; in the former case (SHARP), we model the innovations as *i.i.d.* Gaussian, viz.

$$\varepsilon_{j,t} \sim i.i.d.\mathcal{N}\left(0, \sigma_{\varepsilon,j}^2\right), \tag{3.3}$$

independent across $1 \leq j \leq k$, whereas in the latter case (SHARP-SV), we assume a stochastic volatility (SV), equation-by-equation, process, i.e.

$$\varepsilon_{j,t} = h_{j,t}^{1/2}\eta_{j,t}, \tag{3.4}$$

$$\ln h_{j,t} = \gamma_j + \delta_j \ln h_{j,t-1} + u_{j,t}, \tag{3.5}$$

where $\left\{\eta_{j,t}, 1 \leq j \leq k, 1 \leq t \leq n\right\}$ is a mean zero, unit variance process independent across j and t, and independent of $\left\{u_{j,t}, 1 \leq j \leq k, 1 \leq t \leq n\right\}$, with $u_{j,t} \sim i.i.d.\mathcal{N}(0, \sigma_{u,j}^2)$ also independent across j, and $\left|\delta_j\right| < 1$. While the unconditional first and second moments of the stationary solution of (3.2) can be shown to be constant, its conditional second moment can change over time; hence, (3.4)–(3.5) allows for *(a)* conditional heteroskedasticity in the main Eq. (3.1), through the intercept term (say $\beta_{1,t}$); and *(b)* also for conditional heteroskedasticity in the coefficients process. Model (3.4)–(3.5) is a standard (univariate) stochastic volatility model (see Taylor, 1982); we refer to Harvey et al. (1994) for extensions to multivariate settings, and to Kokoszka et al. (2025) for a functional version.

3.2. (Bayesian) estimation of SHARP and SHARP-SV models

We review the estimation technique employed, and the priors used in our empirical exercise.

In order to estimate the model, we use a modified version of the sequential Monte Carlo method known as the Particle Gibbs (PG) sampler discussed in Andrieu et al. (2010). Whilst the details are in Section C of the Supplement, here we offer a bird's-eye view on the estimation algorithm. The latent variables in our model are $\lambda_t = \left(\beta_t', h_t'\right)'$ where $\beta_t = \left(\beta_{1,t}, \ldots, \beta_{k,t}\right)'$ and $h_t = \left(h_{1,t}, \ldots, h_{k,t}\right)'$, whose prior can be described by $p(\lambda_t|\lambda_{t-1}, \theta)$, with θ a vector containing all the static parameters; the joint posterior is denoted as $p(\theta, \lambda_{1:T}|y_{1:T})$ and also studied in Section C of the Supplement; finally, in the PG sampler, we draw the structural parameters, as usual, from their posterior conditional distributions $p\left(\theta|\lambda_{1:T}, y_{1:T}\right)$, reported again in Section C of the Supplement. By doing so, we can avoid mixture approximations or other Monte Carlo procedures that need considerable tuning and may not have good convergence properties. The latent variables can be integrated out of the joint posterior using the procedure by Creal and Tsay (2015), described in Section D of the Supplement.

Our choice of conjugate priors is as follows:

$$\alpha_j, \gamma_j \sim \mathcal{N}(0, 1) \quad \rho_j, \delta_j \sim \mathcal{N}(0.5, 1)\,\mathbb{I}_{\rho_j \in (0,1)}\mathbb{I}_{\delta_j \in (0,1)} \quad \sigma_{\varepsilon,j}^2 \sim \Gamma(6.5, 0.5) \quad \sigma_v^2 \sim \Gamma(6.5, 0.5),$$

for all $1 \leq j \leq k$, where $\Gamma(a, b)$ denotes the Gamma distribution with shape a and scale b. In the above set of priors: we restrict the estimation range of ρ and δ between $(0, 1)$, to guarantee stationarity and also to reflect the belief that coefficients are positively autocorrelated; we note that the prior of γ and δ can be made more flexible in this exercise with a prior mean of 0; we choose the conjugate prior specification of the variances, $\sigma^2_{u,j}$ and σ^2_v, to have a low mean, reflecting the prior belief that while the coefficients are time-varying, we do not expect high jumps in the magnitude of the coefficients from one day to another.[3]

3.3. Forecasting the log-RV using SHARP and SHARP-SV

To reduce the computational cost of the model, we implement an efficient strategy in our MCMC and particle filtering process. Instead of running the full MCMC and particle filtering for each observation, we perform these computations every 10 observations. During these runs, we use 100 particles for the particle filtering, and the estimated posterior statistics are then applied to the following 10 observations. Within these 10 observations, we estimate the states by running the particle filtering using 1000 particles, so as to balance computational efficiency and model accuracy. We use an in-sample estimation window of approximately four years, and conduct 1000 MCMC iterations, with the first 300 iterations discarded as burn-in. By only updating the MCMC estimates every 10 observations, and using the resulting estimates for the subsequent 10, we significantly reduce computational demands while maintaining forecasting capabilities.

We report below the algorithm for the forecasting of RV^l_t using both the SHARP and SHARP-SV models. Recall that n is the total number of observations, and let: is be the in-sample estimation window size (approximately four years), $nsim$ be the number of MCMC iterations, and $nburn$ the burn-in sample.

(1) For $T = is, is + 10, \ldots, n - 1$

  (a) For $i = 1, \ldots, nsim$

    (i) Draw λ_t for $t = is - T + 1, \ldots, T$, as illustrated in Section D of the Supplement.
    (ii) Sample the parameters of Eqs. (3.2) and (3.5) using their posterior distributions (see Section C of the Supplement).
    (iii) Forecast the states in Eqs. (3.2) and (3.5) then forecast $\widehat{RV^l}_{T+1,i}$ for $i = nburn + 1, \ldots, nsim$ accordingly.

  (b) Estimate $\widehat{RV^l}_{T+1} = \frac{1}{nsim - nburn} \Sigma_{i=nburn+1}^{nsim} \widehat{RV^l}_{T+1,i}$

(2) For $T = is + 1, is + 9, is + 11, \ldots$

  (a) Draw λ_t for $t = is - T + 1, \ldots, T$, as illustrated in Section D of the Supplement.
  (b) Forecast the states in Eqs. (3.2) and (3.5) then forecast $\widehat{RV^l}_{T+1}$ accordingly.

Forecasts of the Realised Variance are then computed based on the expectation of a log-normal distribution,[4] as follows:

$$\widehat{RV}_{t+1} = \exp\left(\widehat{RV^l}_{t+1} + \frac{\hat{\omega}^2_{t+1|t}}{2}\right), \tag{3.6}$$

where:

$$\begin{aligned}\hat{\omega}^2_{t+1|t} = \; & \hat{\sigma}^2_v + \widehat{Var}(\beta_{1,t+1|t}) + (RV^l_t)^2 \times \widehat{Var}(\beta_{2,t+1|t}) \\ & + (RV^l_{w,t})^2 \times \widehat{Var}(\beta_{3,t+1|t}) + (RV^l_{m,t})^2 \times \widehat{Var}(\beta_{4,t+1|t}).\end{aligned}$$

Here, the first term in the expression of $\hat{\omega}^2_{t+1|t}$ - denoted as $\hat{\sigma}^2_v$ - is the variance of the measurement equation, whereas the subsequent terms represent the variance of each of the state equations entering through the coefficients in the measurement equation. In the SHARP model, $\widehat{Var(\beta_{j,t+1|t})} = \hat{\sigma}^2_{\epsilon,j}$, for $j = 1, \ldots, 4$; conversely, in the SHARP-SV model, $\widehat{Var(\beta_{j,t+1|t})} = \exp\left(\ln(\widehat{h_{j,t+1}}) + \hat{\sigma}^2_{u,j}/2\right)$.

4. Empirical study

In this section, we: describe the datasets employed in Section 4.1; analyse the estimated coefficients of our volatility model(s) in Section 4.2; and report a comprehensive out-of-sample forecasting analysis in Section 4.3. For brevity, we report only a selection of the main results; the full-blown set of results is in Section A of the Supplement.

[3] We fixed the hyperparameters primarily to allow for more straightforward implementation of the Gibbs sampler and reduce the computational burden. In a set of unreported experiments, we tried varying the hyperparameters, but we noted no material differences in the empirical findings, which seems to suggest that our results are not sensitive to the specific choice of hyperparameters.

[4] We note that we *always* compute forecasts based on the Gaussianity assumption, which can of course be mis-specified. As our simulations in Section 5.2 show, however, this does not impair the predictive accuracy of our methodology.

