Regularized Wishart Stochastic Volatility for High-Dimensional Covariance Forecasting

Roman Liesenfeld*
Department of Econometrics and Statistics, University of Cologne

Guilherme V. Moura
Department of Economics, Universidade Federal de Santa Catarina

Julian Spies
Department of Econometrics and Statistics, University of Cologne

February 20, 2026

Abstract

We extend Uhlig's (1994, 1997) Wishart stochastic volatility (WSV) model by introducing a regularized state transition for the precision matrix that shrinks the covariance forecasts toward a prior reference matrix. This regularization ensures stationarity of the return process and stabilizes the eigenvalues of the covariance forecasts, while preserving closed-form expressions for filtering, prediction, and likelihood evaluation. We provide conditions for stationarity and for the existence of second-order moments, show that the model admits a multivariate GARCH representation, and derive bounds that illustrate how regularization prevents degeneracy and excessive dispersion in the eigenvalues of the covariance matrix forecasts. To account for regime shifts in the correlation structure, we further embed a time-varying directional forgetting scheme into the regularized model, allowing the forgetting rate to differ over time and across directions in the return space. In a high-dimensional application with up to 1,000 assets, the regularized models deliver significantly more accurate covariance forecasts than the baseline WSV and perform well compared to several benchmark models.

Keywords: Directional forgetting; Multivariate GARCH; Multivariate stochastic volatility; Regularization; Shrinkage.

*The authors appreciate the comments of participants at the Workshop on Statistical Learning and High Dimensional Data at Unicamp (Brazil), the 2024 CFE-CMStatistics conference, and seminar participants at the University of Graz, University of Würzburg and University of Hohenheim. We thank the ITCC (IT Center University of Cologne) for providing compute resources on the DFG-funded HPC (High Performance Computing) system RAMSES (Research Accelerator for Modeling and Simulation with Enhanced Security) as well as support (DFG grant: INST 216/512-1 FUGG). G.V. Moura gratefully acknowledges support provided by CNPq (grant: 310327/2022-9)

1 Introduction

Modeling and forecasting the covariance matrix of asset returns is central to empirical finance. Portfolio allocation, risk management, and asset pricing, all rely on accurate estimates of current and future covariances. A large literature has developed time-series models for multivariate volatility, including multivariate GARCH (MGARCH) and multivariate stochastic volatility (MSV) models, as well as more recent approaches that combine dynamic structure with shrinkage or regularization to handle high-dimensional applications.

Early MGARCH models such as the BEKK specification (Engle and Kroner, 1995) provide flexible parameterizations of time-varying covariance matrices but suffer from severe parameter proliferation as the number of assets grows. This has motivated a range of more parsimonious models, most notably the dynamic conditional correlation (DCC) model (Engle, 2002). However, even such more parsimonious MGARCH models can become difficult to estimate and forecast as the dimension increases into the hundreds.

MSV models offer an alternative, placing latent volatility processes in a state-space framework. Popular approaches include factor MSV models Cholesky-based MSV models and matrix-variate MSV formulations (Asai et al., 2006). Since MSV models, unlike MGARCH models, generally do not have closed expressions for the likelihood and relevant moments for forecasting, statistical inference and forecasting are typically performed using simulated maximum likelihood (ML) or Bayesian Markov chain Monte Carlo (MCMC) methods (Asai et al., 2006). Although these methods can handle a wide variety of flexible dynamic structures in multivariate volatility processes, their computational complexity can be a bottleneck in high-dimensional applications and in situations where rapid sequential updating of forecasts is required.

In the literature on MSV models the Wishart Stochastic Volatility (WSV) approach proposed by Uhlig (1994, 1997) offers an attractive compromise between flexibility and

tractability. In this approach, returns are modeled as conditionally normal with a latent precision matrix that evolves according to a Markovian law of motion driven by singular matrix-variate Beta innovations. This construction ensures positive definiteness of both precision and covariance matrices, yields closed-form filtering and prediction through conjugacy between Normal, Wishart, and singular Beta distributions, and involves only a small number of parameters. These properties make the WSV appealing for high-dimensional covariance forecasting and for applications that require fast online updating. For a recent contribution that has applied Uhlig's WSV model to portfolio allocation and illustrated its computational advantages, see Moura et al. (2020).

However, the baseline WSV also has important limitations when applied to covariances of asset returns. First, the precision process implied by its transition equation is non-stationary, which leads to a non-stationary covariance process and ultimately to a non-stationary process for the returns. This contrasts with the widely held view that covariances of asset returns typically behave like stationary processes. For covariance forecasting, this non-stationarity implies that multi-step-ahead WSV forecasts do not return to a well-defined long-term covariance matrix and may therefore deteriorate as the forecast horizon increases when the actual process is mean reverting. Second, the WSV covariance forecasts involve a classical exponential forgetting (EF) scheme, in which past information is discounted at a constant rate over time and uniformly across all directions in the space of the return vector. This scheme may be too restrictive in that it makes it difficult to adapt quickly to the regime changes that are often observed in the correlation structure of asset returns (Fan et al., 2024).

The challenge of modeling time-varying covariance matrices becomes even more acute in high-dimensional settings, where the number of assets is large relative to the available time-series sample. In such environments, shrinkage and regularization play a central role. A large literature has developed shrinkage estimators of static (stationary) covariance ma-

trices, such as linear and nonlinear shrinkage toward simple targets (Ledoit and Wolf, 2004, 2012, 2015; De Nard, 2020), and factor-based covariance estimators (De Nard et al., 2021). These methods are motivated by the poor finite-sample properties of the sample covariance matrix in high dimension, and they achieve substantial gains by systematically shrinking covariance estimators, usually via their eigenvalues, toward a priori reasonable structures. More recently, such ideas have been integrated into dynamic frameworks, with models that align the temporal evolution of covariance matrices with target matrices that are sparsely parameterized or estimated with shrinkage estimators (Engle et al., 2019).

Our contribution is to bring these ideas together by embedding shrinkage and directional regularization into the WSV framework. We extend Uhlig's (1994, 1997) baseline WSV model in two dimensions. First, we introduce a regularized WSV (RWSV) model that modifies the law of motion for the covariance matrix by shrinking it toward a reference matrix. The regularization is implemented directly in the state transition of the precision matrix in a way that preserves the conjugate structure of the model: filtering, prediction, and likelihood evaluation remain available in closed form. At the same time, the regularization induces a stationary return process and stabilizes the eigenvalues of covariance matrix forecasts. Furthermore, the implied return process admits a restricted MGARCH representation with covariance targeting, linking our model to the MGARCH literature.

Second, we augment the regularized RWSV with a directional forgetting (DF) mechanism motivated by recursive least squares (RLS) methods (Goel et al., 2020). Unlike scalar EF, DF adjusts the effective discount rate across directions in the return space and over time. Information from previous returns is discounted more aggressively when the current return vector is well aligned with previous return vectors, and less aggressively when it points in a direction that has been less travelled by previous returns. This selective forgetting is particularly relevant in environments with regime changes or structural breaks in the correlation structure, where the dominant eigenvectors of the covariance matrix can

shift rapidly, as considered, for example, in Fan et al. (2024). By combining DF with the regularized EF scheme in the RWSV, we obtain models that retain the stabilizing effect of shrinkage toward a reference covariance matrix, whilst enabling a quicker adaptation to changes in the dependence structure.

Our contributions relative to the existing literature can be summarized as follows. (i) We propose a regularized WSV model that restores stationarity of the return process, provides explicit conditions for the existence of second moments for the returns, and preserves the closed-form filtering and predictive features of the original WSV specification. (ii) We show that the RWSV admits an MGARCH representation with covariance targeting, which facilitates interpretation of parameter roles and connects our approach to the broader GARCH literature. (iii) We introduce directional forgetting into the RWSV framework that allow the effective forgetting rate to vary across directions and over time, thereby improving adaptability to structural changes in the covariance matrix. (iv) We demonstrate through a large-scale portfolio application with up to 1,000 assets that the regularized WSV specifications deliver more accurate covariance and density forecasts than the baseline WSV and compare favorably to a range of benchmark models.

Our contribution is most closely related to the works of Moura et al. (2020), Windle and Carvalho (2014) and Philipov and Glickman (2006). However, Moura et al. (2020) focuses on the application of the baseline WSV model for portfolio selection, while Windle and Carvalho (2014) extends the baseline model for use in modeling realized covariance matrices in a Bayesian context. Finally, Philipov and Glickman (2006) propose a latent autoregressive Wishart model that is similar to the WSV model. However, it is neither sparsely parameterized nor does it provide a closed form for the likelihood, making it very difficult to implement in high-dimensional applications.

The remainder of the paper is organized as follows. Section 2 introduces notation. Section 3 presents the regularized RWSV and the DF extensions. Section 4 reports the

results of Monte Carlo experiments investigating their forecast performance. Section 5 presents an empirical application based on large cross-sections of U.S. stocks. Section 6 concludes. Proofs and implementation details are deferred to the online Appendix.

