Modeling and Forecasting of Realized Covariance Matrices of Asset Returns using State-Space Models

Inauguraldissertation zur Erlangung des Doktorgrades der

Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln

vorgelegt von

Jan Patrick Hartkopf

aus

Solingen

Köln, 2020

Referent:	Prof. Dr. Roman Liesenfeld
Korreferent:	Prof. Dr. Jörg Breitung
Tag der Promotion:	29. April 2021

Für meinen Bruder

Contents

List of Tables

List of Figures

Chapter 1.

Introduction

This thesis is concerned with the modeling and forecasting of realized covariances of asset returns in high dimensions. Conditional covariances are key ingredients in many financial applications as, e.g., optimal portfolio allocation, hedging, the assessment of systematic risk and option pricing. Due to the relevance of accurate forecasts for future (co)variances, the modeling of temporal and cross-sectional dependencies of the (co)variation of asset returns arose as an important field of research in financial econometrics. The inherent problem that both practitioners and researchers are facing is that covariance matrices are not directly observable. Traditional modeling approaches therefore either assume the (co)variances to be measurable given past daily return information within observation-driven multivariate GARCH (MGARCH) models, or assume an inherently latent (co)variance process that is modeled together with the underlying returns within parameter-driven multivariate stochastic volatility (MSV) models (see, e.g., the surveys of Bauwens et al., 2006; Asai et al., 2006).

More recent approaches have increasingly focused on the modeling of daily realized covariance matrices. These realized covariance matrices depict non-parametric consistent ex-post estimates of the 'true' daily (co)variation of asset prices, computed from intraday asset return information (see, e.g., Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2004). It has been widely documented that observation- and parameter-driven models for realized covariance matrices provide more precise forecasts than MGARCH or MSV models, which only exploit daily asset return information (see, e.g., Golosnoy et al., 2012; Jin and Maheu, 2013, 2016; Hautsch et al., 2015; Callot et al., 2017).

Despite the superiority of this approach, the development of dynamic models for

the prediction of realized covariance matrices in empirically realistic scenarios of high-dimensional systems still remains a challenging task for three reasons: (i) Realized covariance models have to deal with the so-called 'curse of dimensionality', arising from the fact that the number of model parameters quadratically increases in the asset dimension, and thus become unmanageable even for a handful of assets. (ii) Realized covariance models need to ensure positive definiteness of the covariance prediction. (iii) Consistent and well-conditioned high-dimensional realized covariance data has to be available.

Solving the curse of dimensionality requires sparsity assumptions on the model in order to reduce the number of parameters. This can, e.g., be achieved by using LASSO (least absolute shrinkage and selection operator) regressions and employing carefully designed models which allow for multi-step estimation and covariance targeting. Positive-definiteness is often imposed by using matrix-transformations like the Cholesky decomposition or the matrix-logarithm. A natural approach, which simultaneously assures sparsity and positive-definiteness by construction, is to exploit common factor structures in the covariances. The factor approach has a long-established history in economics and finance and dates back to the single index model of Sharpe (1964) and the Arbitrage Pricing Theory of Ross (1976).

Even though the third challenge applies to the realized covariance data itself, it is nevertheless interconnected to the econometric modeling approach. Simple realized covariance measures like the realized covariance estimator or the multivariate realized kernel estimator (Barndorff-Nielsen and Shephard, 2004; Barndorff-Nielsen et al., 2011) tend to behave poorly if the asset dimension increases, as their construction relies on synchronization of the intraday prices. This drastically reduces the intraday data points available for the construction of the realized measure, and results in noisy or even inconsistent estimates of the 'true' integrated covariance matrix. Particularly for observation-driven models, reliable realized covariance estimates play a crucial role as these models treat them as the 'true' daily asset covariance. In contrast, state-space frameworks mitigate the impact of the noisiness in that realized

(co)variances only serve as a measurement for the latent integrated covariance (see, e.g., Asai et al., 2012a,b).

Nevertheless, in high-dimensional scenarios more refined realized covariance estimators have to be applied. State-of-the-art examples are, e.g., the composite realized kernel estimator of Lunde et al. (2016) and the CholCov estimator of Boudt et al. (2017) which allow for asynchronicity in the intraday price data when constructing the realized measure. This approach yields more efficient estimates, but the econometric model performance can again be negatively affected when the differences in synchronicity are too pronounced.