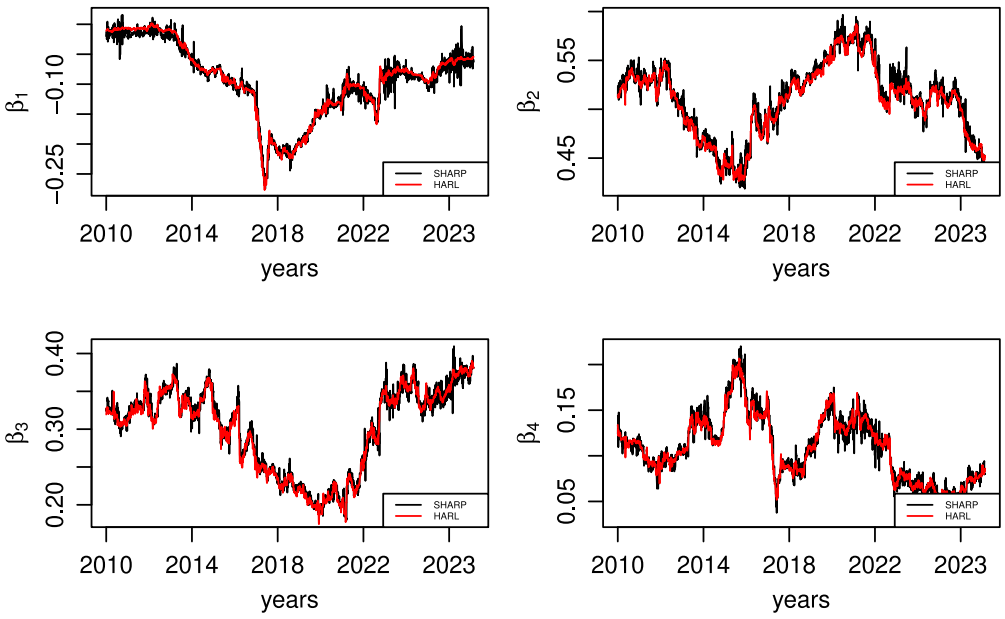

Fig. 4.1. Out-of-sample estimated coefficients, $\hat{\beta}_{ij}$, by SHARP and HARL models on SPY-ETF realised variance in the period January 1st, 2006 to September 8th, 2023.

4.1. Data

We use two datasets, and the VIX index. The first dataset comprises 4451 trading days of SPY-ETF, representing the US stock market, and ten ETFs representing its ten economic sectors, spanning from January 3rd, 2006 to September 8th, 2023. The second dataset includes $4,277$ trading days of twenty individual NYSE stocks, covering the period from January 3rd, 2000 to December 31st, 2016. We also use the VIX index, from July 1st, 2003 to December 29th, 2023, with a sample size of 5161 observations. For both datasets and the VIX index, daily RV was computed from tick-level price observations obtained from TickWrite.[5] The selected stocks represent different market sectors and vary in terms of market activity and volatility: the SPY serves as a widely recognised proxy for aggregate market behaviour; the sector ETFs and stocks offer insight into cross-sectional and sector-specific volatility dynamics; and the inclusion of the VIX adds a fundamentally different type of volatility measure (one that is anticipatory/forward-looking, rather than historical), which contrasts with the RV-based metrics — thus striking a balance between coverage, relevance, and comparability. The chosen sample period includes significant market events such as the 2008 global financial crisis and the COVID-19 pandemic, providing a diverse range of market conditions. Table A.1 in Section A of the Supplement contains the descriptive statistics of RV for all datasets.

4.2. Analysis of temporal volatility coefficients

Using a rolling window approach with an estimation window of 1000 daily observations (approximately four years), we plot the estimated coefficients from the SHARP and HARL models, used in the out-of-sample forecasting period for the realised variances of the SPY-ETF and the VIX respectively, in Figs. 4.1 and 4.2 respectively.

We note that the estimated intercept $\hat{\beta}_{1,t}$ fluctuates significantly, dropping during calm periods and rising during crises (in fact, $\hat{\beta}_{1,t}$ tends to drop below its corresponding average during tranquil periods), while the intercept of the HARL model remains fixed; this suggests that SHARP better adapts to changing baseline levels of realised volatility. Indeed, SHARP shows varying persistence across

[5] TickWrite is a commercial database that provides data for futures, index, and equity markets. Tick data is sourced from NYSE's TAQ (Trade and Quote) database and is adjusted for ticker mapping, code filtering, price splits, and dividend payments. More information can be found at https://www.tickdata.com/ https://www.tickdata.com/

In the construction of our databases, the selection criteria for individual stocks were as follows: (a) only stocks continuously traded over the full sample period were considered; (b) in order to address liquidity and staleness biases, we selected stocks ranking in the top 15th percentile by trading volume, which also tend to fall in the bottom 20th percentile for zero returns; (c) in order to capture potential sector-specific intraday dynamics, we ensured representation from all GICS sectors, choosing two qualifying stocks per sector.

The data were aggregated from the tick level using previous-tick interpolation, and sampled at 5-minute intervals, a standard frequency in the high-frequency literature that balances bias and variance (see Ait-Sahalia et al., 2005; and Hansen and Lunde, 2006).

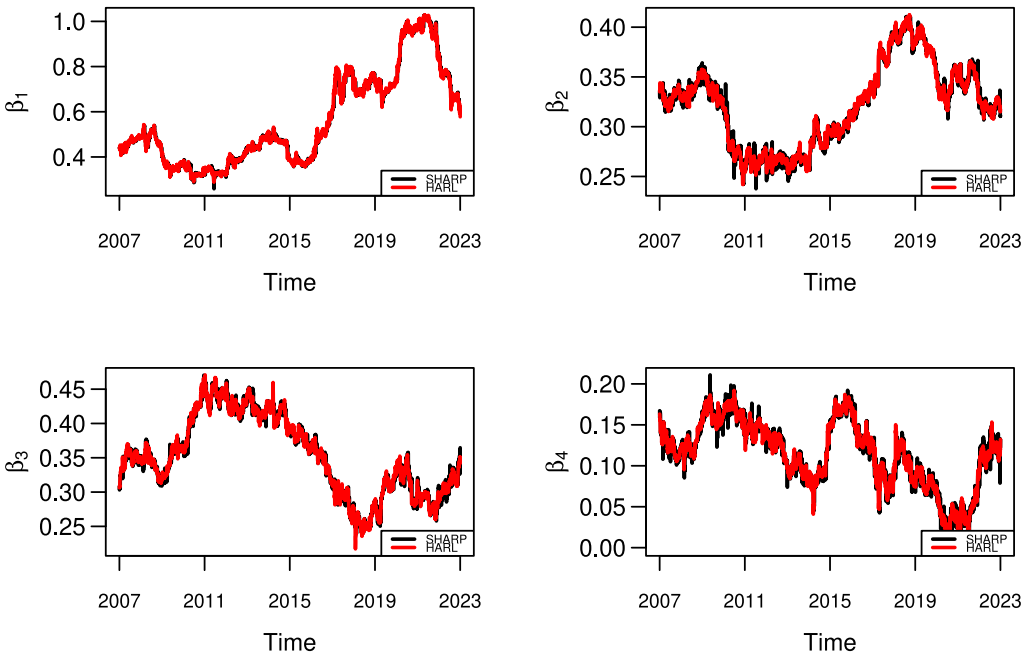

Fig. 4.2. Out-of-sample estimated coefficients, $\hat{\beta}_{tj}$, by SHARP and HARL models on VIX realised variance in the period July 24th, 2007 to December 29th, 2023.

Table 4.1
Correlation matrix of β_{tj} parameters obtained by estimating the SHARP and SHARP-SV models on SPY-ETF realised variance in the period January 1st, 2000 to September 8th, 2023.

SHARP					$SHARP_{sv}$				
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_1$	1	−0.1740	0.7049	−0.0389	$\hat{\beta}_1$	1	−0.1431	0.6355	−0.0243
$\hat{\beta}_2$	−0.1740	1	−0.5772	−0.1763	$\hat{\beta}_2$	−0.1431	1	−0.4675	−0.1212
$\hat{\beta}_3$	0.7049	−0.5772	1	−0.4095	$\hat{\beta}_3$	0.6355	−0.4675	1	−0.3840
$\hat{\beta}_4$	−0.0388	−0.1763	−0.4095	1	$\hat{\beta}_4$	−0.0243	−0.1212	−0.3840	1

Table 4.2
Correlation matrix of β_{tj} parameters obtained by estimating the SHARP and SHARP-SV models on VIX realised variance in the period July 24th, 2007 to December 29th, 2023.

SHARP					$SHARP_{sv}$				
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_1$	1.0000	0.6318	−0.7545	−0.8420	$\hat{\beta}_1$	1.0000	0.6292	−0.7552	−0.8427
$\hat{\beta}_2$	0.6318	1.0000	−0.9344	−0.3613	$\hat{\beta}_2$	0.6292	1.0000	−0.9342	−0.3562
$\hat{\beta}_3$	−0.7545	−0.9344	1.0000	0.4049	$\hat{\beta}_3$	−0.7552	−0.9342	1.0000	0.4063
$\hat{\beta}_4$	−0.8420	−0.3613	0.4049	1.0000	$\hat{\beta}_4$	−0.8427	−0.3562	0.4063	1.0000

daily, weekly, and monthly horizons, reflecting evolving investor responses at different frequencies. HARL, by contrast, imposes static weights, missing this flexibility. The daily coefficient becomes more prominent during periods of uncertainty, such as the COVID-19 period, whereas the weekly coefficient shows a decline during these times; further, the monthly coefficient increases during the COVID-19 period. This phenomenon during times of financial market uncertainty can be attributed to the "primacy" and "recency" effects; the current long-term conditions, reflected in the monthly average of RV^l, can be considered primary information, while recent information is represented by the daily RV^l. When predicting short-term (daily) volatility during periods of uncertainty, primary and recent information become more relevant than intermediate information, such as the weekly average volatility.