2 Notation

We denote by $\mathcal{N}_p(\mu, \Sigma)$ the p-dimensional Normal distribution with mean μ and covariance matrix Σ, and by $\mathcal{W}_p(\upsilon, \Omega)$ the p-dimensional Wishart distribution with υ degrees of freedom (dof) and scale matrix Ω, so that its mean is $\upsilon\Omega$. $\mathcal{IW}_p(\upsilon, \Upsilon)$ denotes the p-dimensional inverse Wishart distribution with υ dof and scale matrix Υ, so that its mean is $\Upsilon/(\upsilon - p - 1)$ when $\upsilon > p + 1$, and $t_p(\mu, \Sigma, \upsilon)$ denotes the p-dimensional t-distribution with υ dof, mean μ (for $\upsilon > 1$) and covariance matrix Σ (for $\upsilon > 2$). $\mathcal{B}_p(\alpha, \beta)$ is the p-dimensional matrix-variate beta distribution with dof α and β. $y \perp z$ is used to indicate stochastic independence of y and z. $\mathcal{F}_t = \sigma(y_s, 0 < s \leq t)$ is the σ-field generated by the observations for the variates y_s up to time period t. For matrices A and B, we write $A > B$ ($A \geq B$) if $A - B$ is positive definite (positive semi-definite). $A = 0$ and $A = \infty$ indicate that A has zero and infinite eigenvalues, respectively. The matrix $A^{1/2}$ denotes the upper-triangular Cholesky factor of A, and I_p is the p-dimensional identity matrix.

3 Wishart stochastic volatility

3.1 Background

We consider a multivariate volatility modelling approach for a vector of m assets returns $r_t = (r_{1,t}, ..., r_{m,t})'$, $t = 1, \ldots, T$, designed to be applicable when m is large. It builds on

Uhlig's (1994, 1997) WSV model, given by the following non-linear state-space system:

$$r_t|X_t \sim \mathcal{N}_m(0, X_t^{-1}), \tag{1}$$

$$X_{t+1} = \frac{1}{\lambda} X_t^{1/2'} \Psi_{t+1} X_t^{1/2}, \qquad \Psi_t \sim \mathcal{B}_m(n/2, 1/2), \qquad \lambda \in (0,1), \tag{2}$$

where λ is a scalar parameter. The innovation Ψ_t, which drives the latent precision matrix X_t, is a serially independent $m \times m$ random matrix with $\Psi_t \perp X_{t-\ell}$, $\ell = 1, 2, \ldots$. It follows a singular matrix-variate Beta distribution with dof $n/2$ and $1/2$ and $n > m-1$, such that $\mathrm{E}(\Psi_t) = [n/(n+1)]I_m$. The initial condition for X_t is

$$X_1 \sim \mathcal{W}_m(n, \Sigma_{1|0}^{-1}), \tag{3}$$

with a fixed positive definite scale matrix $\Sigma_{1|0}^{-1}$. The parameters n and λ jointly determine the variation and dynamics of X_t: The distribution of Ψ_{t+1} becomes more concentrated around I_m as n increases, thereby reducing the conditional variation in X_{t+1} given X_t. The conditional mean of X_t is $\mathrm{E}(X_{t+1}|X_t) = X_t n/[\lambda(n+1)]$, demonstrating that linear serial dependence in X_t increases with decreasing λ.

A key feature of the WSV is that conjugacy between the multivariate Normal, Wishart and singular matrix-variate Beta distributions yields closed-form formulas for filtering and prediction (Uhlig, 1994; Windle and Carvalho, 2014). The resulting covariance forecasts are weighted moving averages with the classical EF scheme, in which older return observations are down-weighted with a discount factor given by the parameter λ (Windle and Carvalho, 2014). Such weighted moving averages based on EF are known to produce robust forecasts and are used in various contexts, including recursive least squares (RLS) for updating covariance matrices to track time-varying parameters in linear models (Ljung and Söderström, 1983; Kulhavỳ and Zarrop, 1993). The combination of these characteristics and its sparse parameterisation makes the WSV an attractive option for modeling and forecasting high-dimensional covariance matrices.

However, the WSV also has important limitations. The precision process defined by the transition equation (2) and hence the implied covariance process are not stationary, which implies a non-stationary return process (Windle and Carvalho, 2014). This non-stationarity is reflected, among other things, in the classical EF scheme for covariance forecasts which cannot prevent the forecasts from degenerating towards a matrix with zero eigenvalues (see Section 3.2.5 below). As a consequence, multi-step-ahead covariance forecasts do not revert toward a positive definite long-run mean and are expected to deteriorate as the forecast horizon increases if the true covariance process is mean reverting.

A further limitation arises from the classical EF scheme in the covariance forecasts. This scheme discounts old information from previous return observations at a rate λ uniformly over time and across all directions in the space of the return vector, regardless of how much information has accumulated along the different directions. However, regime shifts in the correlation structure – usually associated with transitions between different stock market regimes – typically generate excitation patterns in the return vector that are uneven over time and across directions and exhibit a concentration in certain directions over certain periods of time (see, e.g., Fan et al., 2024). The ability to learn the directions associated with a new regime quickly depends crucially on forgetting information about the directions associated with the previous regime. For covariance forecasting, this suggests that a forgetting scheme that discounts more heavily along previously most-excited directions and less along previously least-excited directions may allow forecasts to adapt more quickly to changing correlation structures than classical EF.

These considerations motivate our proposed generalizations of the baseline WSV model, which (i) regularize the transition equation for the precision process (2) by targeting it towards a suitable prior reference matrix, and (ii) generalize the forgetting scheme to vary over time and across directions in the data space.

3.2 Regularized Wishart stochastic volatility (RWSV)

3.2.1 Model specification

We regularize the law of motion for the precision X_t of the WSV by replacing the transition equation (2) with

$$X_{t+1} = \Delta_t X_t^{1/2'} \Psi_{t+1} X_t^{1/2} \Delta_t', \tag{4}$$

$$\Delta_t = \left(\lambda \Sigma_{t|t} + \kappa \Sigma_*\right)^{-1/2} \Sigma_{t|t}^{1/2}, \quad \Sigma_{t|t} = (n-m)\mathrm{E}(X_t^{-1}|\mathcal{F}_t), \quad \lambda \in (0,1),\ \kappa \geq 0. \tag{5}$$

Here Σ_* is a fixed positive definite prior reference matrix and κ is a scalar regularization parameter. $\mathrm{E}(X_t^{-1}|\mathcal{F}_t)$ is the period-t filtered covariance matrix of the returns, defined implicitly by the model. Its explicit form is given below in Lemma 1. The extension of the baseline WSV defined by Equations (1), (3)-(5) will be referred to as the regularized WSV (RWSV) model. Its parameters to be estimated are $\theta = (n, \lambda, \kappa)'$. A natural and convenient choice for the scale matrix in the initial condition (3) is $\Sigma_{1|0} = \Sigma_*$, which we adopt in all our applications. Choices for Σ_* are discussed in Section 3.2.4 below.

To see how the RWSV relates to the baseline WSV, consider the case $\kappa = 0$, so that $\Delta_t = I_m/\sqrt{\lambda}$. Then the transition equation (4) reduces to that of the WSV in Equation (2) with an implied transition for the covariance matrix given by $X_{t+1}^{-1} = \lambda X_t^{-1/2} \Psi_{t+1}^{-1} X_t^{-1/2'}$. Under the RWSV the covariance transition becomes

$$X_{t+1}^{-1} = \Delta_t^{-1'} X_t^{-1/2} \Psi_{t+1}^{-1} X_t^{-1/2'} \Delta_t^{-1}, \quad \Delta_t^{-1} = \Sigma_{t|t}^{-1/2} \left(\lambda \Sigma_{t|t} + \kappa \Sigma_*\right)^{1/2}, \tag{6}$$

assuming that Δ_t is invertible (which holds if Σ_* and $\Sigma_{1|0}$ are positive definite; see proof of Lemma 1 in online Appendix A.1). Since the autoregressive matrix $\Delta_t^{-1} < I_m$ when $(1-\lambda)\Sigma_{t|t} > \kappa \Sigma_*$ and vice versa, the RWSV effectively shrinks the next-period covariance matrix X_{t+1}^{-1} towards the scaled reference matrix Σ_*. When the filtered period-t covariance matrix $\mathrm{E}(X_t^{-1}|\mathcal{F}_t)$ associated with $\Sigma_{t|t}$ is 'large' relative to Σ_*, the process is pulled back

and when it is 'small', the process is pushed away from zero toward Σ_*. This regularization stabilizes the covariance dynamics while preserving the basic WSV structure.

3.2.2 Filtering, prediction, ML estimation and GARCH representation

The property of Uhlig's baseline WSV that filtering and forecasting can be performed in closed form also extends to the RWSV. The next lemma summarizes the filtering and predictive distributions for the precision matrix under the RWSV, obtained using the properties of the singular Beta for the innovation matrix Ψ_t established by Uhlig (1994) and the fact that Δ_t is $\mathcal{F}_t$-measurable and full rank.

Lemma 1. *Under the RWSV model defined by Equations (1), (3)-(5), the period-t filtering distribution of X_t for $t = 1, 2, \ldots$ is*

$$X_t|\mathcal{F}_t \sim \mathcal{W}_m(n+1, \Sigma_{t|t}^{-1}), \qquad \Sigma_{t|t} = \Sigma_{t|t-1} + r_t r_t', \tag{7}$$

and the one-step-ahead predictive distribution is

$$X_{t+1}|\mathcal{F}_t \sim \mathcal{W}_m(n, \Sigma_{t+1|t}^{-1}), \qquad \Sigma_{t+1|t} = \lambda\Sigma_{t|t} + \kappa\Sigma_*. \tag{8}$$

For the baseline WSV with $\kappa = 0$, the filtering and predictive distributions have the same functional form as in Lemma 1, with Σ_* removed from the scale matrix of the predictive distribution, so that $\Sigma_{t+1|t} = \lambda\Sigma_{t|t}$ (Windle and Carvalho, 2014). Thus, the RWSV preserves the WSV's analytical tractability while adding regularization toward Σ_* .