This thesis contributes to the literature by proposing three novel state-space frameworks for the modeling and forecasting of realized covariance matrices of asset returns. These models include two related but distinct factor state-space approaches which tackle the curse of dimensionality by making use of observed risk factors and by imposing different specifications of sparsity. Furthermore, a Cholesky type modeling approach is proposed which introduces and exploits a new matrix-variate distribution capable of coping with liquidity differences in the underlying intraday data. The proposed models are computationally tractable and applicable in large dimensional settings which makes them valuable for practitioners and researchers in the field of finance and financial econometrics.

Overall, this thesis comprises three self-contained essays on the modeling and prediction of realized covariance matrices of asset returns using state-space models. The essay in Chapter 2 has been joint work with Prof. Dr. Roman Liesenfeld and Dr. Bastian Gribisch. The essay in Chapter 4 is joint work with Dr. Bastian Gribisch. Essay 3 is a single-authored project. The essays as well as my contributions to them are summarized in the following.[1]

Chapter 2 corresponds to the paper "Factor state-space models for high-dimensional realized covariance matrices of asset returns" (Gribisch et al., 2020) and was published in the *Journal of Empirical Finance*. This paper proposes a dynamic factor

[1] An additional listing of my contributions to the joint projects is given in the Section 'Kooperationen' at the end of this thesis.

state-space model for high-dimensional covariance matrices of asset returns. The factor state-space model links the measured co-variation of the assets to economically motivated observed risk factors, like in the Fama and French (1993) modeling framework where factors are constructed using portfolios formed by firm characteristics. Thereby, the joint integrated covariance matrix of the assets and factors is modeled as a latent state variable which is assumed to be observed through their noisy realized covariance matrix with a Wishart measurement density. A strict factor structure for the marginal integrated covariance matrix of the assets is imposed, which decomposes the latter into a low-rank component driven by the factors' integrated covariance and a sparse diagonal matrix for the residual components. This factor structure translates into a factorization of the Wishart measurement density in such a way that the proposed multivariate state-space model for the covariance matrix can be devoted into conditionally independent low-dimensional state-space models, which facilitates statistical inference and makes the approach scalable with respect to the number of assets.

Statistical inference of the model at hand is conducted by making use of simulation based Bayesian techniques. For the purpose of parameter estimation a Markov Chain Monte Carlo scheme is proposed which neatly combines state-of-the-art procedures for the sampling of the static parameters and latent state series from their respective full conditional distributions. For its implementation computational parallelization techniques are exploited allowing for a relatively fast exploration of the joint posterior of parameters and latent states in medium to large dimensional applications. In order to provide a complete toolbox for practitioners a Bayesian model comparison approach based on the Deviance Information Criterion (DIC) is proposed. Therefore, a simple Bootstrap Particle Filter algorithm is implemented to estimate the likelihood function of the assets' realized covariance matrices conditional on the risk factors.

In an extensive empirical study the performance of different specifications of the resulting Wishart factor state-space (WFSS) model is investigated. To this end, the WFSS model is applied to daily realized covariance matrices for the returns of 60

NYSE traded stocks, which cover six large industry sectors in the S&P500. The specifications under consideration include WFSS models with either time-constant or time-varying factor loadings, and with either short-memory AR(1) or long-memory type heterogeneous AR dynamics for the logarithmic idiosyncratic and factor variance processes. In addition, different factor specifications are considered. These include a one-factor model using a market factor only, a three-factor model using the Fama-French factors and a nine-factor model using the Fama-French factors as well as sector-specific Exchange Traded Funds (ETFs).

A Bayesian model comparison based on the DIC shows that WFSS models incorporating time-varying dynamics for the factor loadings in combination with long-memory dynamics are extremely useful in explaining the observed dynamic variation in the covariance matrix of the stock returns. Furthermore, the results indicate strong evidence against a one-factor model in favor for the nine-factor specifications. The latter extension is also found to be critical to justify the approximation of the covariance matrix for the residual component by a diagonal matrix.

In an out-of-sample forecasting exercise the one-step-ahead predictive performance of the WFSS approach relative to competing models along several dimensions is illustrated: First, the accuracy of the (co)variance predictions and the reliability of density forecasts is examined. Second, in an economic evaluation of the predictive performance the Value-at-Risk forecasts for an equally-weighted portfolio and the global-minimum-variance-portfolio forecast under short-selling constraints are considered. The out-of-sample results show that the WFSS model performs favorable in nearly all dimensions relative to its competitors.