In Tables 4.1 and 4.2, we report the correlations between estimated parameters in the SHARP and SHARP-SV models for the realised variances of the SPY-ETF and the VIX respectively. The tables indicate a negative correlation between the weekly and daily coefficients, as well as between the weekly and monthly coefficients, whereas the correlation between the daily and monthly coefficients is only weakly negative.

4.3. Out-of-sample comparative analysis

We report an out-of-sample analysis of the SHARP and SHARP-SV models, along with a set of competing models, in forecasting the daily, weekly, and monthly realised Variance (RV) of SPY-ETF, VIX, twenty NYSE individual stocks, and ten economic sector ETFs. We report the out-of-sample forecasts using a rolling window approach with an estimation window of 1000 daily observations (approximately four years).[6] For the NYSE stocks dataset, the in-sample period covers 1000 observations, approximately from January 2000 to December 2003. The out-of-sample period spans from January 2004 to December 2016. For the SPY-ETF and sector ETFs dataset, the in-sample period also uses the first 1000 observations, approximately from January 2006 to December 2009, with the out-of-sample period from January 2010 to September 2023. For the VIX series, the in-sample period also uses the first 1000 observations, approximately from July 2003 to July 2007, with the out-of-sample period from August 2007 to December 2023.

In our horse-race, we evaluate the performance of our SHARP and SHARP-SV models, also in comparisons with other, existing set-ups. For example, the comparison between HARL and HARLQ investigates whether accounting for measurement error in the RV^l series improves the forecasts by HARL, as observed by Bollerslev et al. (2016) when replacing the HAR model with their HARQ specification. Additionally, comparing HARSL and HARLQ outlines whether modelling the daily coefficient as an AR(1) process provides better forecasting performance than the more restrictive yet straightforward HARLQ model. The analysis between TVCHAR, SHARK, and SHARP demonstrates which time-varying specification of the coefficients yields better forecasts when compared to each other and to the aforementioned models. Finally, the comparison between SHARP and SHARP-SV reveals whether the inclusion of the stochastic volatility (SV) feature in the AR process of the coefficients enhances the forecasts. These comparisons are designed to evaluate the robustness and forecasting accuracy of the proposed models in relation to existing benchmarks.

We compare the accuracy of each model in forecasting RV over three horizons: daily, weekly, and monthly. The forecasts are compared with their corresponding actual values $\exp(RV_t^l)$, $\exp(RV_{w,t}^l)$, and $\exp(RV_{m,t}^l)$, respectively. Since we perform direct forecasting for all the models, we do not report the SHARK recursive forecasting for weekly and monthly horizons to maintain consistency across the models.

In order to provide a comprehensive analysis of forecast accuracy, we use the following loss functions: *(i)* the Mean Squared Error (MSE); *(ii)* the Mean Absolute Error (MAE); *(iii)* the heteroskedasticity-adjusted version of the mean squared error (HMSE), designed to adjust for the variability in y_t, thus being less sensitive to periods of high volatility and providing a scale-invariant measure of forecast accuracy (e.g. Wang et al., 2015), defined as

$$HMSE = \frac{1}{n-is}\sum_{t=is+1}^{n}\left(1-\frac{\hat{y}_t}{y_t}\right)^2,$$

where, as customary, we denote the actual and the predicted values as y_t and $\hat{y}_t$ respectively, and is is defined in Section 3.2; *(iv)* the heteroskedasticity-adjusted version of the Mean Absolute Error (HMAE), defined as

$$HMAE = \frac{1}{n-is}\sum_{t=is+1}^{n}\left|1-\frac{\hat{y}_t}{y_t}\right|;$$

and *(v)* the Quasi-Likelihood (QLIKE) loss function, defined as

$$QLIKE = \frac{1}{n-is}\sum_{t=is+1}^{n}\left(\frac{\hat{y}_t}{y_t}-\ln\left(\frac{\hat{y}_t}{y_t}\right)-1\right),$$

which penalises under-predictions of volatility. In all cases, in order to facilitate comparisons, we report the *relative* loss measures (say RL_M, referring to model M), where we evaluate the relative loss compared to the benchmark HARL specification, viz.

$$RL_M = \frac{LF_M}{LF_{HARL}}, \tag{4.1}$$

where LF_M denotes the value of the relevant loss function of model M, and LF_{HARL} the value taken by the same loss function for the HARL model. We present the RL_M for SPY-ETF and VIX data in Tables 4.3 and 4.4, respectively, where we also use the Model Confidence Set methodology of Hansen et al. (2011).[7]

Figs. 4.3 and 4.4 summarise the RL_M for the sector ETFs and the twenty NYSE individual stocks respectively; the full-blown set of numerical results is also reported in Tables A.2 and A.3 in Section A the Supplement, respectively.[8]

[6] Specifically, this involves continuously updating the dataset by removing the oldest observation and adding the most recent one, thus maintaining a constant window of 1000 observations. While in real-time forecasting one should ideally jump 5 and 22 observations for weekly and monthly forecasts respectively, we adopted the aforementioned method for the sake of simplicity — especially when updating the state equations for longer horizons. This ensures continuous updating of the dataset with the latest data, thereby maintaining the robustness and uniformity of our forecasts. In an unreported set of experiments, we note that the relative performance of our SHARP and SHARP-SV models compared to their competitors remains unaltered even when using the alternative approach.

[7] In essence, this approach tests the null hypothesis that all models are equally effective against the alternative that a smaller subset of models is superior. We select a p-value threshold of 0.25 based on range statistics: models with p-values below the 0.25 threshold are excluded from the superior subset, denoted by $\hat{M}_{75\%}$. In Tables A.2 and A.3 in Section A of the Supplement, we also report the frequency with which each model is included in the $\hat{M}_{75\%}$ for the sector ETFs and individual NYSE stocks, respectively.

[8] In Section A of the Supplement, we also present detailed results for the individual ETFs (Tables A.4–A.8) and the individual stocks (Tables A.9–A.13).

Table 4.3
Out-of-sample relative loss measure (4.1) of the models (with HARL being the benchmark) obtained by estimating the models on SPY-ETF realised variance over a rolling window of 1000 observations in the period January 3, 2006, to September 8, 2023.

Panel (a): Daily							
	HARL	HARLQ	HARSL	TVCHAR	SHARK	SHARP	SHARPsv
MSE	1[a]	1.0631	0.9986[a]	**0.9982**[a]	1.0007[a]	1.0454	1.0479
MAE	1	1.0069	0.9970	1.0000	0.9998	0.9764	**0.9624**[a]
HMSE	1	0.9994	0.9712	1.0002	0.9787	0.8579	**0.7121**[a]
HMAE	1	0.9990	0.9845	0.9998	0.9890	0.9290	**0.8549**[a]
QLIKE	1	0.9981	0.9833	0.9999	0.9873	0.9163	**0.8325**[a]
Panel (b): Weekly							
MSE	1	0.9921	0.7744	0.9937		**0.7138**[a]	0.9561[a]
MAE	1	0.9975	0.9596	0.9991		0.7706	**0.7022**[a]
HMSE	1	0.9993	0.6262	0.9984		0.4452	**0.2866**[a]
HMAE	1	0.9994	0.7583	0.9992		0.6867	**0.5550**[a]
QLIKE	1	0.9999	0.6726	0.9989		0.5353	**0.3820**[a]
Panel (c): Monthly							
MSE	1	0.9916	0.8111	0.9973		0.6109	**0.4890**[a]
MAE	1	0.9967	0.8126	0.9990		0.6151	**0.4381**[a]
HMSE	1	0.9992	0.5094	0.9958		0.2632	**0.1015**[a]
HMAE	1	0.9993	0.6089	0.9975		0.5230	**0.3183**[a]
QLIKE	1	0.9995	0.4977	0.9968		0.3336	**0.1379**[a]

[a] Indicates that the model is included in the $\hat{M}_{75\%}$ Model Confidence Set.

Table 4.4
Out-of-sample relative loss measure (4.1) of the models (with HARL being the benchmark) obtained by estimating the models on VIX realised variance over a rolling window of 1,000 observations in the period July 1, 2003, to December 29, 2023.