Lemma 1 implies that the filtering and predictive distributions of the covariance matrix X_t^{-1} are inverse Wishart distributions and that its one-step-ahead covariance forecast is

$$\mathrm{E}(X_{t+1}^{-1}|\mathcal{F}_t) = \frac{1}{n-m-1}(\lambda\Sigma_{t|t} + \kappa\Sigma_*). \tag{9}$$

Thus, the forecast is a convex combination of the prior Σ_* and $\Sigma_{t|t}$, incorporating the information from the return data up to period t. For $\kappa = 1 - \lambda$, it corresponds to the

covariance forecast under the regularized EF scheme proposed by Kulhavỳ and Zarrop (1993) for RLS applications. Under this constraint, the weights λ and $\kappa = 1 - \lambda$ in the forecast (9) can be viewed, in a Bayesian sense, as the probabilities associated with the prior and data information.

From the (closed-form) predictive Wishart distribution for X_t in Lemma 1, it follows that the likelihood function for an ML analysis of the RWSV also has a simple closed form. Specifically, combining the Wishart distribution for X_t in Equation (8) with the conditional normal distribution for r_t in Equation (1) yields for r_t given $\mathcal{F}_{t-1}$ the following multivariate t-distribution with zero mean, covariance matrix Σ_t and dof ν:

$$r_t|\mathcal{F}_{t-1} \sim t_m(0, \Sigma_t, \nu), \qquad \Sigma_t = \frac{1}{\nu - 2}\Sigma_{t|t-1}, \qquad \nu = n - m + 1. \tag{10}$$

The likelihood function is therefore $L(\theta) = \prod_{t=1}^{T} f(r_t|\mathcal{F}_{t-1})$, where $f(r_t|\mathcal{F}_{t-1})$ is the density function of the t-distribution in Equation (10) and $\Sigma_{t|t-1}$ is calculated using the recursion defined in equations (7) and (8), i.e. $\Sigma_{t|t-1} = \kappa\Sigma_* + \lambda r_{t-1}r'_{t-1} + \lambda\Sigma_{t-1|t-2}$.

Furthermore, the conditional return distribution in Equation (10) with the recursion for $\Sigma_{t|t-1}$ immediately results in the notable property of the latent non-linear state-space model defined by the RWSV that it can be represented as the following restricted scalar diagonal MGARCH(1,1) process with covariance targeting and t-distributed innovations:

$$r_t = \Sigma_t^{1/2\prime}\eta_t, \quad \Sigma_t = \frac{\kappa}{\nu - 2}\Sigma_* + \frac{\lambda}{\nu - 2}r_{t-1}r'_{t-1} + \lambda\Sigma_{t-1}, \quad \eta_t \sim t_m(0, I_m, \nu), \tag{11}$$

The target for Σ_t is the stationary covariance matrix $\mathrm{E}(r_t r'_t)$ implied by the prior reference matrix Σ_*, which is given by $\mathrm{E}(r_t r'_t) = \kappa\Sigma_*/[(\nu - 2) - \lambda(\nu - 1)]$ (see Proposition 1 below). This GARCH representation facilitates the interpretation of the role of the RWSV parameters, for example, the fact that $\lambda(\nu - 2)^{-1}$ measures the effect of the GARCH innovations and that the sum of the GARCH coefficients $\lambda(\nu - 2)^{-1} + \lambda$ measures the persistence in the conditional covariance process.

3.2.3 Conditions for stationarity

We next establish conditions for stationarity and the existence of the second-order moments of the returns r_t. We assume that r_t is serially uncorrelated and have a mean of zero. Thus, covariance stationarity only requires that the second moments of r_t are finite, or equivalently, that the mean of X_t^{-1} is finite. The following result provides conditions for this finiteness.

Proposition 1. *Under the RWSV model defined by Equations (1), (3)-(5) with* $\nu = n - m + 1$ *and* $\lambda > 0$ *the expectation of* X_t^{-1} *and the second moment of* r_t *are finite iff* $\nu > 2$ *and* $\lambda < (\nu - 2)/(\nu - 1)$. *In that case*

$$E(r_t r_t') = E(X_t^{-1}) = \frac{\kappa}{(\nu - 2) - \lambda(\nu - 1)} \Sigma_*. \tag{12}$$

According to this proposition, the existence condition for $\mathrm{E}(r_t r_t')$ requires a smaller λ for a decreasing ν. This reflects the fact that the variation of the covariance X_t^{-1} of the conditional return distribution (1) increases as the dof $\nu = n - m + 1$ decreases (see Section 3.1) and as λ increases (see Equation 6), causing the tails of the unconditional return distribution to become fatter. Furthermore, Equation (12) shows that stationarity also depends on the regularizing parameter κ. When $\kappa \to 0$ (recovering the baseline WSV), $\mathrm{E}(r_t r_t')$ tends to a zero matrix, reflecting the non-stationarity of the WSV.

3.2.4 Choice of the prior reference covariance matrix

To implement the RWSV model, we must select a reference matrix Σ_* in the transition equation (5), which we then propose to use also as the scale $\Sigma_{1|0}$ of the initial condition (3). As the RWSV shrinks the covariance predictions towards Σ_* and $\Sigma_{1|0}$ according to Equation (9), it is advisable to select them based on reasonable a-priori assumptions while maintaining scalability to high-dimensional settings.

We specify Σ_* such that the implied stationary covariance matrix $E(r_t r_t')$ is diagonal with zero off-diagonal elements. Denote this diagonal covariance matrix by S. Using Equation (12), this leads to the specific choice

$$\Sigma_* = \Sigma_{1|0} = \frac{(\nu-2) - \lambda(\nu-1)}{\kappa} S. \tag{13}$$

For the diagonal elements of S, we use the usual sample variances of each of the return series computed over the in-sample estimation period. For the baseline WSV, which has no well-defined stationary covariance matrix, we set $\Sigma_{1|0} = (\nu-2)S$. The scaling by $(\nu-2)$ ensures that the predictive distribution of X_t^{-1} under the WSV implies $E(r_1 r_1'|\mathcal{F}_0) = S$.

These choices for Σ_* and $\Sigma_{1|0}$, which are based on the prior assumption of zero return covariances, are consistent with the specifications of Uhlig (1997) and Moura et al. (2020) for $\Sigma_{1|0}$ in the baseline WSV. They are also in the spirit of the shrinkage approach of Ledoit and Wolf (2004), which aims to reduce estimation error in high-dimensional covariance matrices, by shrinking the sample covariance matrix toward a diagonal target. In Ledoit and Wolf (2004), this target is a scalar multiple of the identity matrix, whereas our choice allows for heterogeneous diagonal elements, and hence different variances across assets[1].

3.2.5 Forecasts

To assess the effect of regularization in the RWSV on covariance forecasts, we compare the RWSV forecasts and their bounds to those of the baseline WSV. Combining Equation (9) with the recursion for $\Sigma_{t+1|t}$ from Lemma 1 and setting $\Sigma_{1|0} = \Sigma_*$, the one-step-ahead RWSV covariance forecast based on the T in-sample return observations takes the form

$$\mathrm{E}(X_{T+1}^{-1}|\mathcal{F}_T) = \frac{1}{n-m-1}\left[\frac{(1-\lambda-\kappa)\lambda^T + \kappa}{1-\lambda}\Sigma_{1|0} + \sum_{\ell=0}^{T-1}\lambda^{\ell+1} r_{t-\ell} r_{t-\ell}'\right]. \tag{14}$$

[1]We also explored alternative selections for Σ_* inspired by other shrinkage approaches, however, these did not deliver substantial forecast improvements over the simple diagonal specification in Equation (13)

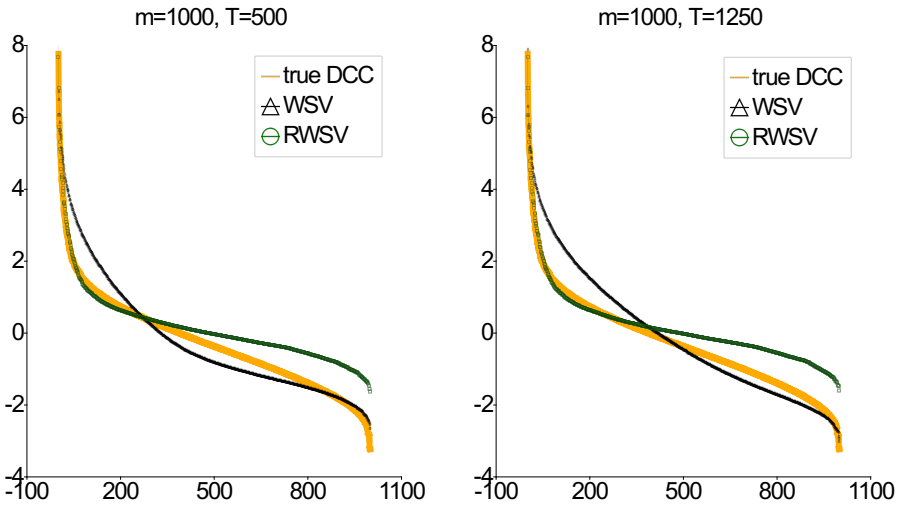

Figure 1: Log eigenvalues of out-of-sample conditional covariance matrix forecasts from the baseline WSV and RWSV and the true eigenvalues obtained from the Monte-Carlo experiment described in Section 4. The data-generating process is a Gaussian DCC model and the true eigenvalues are those of its simulated conditional covariance matrix for a randomly selected time period in the out-of-sample window. Eigenvalues are sorted from largest to smallest and then plotted against their rank.