My contributions to Chapter 2 are as follows: Besides assisting the writing process by steady revision of the draft of the paper, I contributed to the development of the model dynamics and to the development of the proposed Markov Chain Monte Carlo (MCMC) procedure used for parameter estimation and forecasting purposes. I implemented all of the proposed algorithms in MATLAB and enabled a fast MCMC sampling scheme by exploiting parallel computation techniques. Furthermore, I im-

plemented the bootstrap particle filter algorithm needed for the computation of the DIC and I did the calculations for the statistical and economic evaluation of the predictive performance of the WFSS model. The latter also includes the implementation of the benchmark models.

Chapter 3 proposes a new flexible factor state-space framework for the prediction of realized covariance matrices of asset returns which is capable of efficiently handling dynamic systems of vast dimensions including 200 assets and more. Similar to the WFSS model discussed in Chapter 2, this approach proposes to model the realized asset covariance matrices using an observed factor structure. However, instead of imposing a joint modeling framework for the realized asset and factor covariance matrices, it is proposed to independently model the individual components of the realized factor decomposition.

This separate modeling approach theoretically has certain drawbacks as it ignores potential data-imposed dependencies in the measurements. However, it is found to be practically advantageous as it drastically reduces the model complexity and enables an entirely parallel handling of the individual components without any bottleneck like, e.g., a single d.o.f. parameter. Moreover, strict assumptions as the often imposed diagonal residual sparsity can be relaxed without effort, allowing to model empirically more realistic scenarios.

The method relies on a joint construction of the realized measure for the factors and assets. After observing the joint covariance matrix, realized factor loadings and realized residual components are constructed from standard matrix decompositions resulting in a time-series for each of the individual parts. Now, independent tractable state-space frameworks are applied to model and predict the individual time-series of the factors, factor loadings and residual components separately. Forecasts of the full covariance matrix of the assets are then obtained by combining the separate predictions via the factor structure.

For the factor and residual covariance components of the resulting composite factor state-space (CFSS) model the state-space approach of Windle and Carvalho (2014)

is adopted, which combines Wishart measurement densities with parsimoniously parameterized Matrix-Beta processes for the corresponding latent integrated precision matrices. For the proposed model several parameter restrictions are discussed and their usefulness is examined in an out-of-sample study. As further contribution to the literature the paper at hand provides additional insights on the dynamics of this modeling approach. For the factor loadings well established models from the time-varying parameter vector autoregression literature are adopted (Moura and Noriller, 2019).

The CFSS model is entirely tractable in a sense that the predictive distribution, and hence, the likelihood function, is available in closed-form for each individual component. In contrast to other state-space approaches that require high computational effort for parameter estimation and the prediction of future realizations, this approach consequently enables parameter estimation and prediction within only few minutes.

In an extensive out-of-sample forecasting and portfolio selection exercise for 225 stocks the performance of the CFSS model is examined. Thereby, different specifications regarding the residual sparsity as well as different restrictions to the dynamics of the realized covariance components are imposed. The residual specifications include block diagonality assumptions based on the Global Industry Classification Standard sector classifications and strict diagonality assumptions as well as combinations of both using linear shrinkage. Moreover, different sets of risk factors are considered.

In terms of predictive ability measured through different loss functions for point predictions of the covariance matrices, the CFSS model is found to significantly outperform several state-of-the-art competitors over four forecasting horizons ranging from one-day-ahead to one-month-ahead. Particularly the CFSS model incorporating a set of twelve observed risk factors in combination with a linear shrinkage-type residual specification shows outstanding performance. An economic forecast evaluation shows that the CFSS model is a valuable tool for optimal portfolio allocation decisions in a global-minimum-variance and a mean-variance framework, which is confirmed under different side restrictions.

Chapter 4 corresponds to the paper “Modeling realized covariance measures with

heterogeneous liquidity: A generalized matrix-variate Wishart state-space model" (Gribisch and Hartkopf, 2020). This paper introduces a new matrix-variate distribution to the literature on the modeling of realized covariance matrices of asset returns, i.e., the so-called Riesz distribution of Hassairi and Lajmi (2001).[2] The Riesz distribution depicts a generalization of the Wishart that allows to capture more complex measurement error structures. This is achieved by incorporating additional degree of freedom parameters (d.o.f.) in contrast to the central Wishart, which depends on a single d.o.f. driving the conditional covariance structure. Under the assumption of equal d.o.f. parameters the Riesz collapses to the Wishart.