Panel (a): Daily							
	HARL	HARLQ	HARSL	TVCHAR	SHARK	SHARP	SHARPsv
MSE	1	1.0132	1.2956	1.0001	1.0086	**0.9998**	1.0055[a]
MAE	1	0.9970	1.3442	1.0001	1.0044	0.9708	**0.9559**[a]
HMSE	1	0.9399	4.7839	0.9995	1.0099	0.7337	**0.6619**[a]
HMAE	1	0.9818	1.6925	1.0000	1.0048	0.8622	**0.8206**[a]
QLIKE	1	0.9719	2.2312	0.9999	1.0057	0.8394	**0.7929**[a]
Panel (b): Weekly							
MSE	1	0.9919	7.0596	1.0000		0.8219	**0.6069**[a]
MAE	1	0.9924	0.8709	1.0000		0.9396	**0.7443**[a]
HMSE	1	0.9628	0.4948	0.9999		0.8002	**0.3729**[a]
HMAE	1	0.9861	**0.5660**[a]	1.0000		0.9051	0.6341
QLIKE	1	0.9777	**0.4192**[a]	0.9999		0.8137	0.4675
Panel (c): Monthly							
MSE	1	0.9916	1.2366	0.9999		0.7417	**0.6994**[a]
MAE	1	0.9971	**0.4589**[a]	0.9999		0.7727	0.8311
HMSE	1	0.9973	**0.1573**[a]	0.9996		0.3808	0.5881
HMAE	1	0.9964	**0.3508**[a]	0.9999		0.6599	0.8018
QLIKE	1	0.9918	**0.1724**[a]	0.9997		0.5090	0.6427

[a] Indicates that the model is included in the $\hat{M}_{75\%}$ Model Confidence Set.

We also use the *unconditional* predictive ability (uCPA) test by Giacomini and White (2006) to evaluate the out-of-sample predictions produced by the models. We perform pairwise tests of unconditional predictive ability over the full out-of-sample (OOS) period with a significance level of $\alpha = 0.05$. To ensure robustness, we conduct the same test using five different loss functions and examine the results across three forecasting horizons (see Fig. 4.5).

Overall, our results obtained using the RV^{l} of SPY-ETF, sector ETFs, individual stocks and the VIX are consistent throughout. For most individual stocks, incorporating some form of time-varying specification of the coefficients generally enhances the models forecasting accuracy, especially for longer forecasting horizons. With the exception of the MSE loss function in the case of daily forecasting, the SHARP and SHARP-SV models almost always exhibit the lowest relative loss (RL) using the SPY-ETF data and the average relative loss using both the sector ETFs and the twenty individual stocks data. A partial exception is encountered with the VIX, where, at lower frequencies, results are less clear-cut - although, at a daily frequency, our proposed models still deliver a superior forecasting performance than the other models. Introducing the SV feature to the AR process of the coefficients yields a further moderate improvement compared to the forecasts by the SHARP model. Thus, we find that the general specification of the

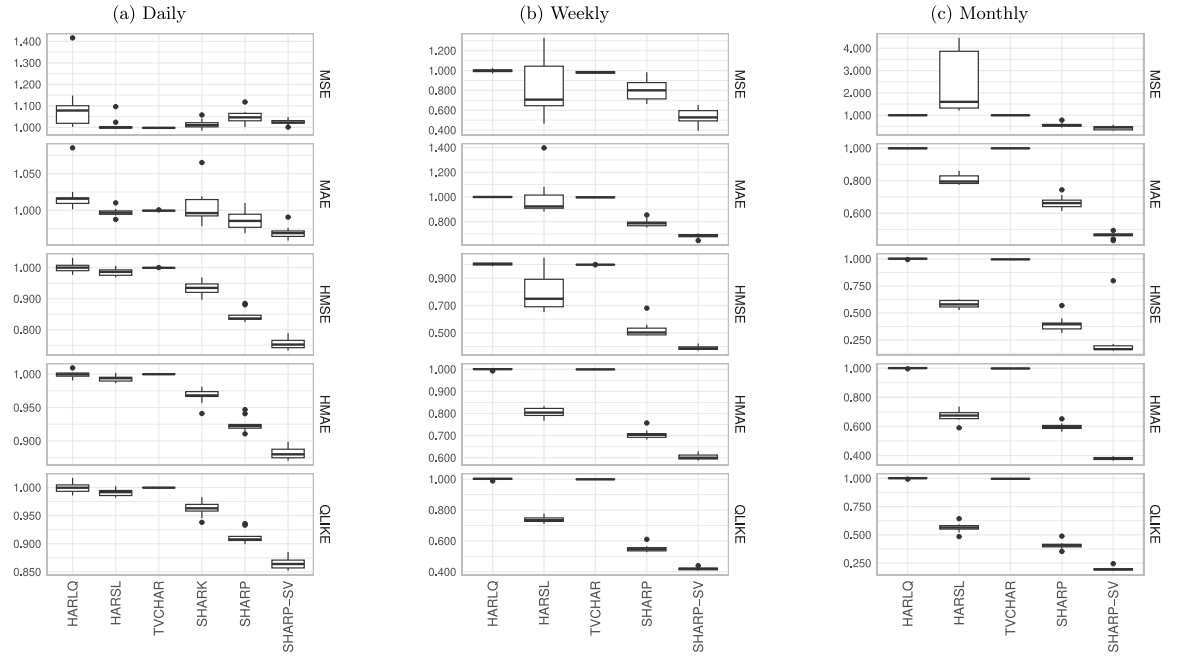

Fig. 4.3. Box plots of relative loss functions comparison — using the ten-sectors ETFs dataset.
Box plots of relative loss measures for five loss functions (MSE, MAE, HMSE, HMAE, QLIKE) across three forecasting horizons (Daily, Weekly, and Monthly) for each model, using the HARL as the benchmark model. The models were estimated on the realised variance for each of the ten sector ETFs over a rolling window of 1000 observations during the period from January 3, 2006, to September 8, 2023. One outlier point was removed from each series for the weekly and monthly forecasting horizons when computing the MSE for the HARSL model to prevent inflation of the measure. Applying the same principle to other models did not impact the MSE, so this adjustment was only done for the HARSL model.

time-varying coefficients in the SHARP-SV model leads to the most significant improvement in the forecasts of RV across all three forecasting horizons. These results are also robust across the five loss function measures used. As far as the other specifications are concerned, the HARSL model shows some moderate improvements – mainly at weekly and monthly horizons – suggesting that accounting for time variation in coefficients is more effective than simple corrections for measurement error. By contrast, HARLQ, TVCHAR, and SHARK generally fail to deliver meaningful improvements over the baseline HARL. Taken together, these results demonstrate that the flexibility of time-varying parameters, and especially the incorporation of stochastic volatility, is critical for achieving reliable forecast improvements across a wide cross-section of assets. Note also that our models, particularly the SHARP-SV model, are almost always included in the confidence set, while other models are frequently excluded with few exceptions.

In conclusion, we note that our model assumes Gaussianity of all innovations. Whilst assuming Gaussianity may not have an impact on the inference on coefficients, inference on the volatility process may be distorted in the presence of departures from normality such as excess kurtosis or nonzero skewness.[9] Seeing as our data are likely to exhibit such features, by way of misspecification analysis we carry out a test for Gaussianity inspired by Koopman and Scharth (2012). Whilst originally designed for a different specification, the test is based on a parametric bootstrap which we also use here, generating the pseudosamples using the estimated parameters and simulating all innovations from Gaussian distributions. Results in Tables A.14 and A.15 in Section A.1 in the Supplement show that the assumption of Gaussianity is indeed most often rejected; on the other hand, our empirical results show some degree of robustness to such misspecification, which are also reinforced by our Monte Carlo analysis in Section 5.2.

5. Monte Carlo study

We evaluate the performance of our models (and of Bayesian estimation) via a comprehensive set of simulations. In Section 5.1, we assess the performance of the Particle Gibbs estimator under correct model specification; in Section 5.2, we evaluate predictive accuracy under a possibly misspecified model.

[9] We are grateful to an anonymous Referee for pointing this out to us, and suggesting the testing approach used herein.