For the baseline WSV with $\kappa = 0$, the corresponding forecast is

$$\mathrm{E}(X_{T+1}^{-1}|\mathcal{F}_T) = \frac{1}{n-m-1}\left[\lambda^T \Sigma_{1|0} + \sum_{\ell=0}^{T-1} \lambda^{\ell+1} r_{t-\ell} r_{t-\ell}'\right]. \tag{15}$$

In both cases, forecasts are exponentially weighted averages of current and past returns with a shrinkage towards the initial (prior) matrix $\Sigma_{1|0}$. The difference is that the shrinkage component in the WSV vanishes as T grows, whereas it remains bounded away from zero in the RWSV, at a level determined by κ. Furthermore, for given (n, λ, κ) and T, the regularization implies a stronger shrinkage towards $\Sigma_{1|0}$ under the RWSV.

As a special case of the (R)WSV forecasts in Equations (14) and (15), letting $\lambda \to 1$ and eliminating shrinkage towards $\Sigma_{1|0}$ recovers the standard sample covariance matrix. As documented by Ledoit and Wolf (2004, 2012), forecasts based on the sample covariance

matrix in high dimension with small T perform poorly, typically exhibiting upward bias in the largest eigenvalues and downward bias in the smallest ones. This indicates that too much information accumulates in the most excited directions of r_t and too much information is ignored in their least excited ones. By shrinking towards $\Sigma_{1|0}$, both the WSV and RWSV can narrow the eigenvalue spread in the forecasts of the covariance matrix, and thus can be expected to improve forecast accuracy, similarly to the shrinkage strategies for the sample covariance matrix in Ledoit and Wolf (2004, 2012).

Assuming that returns are bounded so that $0 \leq r_t r_t' \leq C < \infty$, the lower and upper bounds of the RWSV forecasts are

$$\left[\ \frac{1}{n-m-1}\frac{(1-\lambda-\kappa)\lambda^T+\kappa}{1-\lambda}\Sigma_{1|0} \ , \ \frac{1}{n-m-1}\left(\frac{(1-\lambda-\kappa)\lambda^T+\kappa}{1-\lambda}\Sigma_{1|0}+\frac{\lambda-\lambda^T}{1-\lambda}C\right) \ \right], \tag{16}$$

and those for the WSV forecasts

$$\left[\ \frac{1}{n-m-1}\lambda^T\Sigma_{1|0} \ , \ \frac{1}{n-m-1}\left(\lambda^T\Sigma_{1|0}+\frac{\lambda-\lambda^T}{1-\lambda}C\right) \ \right]. \tag{17}$$

These bounds determine the range of eigenvalues of the covariance matrix forecasts and show how λ, n, and κ control the trade-off between the risk of excessively small and excessively large eigenvalues: Increasing λ and decreasing n in the WSV and RWSV reduces the risk that the smallest eigenvalues become too small, but raises the risk that the largest eigenvalues become too large. In the RWSV, increasing κ ceteris paribus moves the lower bound further away from a matrix with zero eigenvalues, without substantially affecting the upper bound (since in the most excited directions return data typically bring substantially more information than contained in a reasonably selected prior $\Sigma_{1|0}$). It is to be expected that this additional regularization margin in the RWSV allows a better trade-off between biases in the smallest and largest eigenvalues than is possible under the baseline WSV.

Figure 1 illustrates these effects using results of the Monte Carlo experiment from Section 4 below, where the true conditional covariance matrix is generated by a Gaussian

DCC model of Engle (2002). The figure plots the log of the sorted eigenvalues of out-of-sample covariance matrix forecasts from the WSV and RWSV with $\kappa = 1 - \lambda$, together with the eigenvalues of the true conditional covariance matrix for $m = 1000$ assets and in-sample estimation window lengths $T \in \{500, 1250\}$. The eigenvalues are shown for a randomly selected time period. However, their key characteristics, which can be seen in Figure 1 for this time period, are identical for all time periods in the out-of-sample window. The WSV forecasts exhibit a pronounced upward bias for the large eigenvalues (> 1) and a moderate downward bias for the small ones (< 1), while the RWSV predicts the largest eigenvalues much more accurately and shows only a moderate upward bias for the smallest ones. Thus, the regularization in the RWSV helps to capture the variation along the dominant directions of the return data considerably better than the WSV.

3.3 Wishart stochastic volatility with directional forgetting (DF)

In the covariance forecasts $\mathrm{E}(X_{t+1}^{-1}|\mathcal{F}_t)$ of the RWSV and WSV (Equation 9), the parameter λ discounts past information uniformly – both over time and across all directions in the data space – regardless of how much information has already accumulated along each direction.

An alternative discounting scheme used in RLS approaches for covariance updates when new information arrives unevenly across directions is directional forgetting (DF). DF is a selective forgetting scheme: old information is forgotten only when it can be replaced by incoming data. In our context, such a DF scheme can be expected to allow covariance forecasts to adapt more quickly to sudden shifts in the correlation structure than classical or regularized EF. To assess its value for forecasting return covariances, we combine Wishart stochastic volatility with the DF mechanism of Kulhavỳ and Zarrop (1993). For other DF variants, see Goel et al. (2020). We first replace the classical EF in the baseline WSV by the DF scheme, obtaining the WSV-DF model, and then combine the DF with regularized EF in the RWSV, obtaining the RWSV-DF.

3.3.1 WSV-DF model

The WSV-DF is obtained by replacing the autoregressive matrix Δ_t of the RWSV transition equation (5) with

$$\Delta_t = \Big(\Sigma_{t|t} - (1-\lambda)\delta_t r_t r_t'\Big)^{-1/2}\Sigma_{t|t}^{1/2}, \quad \delta_t = (r_t'\Sigma_{t|t}^{-1}r_t)^{-1} - (r_t'\Sigma_{1|0}^{-1}r_t)^{-1}, \quad \lambda \in (0,1). \quad (18)$$

The full WSV-DF model is defined by Equations (1), (3), (4), and (18), and involves the parameters $\theta = (n, \lambda)$. Its filtering and predictive distributions for the precision X_t are summarized in Lemma 2. Their derivation is analogous to that for the RWSV distributions in Lemma 1, since the WSV-DF and RWSV differ only in the $\mathcal{F}_t$-measurable matrix Δ_t.

Lemma 2. *Under the WSV-DF model defined by Equations (1), (3), (4), and (18), the period-t filtering distribution of X_t for $t = 1, 2, \ldots$ has the same form as that for the RWSV model given in Equation (7), and the one-step-ahead predictive distribution is*

$$X_{t+1}|\mathcal{F}_t \sim \mathcal{W}_m(n, \Sigma_{t+1|t}^{-1}), \qquad \Sigma_{t+1|t} = \Sigma_{t|t} - (1-\lambda)\delta_t r_t r_t'. \quad (19)$$

From Equation (19) it follows that the one-step-ahead forecast of the covariance is

$$\mathrm{E}(X_{t+1}^{-1}|\mathcal{F}_t) = \frac{1}{n-m-1}\Big(\Sigma_{t|t} - (1-\lambda)\delta_t r_t r_t'\Big). \quad (20)$$

Thus, the DF scheme allows discounting of the accumulated past information in $\Sigma_{t|t}$ only in the currently excited direction of the return vector r_t. The effective period-t discount factor is $(1-\lambda)\delta_t$, where δ_t is defined in Equation (18). The latter measures how well the current direction of r_t aligns with the directions that have been most strongly excited by the returns observed up to period t, relative to its alignment with the 'prior directions' represented by $\Sigma_{1|0}$. Better alignments of r_t with the directions that are most excited up to t yields larger δ_t and hence stronger discounting, and vice versa.

The recursion resulting from Lemma 2 for $\Sigma_{t+1|t}$ gives the following representation of the covariance forecast:

$$\mathrm{E}(X_{T+1}^{-1}|\mathcal{F}_T) = \frac{1}{n-m-1}\left[\Sigma_{1|0} + \sum_{\ell=0}^{T-1} \omega_\ell r_{T-\ell} r'_{T-\ell}\right], \tag{21}$$

$$\omega_\ell = \lambda - (1-\lambda)\Big\{(r'_{T-\ell}\Sigma^{-1}_{T-\ell|T-\ell-1} r_{T-\ell})^{-1} - (r'_{T-\ell}\Sigma^{-1}_{1|0} r_{T-\ell})^{-1}\Big\}. \tag{22}$$

Thus, the lower bound for the WSV-DF forecasts is $\Sigma_{1|0}/(n-m-1)$, which is greater than that of the WSV and equal to that of the RWSV with restriction $\kappa = (1-\lambda)$ (see Equations 16 and 17). A finite upper bound would result if the sum of the weights $\sum_{\ell=0}^{T-1} \omega_\ell$ is finite, which, however, is generally not guaranteed for any T (Kulhavỳ and Zarrop, 1993). Hence, although the WSV-DF can be expected to adapt quickly to changes in the correlation structure, it may accumulate too much information for the covariance forecasts.

3.3.2 RWSV-DF model

To prevent the DF mechanism in the WSV from accumulating too much information, we follow Kulhavỳ and Zarrop (1993) and combine the DF with the regularized EF in the RWSV. The resulting RWSV-DF model expands the autoregressive matrix Δ_t of the WSV-DF in Equation (18) as follows:

$$\Delta_t = \Big(\lambda\Sigma_{t|t} + (1-\lambda)\Big[\tau\{\Sigma_{t|t} - \delta_t r_t r'_t\} + (1-\tau)\Sigma_*\Big]\Big)^{-1/2}\Sigma_{t|t}^{1/2}, \quad \tau \in [0,1], \tag{23}$$

with $\lambda \in (0,1)$, where δ_t is defined as in the WSV-DF (Equation 18) and Σ_* is the prior matrix as used in the RWSV. The parameter τ controls the strength of regularized EF relative to DF. For $\tau = 1$, the RWSV-DF becomes the WSV-DF, and for $\tau = 0$, it corresponds to the RWSV with the constraint $\kappa = 1-\lambda$.