In the context of realized covariance measures the need for such a generalized Wishart is motivated by the presence of heterogeneous liquidity across assets mirrored by missing intraday return data. For standard multivariate realized measures, like the simple realized covariance estimators (Barndorff-Nielsen and Shephard, 2004) or the more robust realized kernel estimators (Barndorff-Nielsen et al., 2011) liquidity differences are not present due to the fact of price synchronization. However, since these basic measures induce severe data loss in increasing dimensions and biased covariance estimates due to stale prices and zero returns, up-to-date realized measures like the composite realized kernel estimators (Lunde et al., 2016) have been developed to exploit a maximum of available data by implementing two-dimensional realized measures for all pairs of assets. On one hand this approach has the advantage of yielding efficient estimates of the covariance of asset returns in high dimensions, on the other hand, potential liquidity differences in the assets find their way into the realized measure.

In order to give a better understanding of the distribution at hand, this paper first analyzes the stochastic properties of the Riesz. Identities for the characteristic function of the central Riesz and expressions for the conditional first and second order moments are derived. The Riesz density function is based on a Cholesky decomposition of the realized measure. Unfortunately, it is thus depending on the ordering of

[2]It is worth to note, that the Riesz distribution is sometimes also known as Bellman-Gamma distribution.

the assets. However, based on the findings for the conditional covariance structure and the Riesz' data generating process a natural ordering of the assets by increasing liquidity is proposed. Furthermore, it is shown that the density neatly factorizes in the product of conditionally independent Gamma and Normal densities for the individual parts of the Cholesky decomposition.

The benefits of the Riesz distribution in comparison to the Wishart are examined in an empirical application to 30 NYSE traded stocks with heterogeneous liquidity. Therefore, a Riesz state-space (RSS) model is developed. The model exploits the Cholesky-structure and the factorization of the Riesz measurement density in order to obtain a conditionally independent sequence of state-space models for the conditional Cholesky coefficients. For fast and efficient statistical inference on the independent state-space models a Bayesian MCMC scheme is proposed, which is completely parallelized over the number of assets. Moreover, to cope with potential overfitting problems in high-dimensional state-space settings the straightforward implementation of shrinkage prior restrictions (Bitto and Frühwirth-Schnatter, 2019) on the conditional state variances is illustrated.

The empirical results indicate that a Wishart-implied attenuation of measurement errors for less liquid assets due to its single d.o.f. parameter downwards biases the persistence and upwards biases the conditional variance of the latent state processes. This bias is successfully alleviated by the Riesz. Distinct disparities in the estimated Riesz d.o.f. parameters under the proposed liquidity sorting also show evidence that the Wishart is empirically rejected by the data. This finding is confirmed in a Bayesian model comparison exercise based on the DIC for six randomly selected five-dimensional subsets of the full data set.

An out-of-sample forecasting exercise shows that the proposed RSS model significantly outperforms its Wishart competitor as well as several other competing models in terms of one-step-ahead predictive accuracy. In an economic evaluation based on global-minimum-variance portfolios the Riesz is found to perform superior in terms of portfolio standard deviations, Sharpe ratios and turn-over when compared to the

Wishart.

My contributions to Chapter 4 are as follows: First, I discovered the Riesz distribution in a differing strand of the literature and conducted a thorough review of the theoretical and applied papers connected to the Riesz. Second, I proved the propositions stated in the paper. Third, besides participating in the development of the RSS model dynamics, I fully developed and implemented the proposed MCMC scheme for parameter estimation and forecasting. This particularly includes the translation of the GIG random number generator of Hörmann and Leydold (2014, 2015) needed for the shrinkage prior implementation from `C/C++` to MATLAB. I enhanced the translated random number generator to properly handle limiting cases of the GIG distribution.[3] Furthermore, I proposed and implemented the Rao-Blackwellized particle filter algorithm required for the computation of the DIC and I performed the calculations for the statistical and economic evaluation of the (predictive) performance of the RSS model. The latter also includes the implementation of the benchmark models. Finally, I contributed by authoring several sections and steadily revising the draft of the paper.