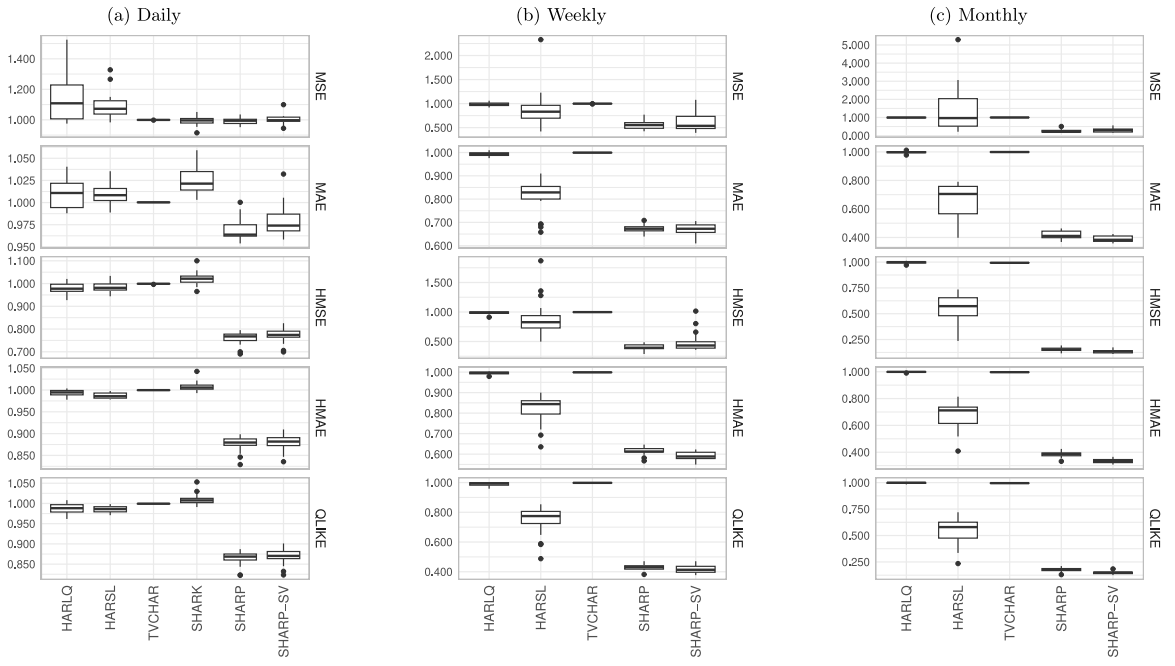

Fig. 4.4. Box plots of relative loss functions comparison — using the twenty individual NYSE stocks dataset.
Box plots of relative loss measures for five loss functions (MSE, MAE, HMSE, HMAE, QLIKE) across three forecasting horizons (Daily, Weekly, and Monthly) for each model, using the HARL as the benchmark model. The models were estimated on the realised variance for each of the twenty NYSE individual stocks over a rolling window of 1000 observations during the period from January 3, 2000, to December 31, 2016. One outlier point was removed from each series for the weekly and monthly forecasting horizons when computing the MSE for the HARSL model to prevent inflation of the measure. Applying the same principle to other models did not impact the MSE, so this adjustment was only done for the HARSL model.

5.1. The performance of the Particle Gibbs estimator

We study the performance of Particle Gibbs estimation under correct model specification. In particular, we evaluate *(a)* the ability to recover true parameters and latent states, and *(b)* the comparative performance of SHARP estimation via Particle Gibbs versus a Kalman filter based estimation.

Data are generated according to Eqs. (3.1)–(3.3), reported here for convenience

$$\begin{aligned} y_t &= x_t'\beta_t + v_t, \quad v_t \sim i.i.d.\mathcal{N}(0,\sigma_v^2), \\ \beta_{j,t} &= \rho_j \beta_{j,t-1} + \varepsilon_{j,t}, \quad \varepsilon_{j,t} \sim i.i.d.\mathcal{N}\left(0,\sigma_{\varepsilon,j}^2\right), \end{aligned}$$

with $1 \le t \le n$. Recall that the vector $x_t = (1,\ RV_{t-1}^l,\ RV_{w,t-1}^l,\ RV_{m,t-1}^l)'$ includes the daily, weekly, and monthly (lagged) log-realised RV, and that the innovations v_t and $\varepsilon_{t,j}$ are mutually independent. The parameter values are set as $\rho_j = 0.96$ for $1 \le j \le 4$, and $\left(\sigma_{\varepsilon,1}, \sigma_{\varepsilon,2}, \sigma_{\varepsilon,3}, \sigma_{\varepsilon,4}\right) = (0.15, 0.08, 0.08, 0.08)$; further, we set $\sigma_v = 0.02$ for the measurement error standard deviation. Initial values are set to $\beta_{1,1} = -0.5$, $\beta_{2,1} = 0.4$, $\beta_{3,1} = 0.3$, and $\beta_{4,1} = 0.15$, and are based on estimates from fitting a standard HARL model to the daily log-RV of the SPY index. These choices also align with decay structures documented in the literature (e.g. Andersen et al., 2007; and Corsi, 2009). We initialise the log-RV process y_t, by generating the first $1 \le t \le 22$ observations as $i.i.d.\mathcal{N}(0,1)$. Our experiments are based on a sample size $n = 1000$, and on 1000 replications.

Fig. 5.1 summarises the estimation accuracy for the four time-varying coefficients $\beta_{t,j}$ across three key metrics: normalised RMSE, 95% interval coverage, and interval width.[10] As can be seen, using the SHARP specification with Particle Gibbs systematically achieves the lowest estimation error, with average normalised RMSEs of 0.428, 0.724, 0.894, and 0.949 for $\beta_{t,1}$ through $\beta_{t,4}$, respectively. Estimation via Kalman filter performs comparably well for the short-horizon coefficients $\beta_{t,1}$ and $\beta_{t,2}$, but it exhibits higher errors for $\beta_{t,3}$ and $\beta_{t,4}$ (1.101 and 1.570 respectively), reflecting its limited flexibility in tracking longer-term dynamics. Turning to interval calibration, using the Particle Gibbs estimator achieves reasonably balanced coverage across all coefficients, with coverage rates ranging from 0.707 to 0.781, accompanied by normalised interval widths of 0.774, 1.541, 2.165, and 2.367

[10] The *normalised* RMSE is defined by computing (the square root of) $\sum_{t=1}^{T}\left(\hat{\beta}_{t,j} - \beta_{t,j}\right)^2/T$, and then dividing it by $\sigma_{\varepsilon,j}$; we use the same normalisation also for coverage.

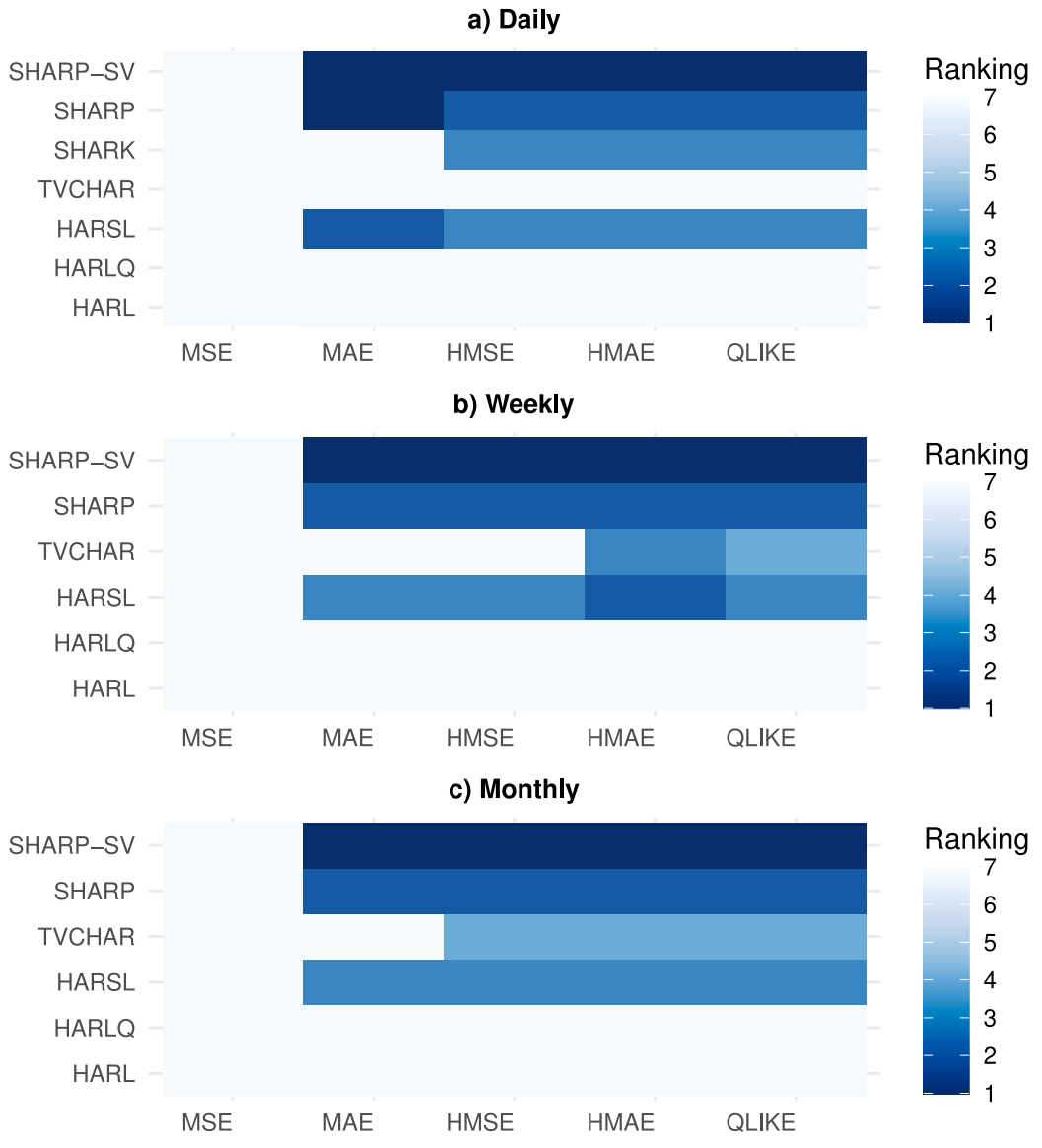

Fig. 4.5. Heatmap of Model Rankings Based on uCPA (Unconditional Conditional Predictive Ability).
Results of the uCPA test for out-of-sample forecasting performance between models, with each selected loss function obtained by estimating the models on SPY-ETF realised variance over a rolling window of 1000 observations during the period from January 1st, 2006 to September 8th, 2023. The rankings range from 1 (best) to 7 (least). A darker colour indicates a better model performance. If two models have the same rank (i.e., colour), they are considered equally good. The results are shown for three forecasting horizons in panels a, b, and c, representing daily, weekly, and monthly forecasts, respectively.