The filtering and predictive distributions of the precision X_t under the RWSV-DF are analogous to those for the RWSV and WSV-DF and are given in the following lemma.

Lemma 3. *Under the RWSV-DF model defined by Equations (1), (3), (4), and (23), the period-t filtering distribution of X_t for $t = 1, 2, \ldots$ has the same form as that for the RWSV model given in Equation (7), and the one-step-ahead predictive distribution is*

$$X_{t+1}|\mathcal{F}_t \sim \mathcal{W}_m(n, \Sigma_{t+1|t}^{-1}), \quad \Sigma_{t+1|t} = \lambda\Sigma_{t|t} + (1-\lambda)\Big[\tau\{\Sigma_{t|t} - \delta_t r_t r_t'\} + (1-\tau)\Sigma_*\Big]. \quad (24)$$

For $\Sigma_* = \Sigma_{1|0}$, the recursion resulting from Lemma 3 for $\Sigma_{t+1|t}$ implies that the covariance forecast can be expressed as

$$\mathrm{E}(X_{T+1}^{-1}|\mathcal{F}_T) = \frac{1}{n-m-1}\left[\Sigma_{1|0} + \sum_{\ell=0}^{T-1}\omega_\ell r_{T-\ell}r_{T-\ell}'\right], \quad (25)$$

$$\omega_\ell = \Big[\lambda + \tau(1-\lambda)\Big]^{\ell+1}\Big[\lambda - \tau(1-\lambda)\Big\{(r_{T-\ell}'\Sigma_{T-\ell|T-\ell-1}^{-1}r_{T-\ell})^{-1} - (r_{T-\ell}'\Sigma_{1|0}^{-1}r_{T-\ell})^{-1}\Big\}\Big].$$

Thus, forecasts are bounded below as with the WSV-DF. If $\lambda+\tau(1-\lambda) < 1$ (or equivalently $\tau < 1$), then $\sum_{\ell=0}^{T-1}\omega_\ell < \infty$. This means that, in contrast to the WSV-DF, the RWSV-DF forecasts have a finite upper bound, the value of which depends on the actual observed data (Kulhavỳ and Zarrop, 1993). Thus, the RWSV-DF forecasts inherit the stabilizing effect of regularized EF controlling the eigenvalue spread while retaining the ability of DF to adjust more quickly to changes in dominant directions.

As with the RWSV, the predictive return distribution under the RWSV-DF (and the WSV-DF) is a t-distribution (see Equation 10) and leads to a scalar diagonal GARCH representation (see Equation 11). Under the RWSV-DF, the specific form of conditional covariance matrix Σ_t in Equations (10) is

$$\Sigma_t = \frac{(1-\lambda)(1-\tau)}{\nu-2}\Sigma_* + \frac{1}{\nu-2}[\lambda + (1-\lambda)\tau(1-\delta_{t-1})]r_{t-1}r_{t-1}' + [\lambda + (1-\lambda)\tau]\Sigma_{t-1}. \quad (26)$$

Therefore, in the RWSV-DF-implied MGARCH process, the innovations have a time-varying effect depending on their direction via δ_t[2].

[2]The fact that δ_t in the recursion for the scale matrix $\Sigma_{t+1|t}$ (Equation 24) is a non-linear function

For the implementation of the (R)WSV-DF, we select the reference matrix Σ_* and the initial condition $\Sigma_{1|0}$ according to the same principles that are used for the (R)WSV and described in Section 3.2.4. However, for the RWSV-DF, we do not have a closed-form formula for the stationary covariance matrix of returns (assuming it exists) to calibrate Σ_* and $\Sigma_{1|0}$ to a diagonal stationary covariance matrix with zero covariances, as in the RWSV. We therefore use an approximation to the stationary covariance matrix, assuming that $(\delta_t - 1)r_t r_t' = \{(r_t'\Sigma_{t|t-1}^{-1} r_t)^{-1} - (r_t'\Sigma_{1|0}^{-1} r_t)^{-1}\} r_t r_t'$ is on average close to zero, so that $\mathrm{E}[(\delta_t - 1)r_t r_t'] \simeq 0$. This approximation can be justified by the fact that if $\Sigma_{t|t-1}$ is a stable process, it fluctuates around $\Sigma_{1|0} = \Sigma_*$. As shown in online Appendix A.5, this then yields,

$$\Sigma_* = \Sigma_{1|0} = \frac{(\nu-2)(1-\tau) - \lambda(\nu-1) + \tau(\nu-2)\lambda}{(1-\lambda)(1-\tau)} S. \tag{27}$$

For $\tau = 0$, for which the RWSV-DF reduces to the RWSV, this selection matches that for the RWSV in Equation (13). For the WSV-DF, we set $\Sigma_{1|0} = (\nu-2)S$, as in the WSV.

4 Monte Carlo simulations

In this section, we examine the out-of-sample forecasting performance of the WSV models in a setting where the true covariance is known. To this end, we conduct Monte-Carlo (MC) experiments based on two data-generating processes (DGPs). The first DGP is a DCC model (Engle, 2002). This DGP represents a scenario in which the correlation structure evolves relatively slowly and smoothly over time. To study settings with more abrupt changes in the correlation structure we consider a second DGP that features structural breaks in the eigenvectors of the covariance matrix. This design is intended to assess,

of past returns means that analysing the stationarity of the RWSV-DF, as was done in Section 3.2.3 for the RWSV, is mathematically challenging. We leave this to future research, in which we can draw on the results of Meitz and Saikkonen (2008) on the stability of univariate GARCH models with nonlinear conditional GARCH equations, for example.

in particular, how well WSV models can adapt to information about the new directions following sudden changes in the covariance matrix. The design follows Fan et al. (2024) and is given by $r_t \sim \mathcal{N}(0, \Sigma_t)$, where $\Sigma_t \in \{\Sigma_{(0)}, \Sigma_{(1)}\}$. Σ_t switches between $\Sigma_{(0)}$ and $\Sigma_{(1)}$ every 50 periods, whereby the eigenvector matrices of $\Sigma_{(0)}$ and $\Sigma_{(1)}$ are orthogonal to each other. Additional details of the two DPGs are given in online Appendix B.

4.1 Performance measures

We use three performance measures. The first is the Frobenius distance of the one-step-ahead forecast of the conditional covariance matrix, $\text{FROB}_{t+1} = \text{vec}(\Sigma^f_{t+1} - \Sigma_{t+1})'\text{vec}(\Sigma^f_{t+1} - \Sigma_{t+1})$, where Σ^f_{t+1} is the model's forecast of the true covariance matrix Σ_{t+1}. The second measure is the QLIKE loss, $\text{QLIKE}_{t+1} = \text{tr}\left[\left(\Sigma^f_{t+1}\right)^{-1}\Sigma_{t+1}\right] + \ln|\Sigma^f_{t+1}|$. Formal justifications for these performance measures can be found in Patton and Sheppard (2009). The third measure is the log predictive density (LPD), which assesses the overall out-of-sample fit of the model's return density forecasts (Geweke and Amisano, 2010). It is computed as the log conditional density of r_{t+1} given $\mathcal{F}_t$ assumed by the model, evaluated at its covariance forecast Σ^f_{t+1} and the parameter estimates $\hat{\theta}$, i.e. $\text{LPD}_{t+1} = \ln f(r_{t+1}|\Sigma^f_{t+1}, \hat{\theta})$.

4.2 MC results

For both DGPs, we simulate return series to obtain 3000 out-of-sample covariance forecasts. We consider an investment universe with $m = 1000$ assets and three estimation window lengths, $T = \{250, 500, 1250\}$, corresponding to roughly one, two, and five years of daily data. These choices imply concentration ratios $m/T = \{0.8, 2, 4\}$. In the DCC experiment, forecasts are obtained by re-estimating the model parameters every 21 periods using a rolling window of length T, in line with the common practice of treating 21 consecutive days as a virtual month. In the structural break experiment, the parameters are estimated

		WSV	RWSV	RWSV -κ	WSV -DF	RWSV -DF	DCC
$(m,T)=(1000,250)$	FROB	1942.3	561.7	**553.5**	22116.7	563.8	619.2
	QLIKE	1419.3	1103.7	**1101.6**	2361.3	1103.6	2009.4
	LPD	-1546.5	**-1469.7**	-1469.7	-1546.1	-1469.7	–
$(m,T)=(1000,500)$	FROB	1947.6	593.9	589.9	26183.9	595.6	**400.4**
	QLIKE	1490.9	1098.6	**1097.4**	2631.9	1098.6	2177.0
	LPD	-1588.4	**-1466.9**	-1466.9	-1593.9	-1466.9	–
$(m,T)=(1000,1250)$	FROB	1833.4	616.0	614.4	33778.8	617.9	**232.9**
	QLIKE	1565.2	1094.8	**1094.3**	3122.4	1094.8	1966.0
	LPD	-1644.8	**-1465.0**	-1465.0	-1688.1	-1465.0	–

Table 1: Results of the MC experiment using the DCC as DGP. Reported are the time average of the out-of-sample losses based on Frobenius distance (FROB), QLIKE, and log predictive density (LPD). Bold numbers indicate the smallest average loss (FROB, QLIKE) across all models and the largest average LPD value across the WSV models.