for $\beta_{t,1}$ through $\beta_{t,4}$, respectively. In contrast, Kalman filter estimation yields wider intervals 1.153, 2.527, 2.677, and 2.770 for $\beta_{t,1}$ through $\beta_{t,4}$; while delivering better coverage for $\beta_{t,1}$ and $\beta_{t,2}$ (0.866 and 0.919 respectively), it still under-covers $\beta_{t,3}$ and $\beta_{t,4}$ (with coverages 0.758 and 0.712 respectively).[11] Hence, Kalman filter estimation does not consistently yield better-calibrated coverage, especially for longer-horizon components, despite substantially wider intervals, which is reinforced by the variability in coverage across coefficients and replications (especially as far as higher-lag coefficients are concerned). Conversely, Particle Gibbs estimation exhibits tighter dispersion and fewer outliers across all metrics, suggesting robustness across simulated series.

[11] Naturally, the Kalman filter produces frequentist confidence intervals, based on the conditional state covariance matrices, while the particle filter generates Bayesian credible intervals from posterior samples. Although these intervals differ in interpretation, comparing their coverage and width offers insight into how each method quantifies uncertainty in recovering the latent states.

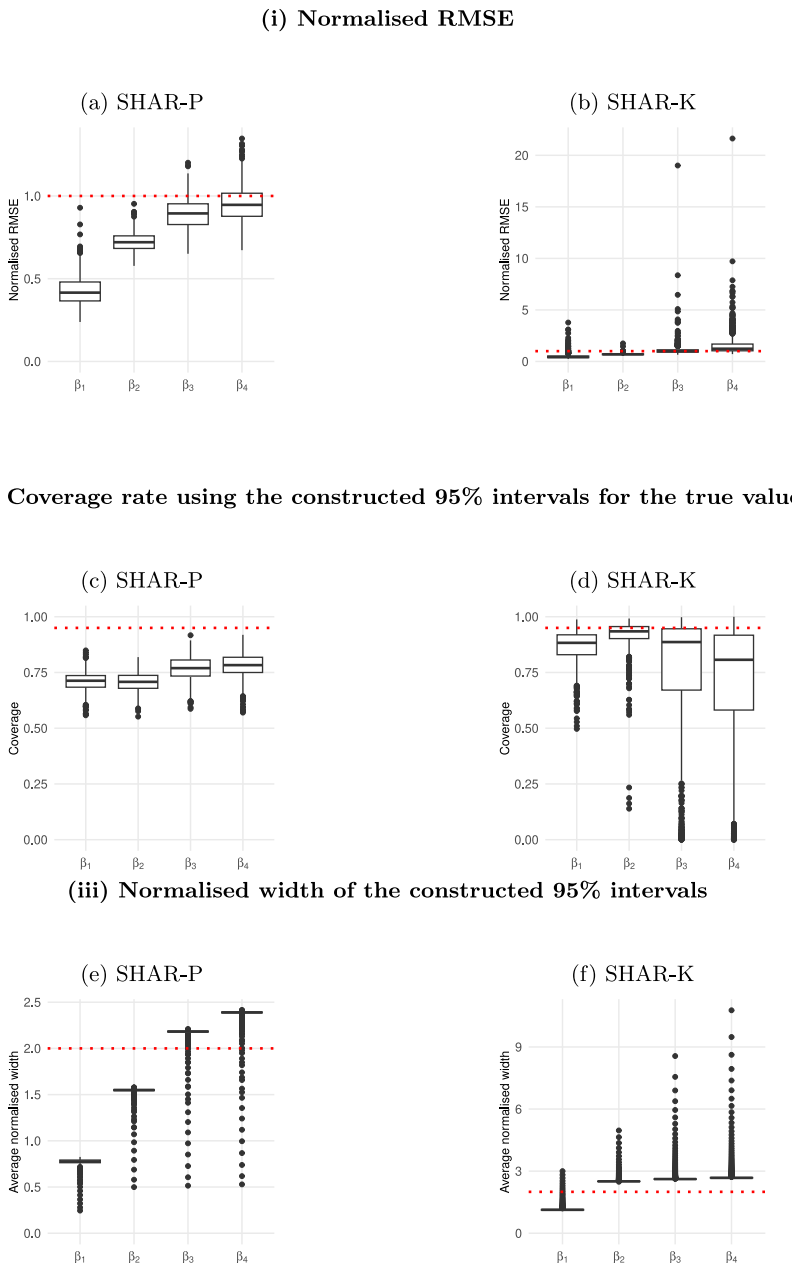

Fig. 5.1. Time-Varying coefficients - estimation performance.
Estimation results obtained under Particle Gibbs and Kalman filter are indicated as SHAR-P and SHAR-K respectively. The dotted red lines represent theoretical benchmark values (1 for RMSE, 0.95 for coverage, and 2 for interval width), corresponding to the expected variability of the latent coefficients under the DGP. Coverage values closer to one reflect correctly calibrated 95% intervals.

In Fig. 5.2, we assess the recovery of fixed parameters persistence (ρ_j, $1 \leq j \leq 4$), innovation volatilities ($\sigma_{\epsilon,j}$), and measurement error (σ_v). Similarly to the previous results, Kalman filter estimation recovers the persistence coefficients ρ_j accurately for short-horizon coefficients (0.958 and 0.954 on average across simulations for ρ_1 and ρ_2) but it underestimates persistence in β_3 and β_4 (with ρ_3 and ρ_4 being, on average, 0.895 and 0.873). Particle Gibbs estimation displays more balanced estimates across all lags (with ρ_j, on average, equal to 0.946, 0.927, 0.918, and 0.913 for $1 \leq j \leq 4$), despite having a slight downward bias. In terms of state innovation volatilities $\sigma_{\epsilon,j}$, Kalman filter estimation yields average values of estimates $(0.146, 0.079, 0.071, 0.071)$, closely matching the true values; Particle Gibbs also performs well, but with a larger downward bias, delivering average estimates $(0.118, 0.069, 0.067, 0.067)$. Finally, turning to the measurement error variance σ_v, Kalman filter recovers the true value nearly exactly, whereas Particle Gibbs slightly overestimates (estimated values are 0.020 and 0.032 respectively). As before, Particle Gibbs corresponds to tight, symmetric

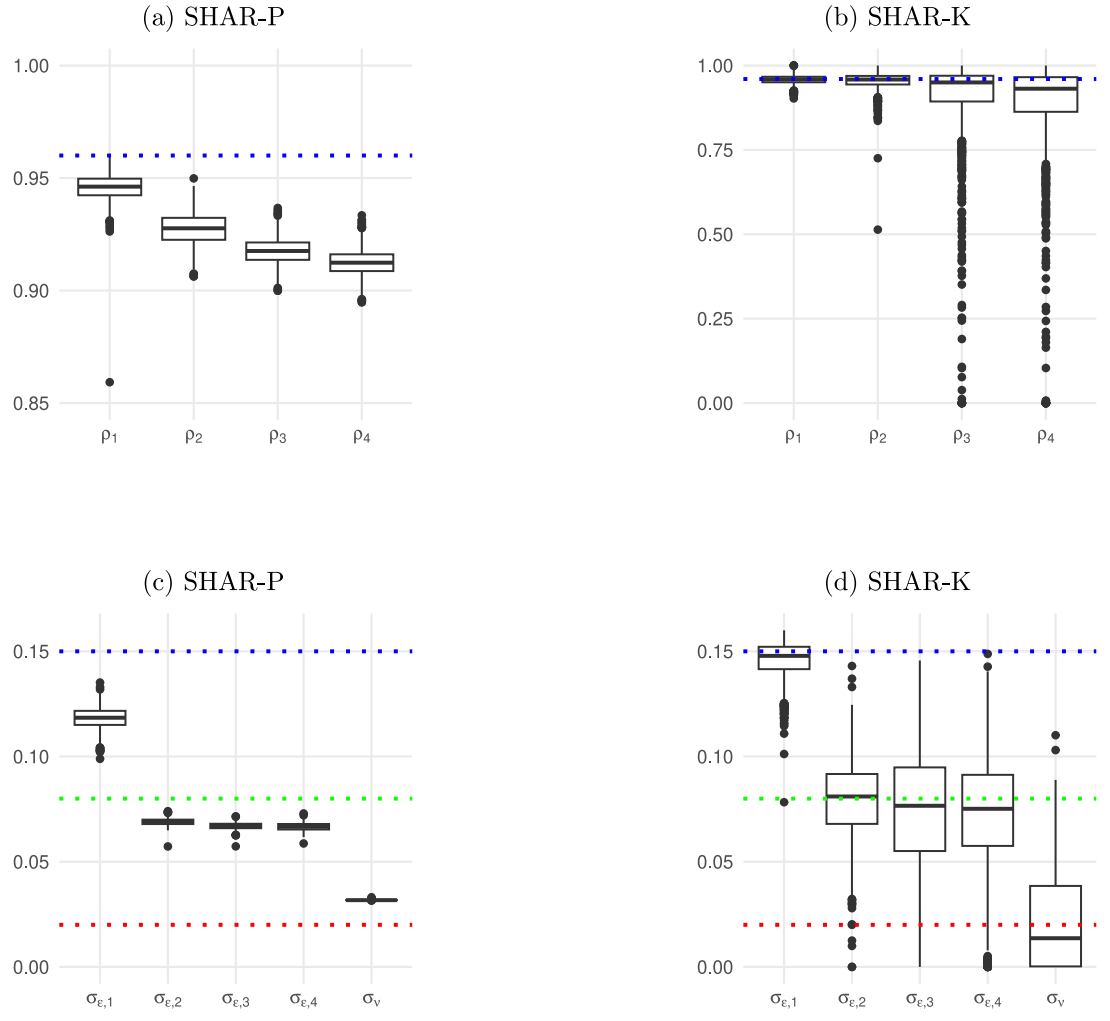

Fig. 5.2. Fixed parameters - estimation performance.
The blue dotted line in panels (a) and (b) indicates the true value of $\rho = 0.96$ used in the data-generating process (DGP). Recall that estimation results obtained under Particle Gibbs and Kalman filter are indicated as SHAR-P and SHAR-K respectively. Panel (c) and (d) show boxplots of the estimated innovation standard deviations $\sigma_{\varepsilon,j}$ and the measurement error standard deviation σ_{ν} under SHAR-P and SHAR-K, respectively; the blue dotted line corresponds to the true value of $\sigma_{\varepsilon,1} = 0.15$, the green line marks the true value of $\sigma_{\varepsilon,j}$, $j = 2, 3, 4$, and the red line marks the true value of the measurement error standard deviation $\sigma_{\nu} = 0.02$.

distributions around the true values, with very few outliers, indicating robust and consistent performance; Kalman filter performs comparably, though it shows slightly greater dispersion in the longer-lag persistence estimates.

The distilled essence of our simulations is that the relative strengths of Kalman filter and Particle Gibbs are *complementary*. The former performs better at estimating fixed parameters, and it is also superior in terms of coverage for the "short-run" parameters - i.e. the intercept $\beta_{t,1}$ and the coefficient multiplying the daily log RV, $\beta_{t,2}$. On the other hand, the Particle Gibbs estimator performs better in terms of RMSE when estimating "long-run" parameters (i.e. $\beta_{t,3}$ and $\beta_{t,4}$, associated to the weekly and monthly log RVs), and it has – in general – less dispersion and fewer outliers across experiments relative to the Kalman filter. As we show in the next section,[12] these relative advantages of Particle Gibbs translate into superior predictive accuracy. Hence, while the Kalman filter remains useful for efficient inference on fixed parameters and near-term components, Particle Gibbs offers a more robust alternative when the primary objectives are accurate state recovery and predictive performance.

[12] See also Section B2 in the Supplement.

5.2. Forecasting accuracy with mis-specified models

In order to assess the robustness of our approach, we now investigate predictive accuracy under a possibly mis-specified model. We consider the following DGP, where we generate a series of returns along similar lines to Jacquier et al. (2004) and Koopman and Scharth (2012), viz.

$$y_t = \mu + \exp\left(\frac{h_t}{2}\right)\sqrt{\lambda_t}\epsilon_t + J_t\mathcal{Z}_t^y. \tag{5.1}$$

In (5.1): λ_t is a scaling factor which we model according to two schemes, with $\lambda_t = 1$ for all t, or drawn from an inverse-Gamma distribution, i.e.

$$\lambda_t \sim i.i.d.\mathcal{IG}\left(\frac{\upsilon}{2}, \frac{\upsilon}{2}\right); \tag{5.2}$$

J_t is an *i.i.d.* Bernoulli random variable representing the jump times with $P\left(J_t = 1\right) = \kappa$, and the jumps $\mathcal{Z}_t^y$ are generated as $\mathcal{Z}_t^y \sim i.i.d.\mathcal{N}\left(\mu_y, \sigma_y^2\right)$. We generate ϵ_t according to $\epsilon_t \sim i.i.d.\mathcal{N}(0,1)$; this entails that, when (5.2) is used, the innovations $\sqrt{\lambda_t}\epsilon_t$ are *i.i.d.* with a Student's t distribution with υ degrees of freedom. We extend the specification in (5.1) by following a similar approach to Stroud and Johannes (2014); we model the log of the diffusive, non-jump variance h_t as

$$h_t = \left(\mu_h + \gamma s_t\right) + x_{1,t} + x_{2,t}, \tag{5.3}$$

where: $\mu_h + \gamma s_t$ is a mean level depending on a regime variable $s_t \in \{0,1\}$ following a Markov chain with transition probabilities $P\left(s_t = i | s_{t-1} = j\right) = p_{ij}$[13]; $x_{1,t}$ is a "slow" variance component generated as

$$x_{1,t+1} = \phi_1 x_{1,t} + \sigma_1 u_{1,t}, \tag{5.4}$$

with a "large" partial autocorrelation coefficient $0 < \phi_1 < 1$ and innovations $u_{1,t} \sim i.i.d.\mathcal{N}(0,1)$; and $x_{2,t}$ is a "fast" variance component generated as

$$x_{2,t+1} = \phi_2 x_{2,t} + \sigma_2\left(\rho\epsilon_t + \sqrt{1-\rho^2}u_{2,t}\right) + J_t\mathcal{Z}_t^2, \tag{5.5}$$

with a "small" partial autocorrelation coefficient $0 < \phi_2 < \phi_1$, innovations $u_{2,t} \sim i.i.d.\mathcal{N}(0,1)$, and jumps $\mathcal{Z}_t^2 \sim i.i.d.\mathcal{N}\left(\mu_2, \sigma_2^2\right)$. In (5.1), (5.4) and (5.5), $\left\{\epsilon_t, -\infty < t < \infty\right\}$, $\left\{\mathcal{Z}_t^y, -\infty < t < \infty\right\}$, $\left\{u_{1,t} - \infty < t < \infty\right\}$, $\left\{u_{2,t}, -\infty < t < \infty\right\}$ and $\left\{\mathcal{Z}_t^2, -\infty < t < \infty\right\}$ are five mutually independent groups; hence, ρ in (5.5) represent a measure of diffusive leverage, via the correlation between the innovations in the returns DGP ϵ_t and the fast variance component $x_{2,t}$.

The DGP in (5.1)–(5.5) thus considers a comprehensive specification designed to capture the key empirical features observed in financial market volatility, and in particular in the dynamics of the realised variance, and it builds – as well as on the references cited above – on several contributions in this literature including, *inter alia*, Harvey and Shephard (1996), Jacquier et al. (2004), Chib et al. (2006), McAleer and Medeiros (2008), and Koopman and Scharth (2012). Several models are nested within (5.1)–(5.5); in Table B.1 in the Supplement, we summarise five variants, each incorporating different empirical features. The baseline "SV" model assumes normal return innovations with two volatility factors; the "SVt" specification extends this by allowing for heavier tails; "SVL" considers the presence of a leverage effect between returns and volatility; "SVLJ" allows for jumps in both returns and volatility (as well as leverage effects); finally, "SVML" combines leverage effects with Markov-switching regimes. In Table B.2 in the Supplement, we report the values of the parameters used in our simulations, which largely follow the estimated values in Table 3 of Stroud and Johannes (2014), allowing our DGP to be more adherent to the features of real data.