		WSV	RWSV	RWSV -κ	WSV -DF	RWSV -DF
$(m,T)=(1000,250)$	FROB	1955.5	2006.9	2006.5	**1390.2**	2138.3
	QLIKE	1937.7	1874.2	**1871.5**	2221.4	1876.9
	LPD	-1871.0	-1852.8	-1852.5	-1582.7	**-1849.9**
$(m,T)=(1000,500)$	FROB	1864.4	1908.4	1894.0	**1237.7**	2146.0
	QLIKE	1912.9	1847.2	**1840.3**	2248.2	1843.6
	LPD	-1859.9	-1839.4	-1838.1	-1839.4	**-1836.2**
$(m,T)=(1000,1250)$	FROB	1620.9	1699.5	1634.7	**1028.1**	2125.7
	QLIKE	1867.0	1814.8	**1799.0**	2317.8	1805.1
	LPD	-1841.6	-1823.1	**-1818.1**	-1824.6	-1820.3

Table 2: Results of the MC experiment with structural breaks in eigenvectors as DGP. Reported are the time average of the out-of-sample losses based on Frobenius distance (FROB), QLIKE, and log predictive density (LPD). Bold numbers indicate the smallest average loss (FROB, QLIKE) across all models and the largest average LPD value across the WSV models.

once and then kept fixed over the entire out-of-sample period.

Table 1 reports time averages of the FROB, QLIKE, and LPD for the WSV models in the DCC experiment. The models are: WSV (baseline), RWSV (regularized with constraint $\kappa = 1 - \lambda$), RWSV-κ (regularized with unconstrained κ), WSV-DF (DF without

regularization), and RWSV-DF (DF with regularization). For comparison, the table also reports the corresponding results for the DCC. Its forecasts are obtained by estimating the model parameters using nonlinear shrinkage estimation for correlation targeting as suggested by Engle et al. (2019). Consistent with the benefits of shrinkage in high dimensions, the regularized WSV models (RWSV, RWSV-κ, RWSV-DF) significantly outperform the baseline WSV across all scenarios. Notably, for the short estimation window ($T = 250$), the RWSV models outperform even the true DCC model in terms of FROB and QLIKE losses. This result highlights that, in high-dimensional settings with limited data history, the estimation uncertainty in the true parameter-rich DCC model outweighs the bias caused by the sparse shrinkage target in the RWSV models. Conversely, the WSV-DF performs significantly worse. This shows that, in contrast to the regularization of the WSV, replacing the classical EF in the WSV with the DF scheme does not improve the forecasts. This can be explained by the fact that the WSV-DF accumulates too much information while its ability to account for abrupt changes in the correlation structure is irrelevant in the DCC experiment involving slowly moving correlations. The inadequacy of the DF scheme in this scenario is also evident in the fitted RWSV-DF, where the rolling-window estimates for τ (not reported here) are close to zero. This makes the fitted RWSV-DF virtually identical to the fitted RWSV, providing essentially the same forecast performance. The RWSV-κ is the best WSV specification. However, the differences in its performance compared to the constrained RWSV are small and show no consistent pattern, indicating that, in the absence of structural breaks, the additional flexibility introduced by an unconstrained shrinkage parameter or by a DF scheme is not picked up by the data.

Table 2 reports the MC results for the DGP that features structural breaks in the covariance matrix. In this setting with abruptly shifting eigenvectors, there is no fixed long-run covariance target. We see that the WSV-DF has the lowest average FROB values of all models. This shows that the DF scheme enhances the ability to learn the new directions

after regime changes compared to the classical and regularized EF. However, this comes at the cost of excessive information accumulation, even in unchanged directions, as indicated by the substantially larger QLIKE values for the WSV-DF compared to the other models. These large QLIKE values for the WSV-DF are primarily due to the comparatively large eigenvalues of the covariance forecasts, which then lead to very large log determinants of Σ_{t+1}^{f} in the QLIKE loss. The RWSV-DF significantly reduces this instability, achieving a QLIKE loss comparable to that of the RWSV-κ, which performs best in terms of QLIKE. It also exhibits superior LPD values for the two shorter estimation windows.This confirms that combining DF with regularized EF successfully balances adaptability with stability.

In summary, the MC results highlight that the relative merits of covariance models depend on the stability of the underlying dependence structure. If the correlations change only gradually over time and there is a stable long-run covariance, as in the DCC experiment, global shrinkage through the parsimonious RWSV model is sufficient. In contrast, when the covariance structure is subject to abrupt regime shifts and eigenvalue changes, as in the second DGP, additional local flexibility becomes essential. These findings underscore that RWSV-DF is a robust model that can excel even in environments characterized by structural instability at the cost of only one additional parameter.

5 Empirical application

In this section, we investigate the out-of-sample properties of the proposed WSV models in an empirical application based on historical stock returns, and compare their performance to that of several benchmark models.

We use daily stock price data from the Center for Research in Security Prices (CRSP) for NYSE, AMEX, and NASDAQ stocks from January 1, 1985, to December 31, 2024. We generate one-day-ahead forecasts of the return covariance matrix Σ_{t+1}, re-estimating

model parameters every 21 trading days. We consider two investment universes with $m \in \{500, 1000\}$ stocks, and three rolling estimation windows with $T \in \{250, 500, 1250\}$ daily return observations. The out-of-sample window begins on January 3, 1990 and ends on December 31, 2024, yielding 8817 forecasts.

Before each re-estimation of the parameters, the m stocks in the investment universe are re-selected following Engle et al. (2019). Specifically, at each re-estimation date we choose the stocks with the largest market capitalization that have a complete return history over the past T trading days and the next 21 trading days. In addition, for any pair of stocks with a sample correlation exceeding 0.95, we exclude the one with lower trading volume.

As in Section 4, we evaluate out-of-sample performance using the FROB, QLIKE, and LPD measures. Since the true covariance matrix Σ_{t+1} cannot be observed, we use $r_{t+1}r'_{t+1}$ as a proxy for Σ_{t+1} when calculating the FROB and the QLIKE (see Section 4.1). In addition to these statistical measures, we consider an economically motivated criterion. This is based on the evaluation of covariance forecasts Σ^f_{t+1} for predicting the weights of the global minimum variance portfolio (GMVP). The GMVP weights are defined as those minimizing the variance of the portfolio returns. For the true covariance matrix, they are given by $w_{t+1} = \Sigma^{-1}_{t+1}\iota_m/(\iota'_m\Sigma^{-1}_{t+1}\iota_m)$. Their forecasts, w^f_{t+1}, are obtained by replacing Σ_{t+1} by Σ^f_{t+1}, and the average GMVP loss is the sample mean of $(r'_{t+1}w^f_{t+1})^2$. Formal justifications for the GMVP loss can be found in Patton and Sheppard (2009).

We compare the WSV-based models with a set of benchmark approaches for covariance forecasting: (i) The sample covariance matrix computed from the T most recent return observations (SCV), i.e. the standard estimator for the stationary covariance matrix Σ. (ii) The sample covariance matrix as in (i), but setting all the covariances equal to zero (SCV-d). (iii) The linear shrinkage estimator (SHR-l) of Ledoit and Wolf (2004) for Σ. (iv) The nonlinear shrinkage estimator (SHR-nl) of Ledoit and Wolf (2012, 2015). (v) The linear constant-variance-covariance shrinkage estimator (SHR-c) of De Nard (2020)

		WSV	RWSV	RWSV -κ	WSV -DF	RWSV -DF	DCC	SCV	SCV -d	SHR -l	SHR -nl	SHR -c
$(m,T)=(500,250)$	FROB	2563.1	2497.8	2497.8	2815.7	**2496.8**	2523.6	–	2535.4	2600.9	2609.9	2604.7
	QLIKE	967.4	926.6	924.9	1027.5	**922.6**	938.1	–	1184.3	1892.1	985.0	1340.2
	GMVP	**0.289**	0.290	0.290	0.318	0.298	0.326	–	0.956	0.422	0.363	0.398
	LPD	-903.4	-894.6	-894.3	-889.6	**-885.8**	–	–	–	–	–	–
$(m,T)=(1000,250)$	FROB	6070.0	**5960.9**	5961.9	6226.4	5967.2	6020.1	–	6041.7	6158.0	6172.7	6164.8
	QLIKE	2164.9	2076.1	2073.3	2200.6	**2072.0**	2102.5	–	2536.0	4794.6	2245.3	3162.4
	GMVP	0.228	0.239	0.238	0.246	0.241	**0.225**	–	1.013	0.327	0.312	0.323
	LPD	-1914.5	-1898.2	-1897.9	-1889.3	**-1884.8**	–	–	–	–	–	–
$(m,T)=(500,500)$	FROB	2559.2	2501.8	2501.3	3064.6	**2500.0**	2510.6	–	2534.5	2614.2	2620.5	2617.1
	QLIKE	967.2	932.1	928.2	1083.1	928.9	**920.4**	–	1206.4	1598.1	978.6	1252.9
	GMVP	0.316	0.298	**0.296**	0.422	0.316	0.336	–	0.982	0.465	0.360	0.410
	LPD	-905.6	-895.5	-894.8	-894.4	**-884.2**	–	–	–	–	–	–
$(m,T)=(1000,500)$	FROB	6039.4	**5961.0**	5964.7	6378.2	5970.1	5991.4	–	6040.7	6181.6	6192.7	6186.6
	QLIKE	2167.1	2091.0	2086.1	2275.1	2085.0	**2064.4**	–	2575.9	5270.1	2234.3	3297.0
	GMVP	0.235	0.240	0.237	0.286	0.247	**0.224**	–	1.042	0.344	0.296	0.323
	LPD	-1920.6	-1899.9	-1899.7	-1891.5	**-1881.5**	–	–	–	–	–	–
$(m,T)=(500,1250)$	FROB	2570.4	2509.5	2508.4	3604.3	**2502.9**	2503.2	2626.7	2533.6	2620.7	2625.1	2622.7
	QLIKE	977.3	952.5	946.9	1197.3	954.9	**918.4**	1248.0	1252.6	1120.0	983.9	1062.3
	GMVP	0.337	0.307	**0.305**	0.786	0.334	0.318	0.460	1.011	0.408	0.369	0.388
	LPD	-913.3	-901.7	-900.6	-925.6	**-887.4**	–	–	–	–	–	–
$(m,T)=(1000,1250)$	FROB	6036.2	**5963.2**	5963.6	6945.9	5976.2	5975.5	6201.7	6040.6	6190.2	6197.8	6193.1
	QLIKE	2198.4	2139.1	2133.5	2458.9	2136.6	**2063.4**	6910.1	2661.0	3261.8	2230.3	2683.2
	GMVP	0.246	0.243	0.241	0.550	0.258	**0.224**	0.782	1.075	0.360	0.289	0.316
	LPD	-1940.7	-1912.4	-1912.1	-1927.8	**-1888.2**	–	–	–	–	–	–

Table 3: Results of the empirical application. Reported are the time average of the out-of-sample losses based on Frobenius distance (FROB), QLIKE and GMVP loss, and log predictive density (LPD) for portfolio sizes $m \in \{500, 1000\}$. Parameter estimation is based on a sample of length T. Bold numbers indicate the smallest average loss (FROB, QLIKE, GMVP) across all models and the largest average LPD value all models. Grey cells indicate that the model belongs to the 90% MCS.

for Σ.(vi) The Gaussian DCC model (described in online Appendix B) based on nonlinear shrinkage estimation for correlation targeting as suggested by Engle et al. (2019).

5.1 Out-of-sample performance

Table 3 reports the out-of-sample time averages of the FROB, QLIKE, and GMVP loss for the WSV models and the benchmark models, as well as the time average of the LPD for the WSV models. To assess the statistical significance of differences in average losses across the competing models, we use the model confidence set (MCS) approach of Hansen et al. (2011). The MCS is constructed to contain, with a given confidence level (here 90%), the set of models that are not significantly outperformed by any other model. We implement the MCS using a block bootstrap with a bootstrap sample size of 10000 and the data-dependent choice for the block length recommended by Hansen et al. (2011).

The results in Table 3 can be summarized as follows: The three regularized WSV models – RWSV, RWSV-κ, and RWSV-DF – consistently outperform the baseline WSV and the WSV-DF under the statistical loss functions. The baseline WSV model performs relatively well only for the GMVP loss and achieves a lower loss than the other WSV models for the shortest estimation window. However, even for this loss, the regularized WSV models are in the 90% MCS for most m/T ratios. Among the WSV models, the RWSV-DF exihibits the best overall performance. It delivers the highest LPD values and mostly falls inside the 90% MCS for the other loss measures. However, the performance differences between the RWSV-DF, RWSV, and RWSV-κ are typically small. The three regularized WSV models also compare favorably to the benchmark approaches. Among the benchmarks, the DCC performs best in all dimensions. The DCC only significantly outperforms the regularized WSV models in terms of the QLIKE for $(m, T) = (1000, 500)$ and $T = 1250$. This is consistent with the MC results under the DCC DGP, where the highly parameterized DCC benefits more from larger sample sizes than the more parsimonious regularized WSV

models. Overall, the empirical results demonstrate that, especially for short estimation windows, the regularized WSV specifications perform comparatively well.

5.2 Out-of-sample performance over time

To compare the covariance forecasts of each of the WSV models with the DCC forecasts over time, we use the fluctuation test for unstable environments of Giacomini and Rossi (2010). We implement the Giacomini-Rossi (GR) test for the FROB and QLIKE losses, using the respective average out-of-sample loss within a rolling window of 882 observations (10% of the full out-of-sample period). Figure 2 plots the time series of the GR test statistic for the FROB loss for each WSV model and all m/T ratios, together with the 5% asymptotic critical value for the Null hypothesis that the performance of the respective WSV model is no better than that of the DCC at each point in time. A positive value of the test statistic indicates that a model's forecast is better than the DCC forecast, and the Null hypothesis is rejected if the statistic exceeds the critical value at least once. Figure 3 shows the corresponding GR test statistics for the QLIKE loss.

For the FROB loss, the GR statistics indicate that the three regularized WSV models frequently outperform the DCC, particularly after the dot-com bubble period (2000–2002) and especially in the stress episodes associated with the European sovereign debt crisis (2011), the oil price shock (2014–2016), the energy crisis associated with the Ukraine war (2021–2022), and the bank failure crisis (2023), suggesting that these important periods contribute to the models' good overall performance under the FROB loss. However, their relative performance deteriorates during the global financial crisis (2007–2008) and at the onset of the COVID-19 pandemic (2020). Notably, the performance deterioration of the RWSV-DF in the COVID-19 episode is less pronounced than that for the RWSV(-κ), suggesting that the DF component makes the regularized EF more robust to sudden shocks in extreme market conditions. The baseline WSV model shows substantial losses in forecast

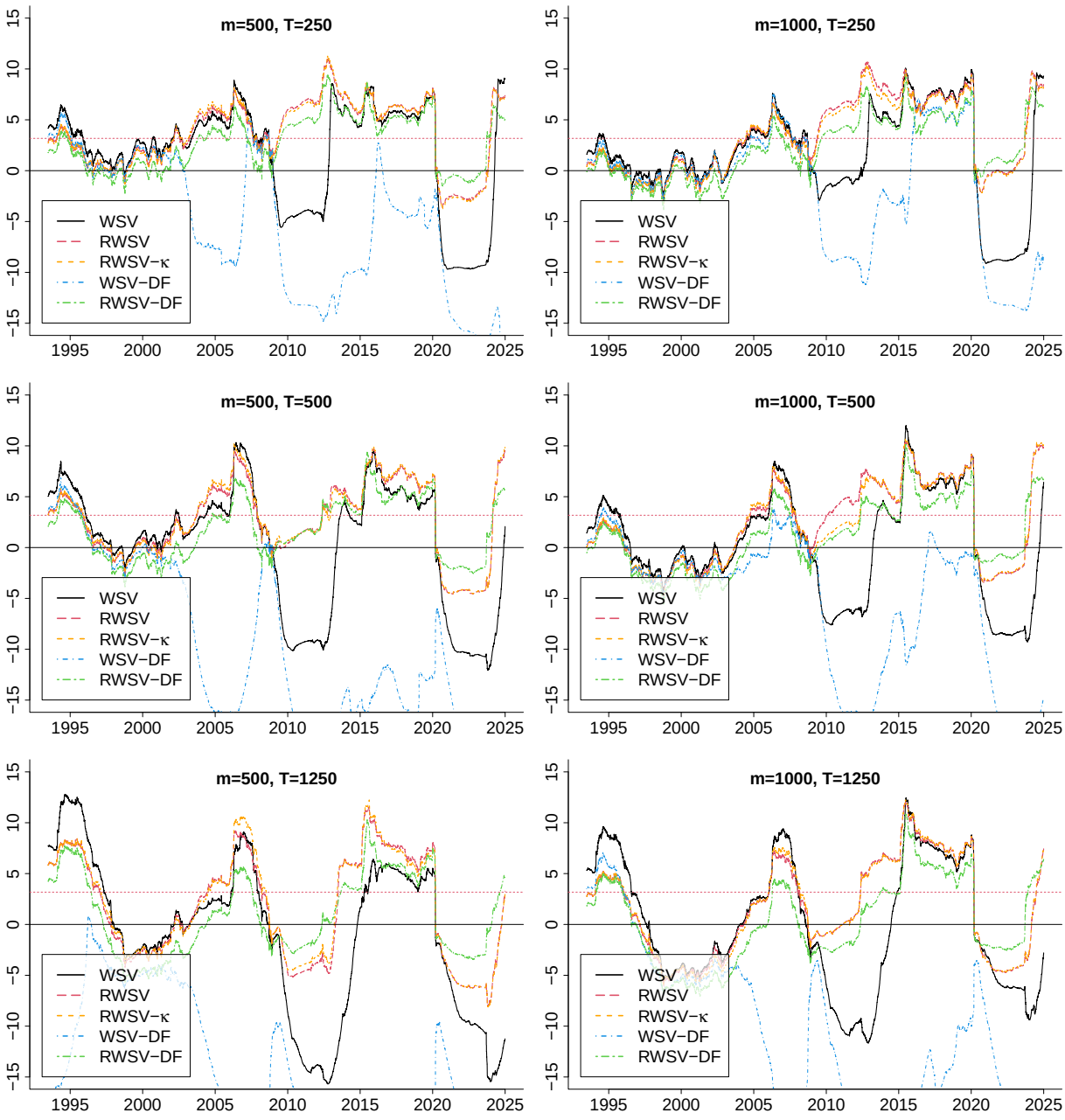

Figure 2: GR fluctuation test statistics based on the Frobenius distance for the hypotheses that the WSV, RWSV, RWSV-κ, WSV-DF and RWSV-DF, each in pairwise comparison, are equal to or worse than the DCC. Red dotted line marks the 5% critical value.