In addition to the above five SV-based DGPs, we also consider a further DGP, based on a HAR-type model to account for correlations between state innovations in a SHARP model:

$$RV_t = \beta_{1,t} + \beta_{2,t}RV_{t-1} + \beta_{3,t}RV_{t-1}^{(w)} + \beta_{4,t}RV_{t-1}^{(m)} + \epsilon_t,$$

$$\beta_{j,t} = \phi_j\beta_{j,t-1} + \nu_{j,t}, \quad j = 1,2,3,4,$$

with state innovations $\nu_t = \left(\nu_{1,t}, \nu_{2,t}, \nu_{3,t}, \nu_{4,t}\right)' \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_\nu\right)$, $\boldsymbol{\phi} = (0.92, 0.92, 0.90, 0.90)'$ and

$$\mathbf{\Sigma}_\nu = \begin{bmatrix} 0.0400 & 0.0008 & -0.0010 & 0.0240 \\ 0.0008 & 0.0014 & -0.0010 & 0.0004 \\ -0.0010 & -0.0010 & 0.0050 & -0.0040 \\ 0.0240 & 0.0004 & -0.0040 & 0.0250 \end{bmatrix} \tag{5.6}$$

For each of the six DGPs mentioned above: we generate a synthetic realised volatility (RV) series of length $n = 1,440$, with the last 440 observations used for out-of-sample forecasting via a rolling window approach; we estimate our proposed models, SHARP and SHARP-SV, against "traditional" alternatives, including the HARL model, as well as more advanced benchmarks like HARSL and TVCHAR; and we compare the forecasting ability of the models mentioned above using standard loss functions (MSE, HMSE, HMAE, and QLIKE). Each experiment is based on 100 Monte Carlo simulations for the sake of computational efficiency.

[13] We introduce a regime-switching dynamics in (5.3) following Vo (2009); this can be viewed as an extension of the model considered in Stroud and Johannes (2014).

We report our main results in Figures B.1 and B.2 in the Supplement. The figures show that both the SHARP and SHARP-SV models consistently deliver the lowest loss function values across all DGPs, thereby demonstrating superior predictive accuracy. The only exception was the simplest SV-based DGP, where the HARLQ model performed marginally better on average but exhibited substantial outliers, making its performance less reliable across different simulations. Among the competing models, HARSL and TVCHAR showed more stable performance across DGPs; however, both the SHARP and SHARP-SV models outperformed them on average, with lower mean loss function values in almost all cases.

Further results are also in the Supplement. In Section B.2, along similar lines as in the previous Section 5.1, we assess the impact of the estimation technique on the performance of our proposed model — in essence, by estimating the SHARP model via MLE *cum* Kalman filtering (Table B.3). Results are decidedly worse, which reinforces the findings in Section 4: the superior forecasting ability associated with the SHARP model and with Particle Gibbs estimation holds also in a set of controlled scenarios using synthetic data. Further, in Table B.4 in Section B.3, we report computational times for the baseline SV DGP, showing that our Bayesian estimator remains feasible for large-scale applications.

6. Conclusion

This paper develops two new dynamic state-space models for realised volatility forecasting – SHARP and SHARP-SV – which extend the heterogeneous autoregressive framework of Corsi (2009) by allowing time-varying coefficients governed by autoregressive processes, with SHARP-SV further incorporating stochastic volatility in the coefficient dynamics. We follow Creal and Tsay (2015), using Gibbs particle filtering as a computationally efficient approach (Andrieu et al., 2010). We study the structure of our proposed models, characterising their dependence and deriving a set of asymptotic results which could be of independent interest. In a comprehensive set of empirical studies, we apply the SHARP and SHARP-SV models to the realised variance of the SPY index, sector ETFs, representative NYSE stocks, and the VIX index over long samples. Across all datasets, horizons, and loss functions (MSE, MAE, HMSE, HMAE, and QLIKE), SHARP and SHARP-SV on balance outperform HARL and its extensions, including HARSL (McAleer and Medeiros, 2008), HARLQ (Wang et al., 2016), TVCHAR (Bekierman and Manner, 2018), and SHARK (Buccheri and Corsi, 2021); these gains are confirmed by predictive ability tests (Giacomini and White, 2006) and Model Confidence Set analysis (Hansen et al., 2011). SHARP and SHARP-SV remain robust even in the presence of non-Gaussian innovations. Monte Carlo evidence further validates these conclusions: under correct specification, Particle Gibbs accurately recovers model parameters and latent states, outperforming Kalman filtering in terms of RMSE and interval calibration for long-horizon coefficients. Under misspecification, we show that SHARP and SHARP-SV maintain superior predictive accuracy – under various forms of misspecification – compared with other, competing models, with SHARP-SV showing particular strength.

These findings establish SHARP and SHARP-SV as versatile and effective tools for volatility forecasting, with direct implications for risk management, asset allocation, and derivative pricing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper was started (and virtually completed, although in a different form than the current one) when Mike was still alive; he proposed the new SHARP and SHARP-SV models, and developed the Bayesian, Particle Gibbs based, inferential procedure. We are grateful to the Editor, Michael Jansson, to the Guest Editor, Subal Kumbhakar, and to two anonymous Referees for very helpful and constructive feedback.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.106146.

References

Ait-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18 (2), 351–416.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39 (4), 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96 (453), 42–55.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (3), 269–342.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Estimating quadratic variation using realized variance. J. Appl. Econometrics 17 (5), 457–477.

Bekierman, J., Manner, H., 2018. Forecasting realized variance measures using time-varying coefficient models. Int. J. Forecast. 34 (2), 276–287.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 (3), 307–327.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. J. Econometrics 192 (1), 1–18.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2018. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. J. Econometrics 207 (1), 71–91.

Buccheri, G., Corsi, F., 2021. HARK the SHARK: Realized volatility modeling with measurement errors and nonlinear dependencies. J. Financ. Econ. 19 (4), 614–649.
Carriero, A., Clark, T.E., Marcellino, M., 2015. Bayesian VARs: Specification choices and forecast accuracy. J. Appl. Econometrics 30 (1), 46–73.
Carriero, A., Clark, T.E., Marcellino, M., 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. J. Econometrics 212 (1), 137–154.
Chen, X.B., Gao, J., Li, D., Silvapulle, P., 2018. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. J. Bus. Econom. Statist. 36 (1), 88–100.
Chib, S., Nardari, F., Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. J. Econometrics 134 (2), 341–371.
Clark, T.E., 2011. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. J. Bus. Econom. Statist. 29 (3), 327–341.
Clark, T.E., Ravazzolo, F., 2015. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. J. Appl. Econometrics 30 (4), 551–575.
Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. Rev. Econ. Dyn. 8 (2), 262–302.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7 (2), 174–196.
Creal, D., Tsay, R.S., 2015. High dimensional dynamic stochastic copula models. J. Econometrics 189 (2), 335–345.
D'Agostino, A., Gambetti, L., Giannone, D., 2013. Macroeconomic forecasting and structural change. J. Appl. Econometrics 28 (1), 82–101.
Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Rev. 3 (1), 1–100.
Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Econ. Stud. 61 (2), 247–264.
Harvey, A.C., Shephard, N., 1996. Estimation of an asymmetric stochastic volatility model for asset returns. J. Bus. Econom. Statist. 14 (4), 429–434.
Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. J. Econometrics 122 (1), 185–212.
Kokoszka, P., Mohammadi, N., Wang, H., Wang, S., 2025. Functional diffusion driven stochastic volatility model. Bernoulli 31 (2), 922–947.
Koop, G., Korobilis, D., 2013. Large time-varying parameter VARs. J. Econometrics 177 (2), 185–198.
Koop, G., Korobilis, D., Pettenuzzo, D., 2019. Bayesian compressed vector AutoRegressions. J. Econometrics 210 (1), 135–154.
Koop, G., Potter, S., 2004. Forecasting in dynamic factor models using Bayesian model averaging. Econom. J. 7 (2), 550–565.
Koopman, S.J., Scharth, M., 2012. The analysis of stochastic volatility in the presence of daily realized measures. J. Financ. Econ. 11 (1), 76–115.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. J. Econometrics 187 (1), 293–311.
McAleer, M., Medeiros, M.C., 2008. A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. J. Econometrics 147 (1), 104–119.
Sims, C.A., 1993. A nine-variable probabilistic macroeconomic forecasting model. In: Business Cycles, Indicators and Forecasting. University of Chicago Press, pp. 179–212.
Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. J. Bus. Econom. Statist. 14 (1), 11–30.
Stroud, J.R., Johannes, M.S., 2014. Bayesian modeling and forecasting of 24-hour high-frequency volatility. J. Amer. Statist. Assoc. 109 (508), 1368–1384.
Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-75. Time Ser. Anal.: Theory Pr. 1, 203–226.
Tsionas, M.G., Izzeldin, M., Trapani, L., 2022. Estimation of large dimensional time varying VARs using copulas. Eur. Econ. Rev. 141, 103952.
Vo, M.T., 2009. Regime-switching stochastic volatility: Evidence from the crude oil market. Energy Econ. 31 (5), 779–788.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: A dynamic model averaging approach. J. Bank. Financ. 64, 136–149.
Wang, X., Wu, C., Xu, W., 2015. Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. Int. J. Forecast. 31 (3), 609–619.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100 (472), 1394–1411.