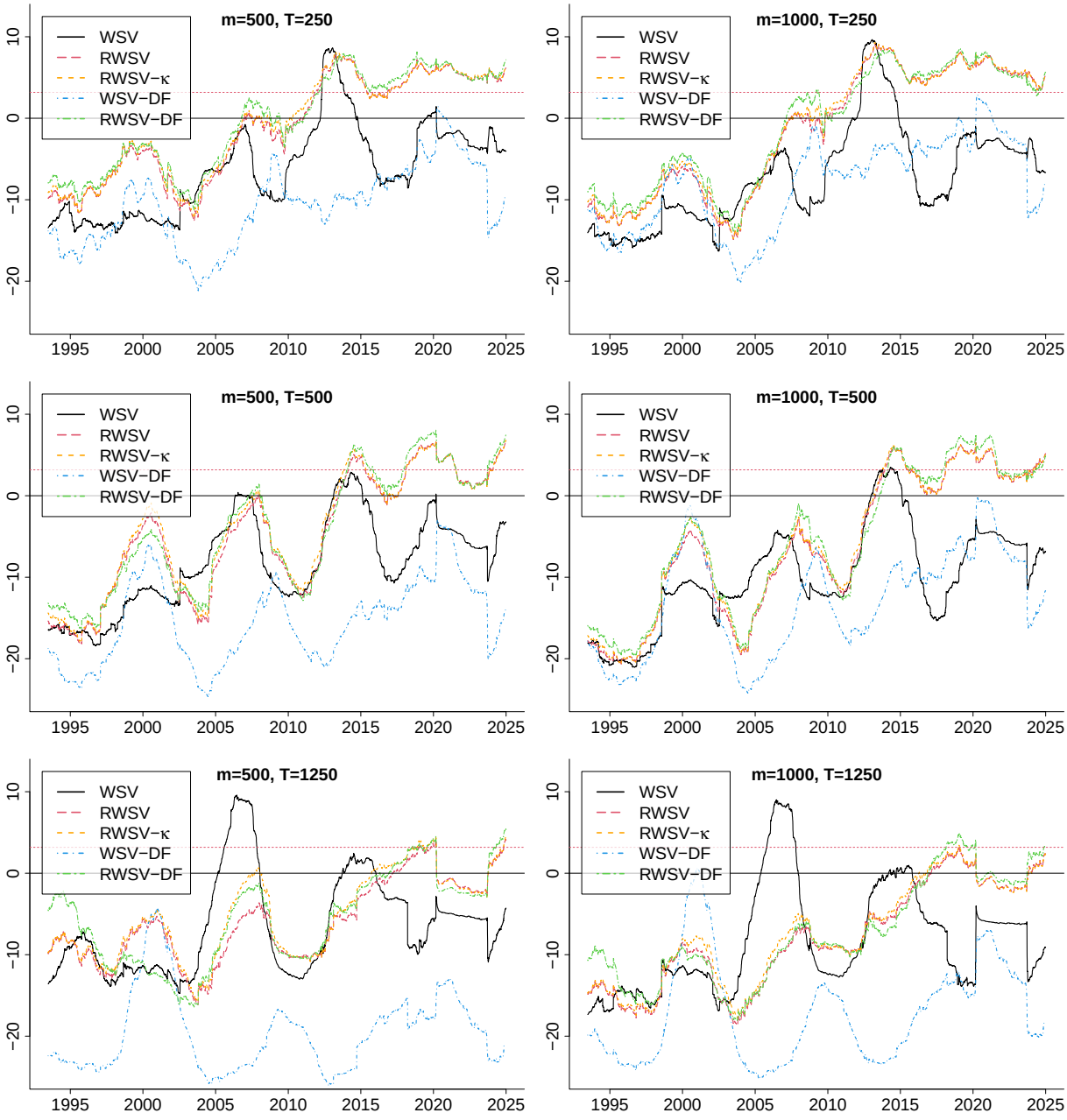

Figure 3: GR fluctuation test statistics based on the QLIKE loss for the hypotheses that the WSV, RWSV, RWSV-κ, WSV-DF and RWSV-DF, each in pairwise comparison, are equal to or worse than the DCC. Red dotted line marks the 5% critical value.

quality compared to the DCC after the global financial crisis and the COVID-19 pandemic, and the WSV-DF generally shows performance losses throughout the entire out-of-sample period, especially for the two longer estimation windows. For the QLIKE loss, the GR statistics show that the good overall performance of the regularized WSV models for shorter estimation windows is driven mainly by gains achieved from the European sovereign debt crisis in 2011 onward up to the end of the out-of-sample period in 2025. For the longest estimation window there is no extended period in which any of the regularized WSV models significantly outperform the DCC. This pattern aligns with the intuition that the DCC benefits more from longer estimation windows, whereas the regularized WSV models are particularly competitive when samples are relatively short and parameter uncertainty is more pronounced.

6 Conclusion

In this article, we have proposed a regularized Wishart stochastic volatility model (RWSV) for forecasting high-dimensional covariance matrices of asset returns, which extends Uhlig's (1994, 1997) Wishart stochastic volatility model (WSV). The RWSV model modifies the law of motion for the latent precision matrix of the WSV, so that the covariance matrix is shrunk towards a prior reference matrix. This regularizes the exponential forgetting (EF) of past returns in the WSV's covariance forecasts and enables stationarity of the return process. We have derived the implied filtering and predictive distributions for the latent covariance process and shown that the model preserves the closed form of these distributions, as well as that of the likelihood, and that it can be represented as a restricted scalar diagonal multivariate GARCH model with covariance targeting.

To improve adaptation to regime changes in the correlation structure, we have further combined the WSV and RWSV model with a time-varying directional forgetting scheme,

resulting in the WSV-DF and RWSV-DF models, respectively. In this directional forgetting, information from previous returns is only discounted in the direction of the current return vector, and the better the alignment between its direction and the directions of previous returns, the more aggressive the discounting.

Our Monte Carlo experiments, based on a DCC-GARCH process and a design with structural breaks in eigenvectors of the covariance matrix, show that the regularized WSV specifications substantially improve covariance and density forecasts relative to the baseline WSV. In the DCC-GARCH experiment, the regularized WSV models achieve lower Frobenius and QLIKE losses and higher log predictive densities across a wide range of dimensions of the asset universe m and sample sizes T, and even outperform the true model when T is small. In the structural break experiment, the WSV-DF achieves significant reductions in Frobenius loss compared to the baseline WSV, but exhibits a comparatively high QLIKE loss because it accumulates too much information. However, this high QLIKE loss is significantly reduced by the RWSV-DF. These results confirm that regularization is effective in stabilizing eigenvalues of the covariance forecasts and reducing forecast errors, while directional forgetting enhances adaptability to shifts in the dependence structure.

In our empirical application with up to $m = 1000$ assets, the regularized WSV models outperform the baseline WSV and compares favorably to several widely used benchmark models, including the DCC. The regularized WSV models deliver lower Frobenius losses and robust performance across different m/T ratios and loss functions. They mostly belong to the model confidence set, with their relative strength being most evident when T is small. These findings highlight the practical relevance of combining the Wishart stochastic volatility with regularization and directional forgetting for large-scale portfolio problems.

Several extensions are left for future research. Firstly, a rigorous analysis of the stationarity and moment conditions for the RWSV-DF would be of interest; a formal treatment could build on recent results for nonlinear GARCH models. Second, it would be of interest

to incorporate additional economic structure, such as factor or network restrictions, into the prior reference matrix and to explore alternative shrinkage targets motivated by economic theory. Finally, while we have focused on the covariance matrix of financial asset returns, the proposed models can be readily applied to other contexts in which high-dimensional covariance forecasts are essential, such as macro-financial vector autoregressive models.

Disclosure Statement

The authors report there are no competing interests to declare.

Data Availability Statement

The data that support the findings of this study were obtained from the Center for Research in Security Prices (CRSP) via Wharton Research Data Services (WRDS). Restrictions apply to the availability of these data, which were used under license for this study. The data are available from CRSP through WRDS (https://wrds.wharton.upenn.edu).

References

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: A review. *Econometric Reviews 25*(2-3), 145–175.

De Nard, G. (2020, 11). Oops! I Shrunk the Sample Covariance Matrix Again: Blockbuster Meets Shrinkage. *Journal of Financial Econometrics 20*(4), 569–611.

De Nard, G., O. Ledoit, and M. Wolf (2021, 01). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics 19*(2), 236–257.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric Theory 11*(1), 122–150.

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*, 363–375.

Fan, J., Y. Li, N. Xia, and X. Zheng (2024). Tests for principal eigenvalues and eigenvectors. Working paper, available at `arxiv.org/abs/2405.06939`.

Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230. Special Issue: Bayesian Forecasting in Economics.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*, 595–620.

Goel, A., A. L. Bruce, and D. S. Bernstein (2020). Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency [lecture notes]. *IEEE Control Systems Magazine 40*(4), 80–102.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*, 453–497.

Kulhavỳ, R. and M. B. Zarrop (1993). On a general concept of forgetting. *International Journal of Control 58*, 905–924.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Ledoit, O. and M. Wolf (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis 139*, 360–384.

Ljung, L. and T. Söderström (1983). *Theory and Practice of Recursive Identification*. MIT Press.

Meitz, M. and P. Saikkonen (2008). Stability of nonlinear ar-garch models. *Journal of Time Series Analysis 29*(3), 453–475.

Moura, G. V., A. A. Santos, and E. Ruiz (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking & Finance 118*, 105882.

Patton, A. J. and K. Sheppard (2009). *Evaluating Volatility and Correlation Forecasts*, pp. 801–838. Berlin, Heidelberg: Springer Berlin Heidelberg.

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via wishart processes. *Journal of Business & Economic Statistics 24*(3), 313–328.

Uhlig, H. (1994). On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, 395–405.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica 65*, 59–74.

Windle, J. and C. M. Carvalho (2014). A tractable state-space model for symmetric positive-definite matrices. *Bayesian Analysis 9*(4), 759–792.