[3]The MATLAB version of the GIG random number generator is publicly available (Hartkopf, 2020).

Chapter 2.

Factor State-Space Models for High-Dimensional Realized Covariance Matrices of Asset Returns

2.1. Introduction

Modeling and forecasting covariance matrices of asset returns is important in risk management and portfolio allocation. Recent contributions to this field increasingly make use of realized covariance matrices which provide non-parametric ex-post estimates for the latent integrated covariance matrices of asset returns, and develop dynamic time-series models for those estimates.[1] Pioneering approaches to modeling and predicting realized covariance matrices are found in Gourieroux et al. (2009), Chiriac and Voev (2011), Bauer and Vorkink (2011), Noureldin et al. (2012), Golosnoy et al. (2012) and Jin and Maheu (2013). However, the models developed in those studies typically suffer from a proliferation of parameters in high dimensional applications and the estimation of their parameters becomes rapidly difficult as the number of assets increases so that they often have limited practical relevance in realistic financial applications. Strategies which have been proposed to overcome those difficulties include the design of models such that their parameters can be iteratively estimated by multistep procedures (Bauwens et al., 2012, 2016), the use of LASSO (least absolute shrinkage and selection operator) type estimation techniques (Callot

This chapter is based on:
Factor State-Space Models for High-Dimensional Realized Covariance Matrices of Asset Returns;
Authors: Bastian Gribisch, Jan P. Hartkopf and Roman Liesenfeld.
Journal of Empirical Finance, doi: 10.1016/j.jempfin.2019.08.003,

[1] For a description of the concept of realized covariance matrices see, for example, Andersen et al. (2003); Barndorff-Nielsen and Shephard (2004); Park and Linton (2012); Lunde et al. (2016).

et al., 2017), or the application of sparse factor structures for the assets' covariance matrix (Tao et al., 2011; Asai and McAleer, 2015; Jin et al., 2019; Sheppard and Xu, 2019). The particular appeal of using a factor approach relative to the other alternatives is that factor models can be economically motivated and have a long established history in explaining the variation of financial returns. Prominent examples thereof are the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), the Arbitrage Pricing Theory (APT) of Ross (1976) and the three-factor model of Fama and French (1993).

In this paper we also adopt a factor approach and propose a dynamic factor state-space model for high-dimensional covariance matrices of asset returns which can be easily statistically analyzed by a combination of fairly standard Bayesian Markov Chain Monte Carlo (MCMC) procedures. It makes use of observed risk factors (such as those in the Fama-French model) and takes the joint integrated covariance matrix of the assets and factors as a latent state variable which is observed through their noisy realized covariance matrix with a Wishart measurement density. For the marginal integrated covariance matrix of the assets we impose a strict factor structure decomposing it into a low-rank component driven by the factors' integrated covariance and a sparse diagonal matrix for the residual components. In this matrix factor decomposition we allow for dynamic variation in the (co)variances of the factors and residual components as well as in the factor loadings. The key to a simple statistical analysis lies in the property of the Wishart measurement density that it factorizes under an observed strict factor structure in such a way that the proposed multivariate state-space model for the covariance matrix can be devoted into conditionally independent low-dimensional state-space models. Thus, we can rely on a simple MCMC approach and exploit for its implementation computational parallelization techniques making the approach fully scalable w.r.t. the number of assets.

Our Wishart factor state-space (WFSS) model which we apply to an analysis of the US stock market builds upon and generalizes the model of Sheppard and Xu (2019). Their factor HEAVY (high-frequency based volatility) model also combines

a Wishart density for the realized covariance matrix with an observed strict factor structure. However, in its application the authors use a single-factor structure based on the market factor in the CAPM model. In our approach we consider additional risk factors and utilize the three Fama-French factors as well as sector-specific Exchange Traded Funds (ETFs). Such an extension can be expected to be critical since a single factor may not suffice to justify the approximation of the covariance matrix for the residual component by a diagonal matrix. Empirical evidence that the CAPM market factor is by far not sufficient to eliminate the correlation in the residual component and that the combination of the three Fama-French factors with sector-specific ETFs substantially improves the sparsity of the residual correlation is provided by Fan et al. (2016) and is also confirmed by our empirical results in Section 2.4. Moreover, recent results of Aït-Sahalia and Xiu (2017) indicate that latent factor structures uncovered by a principal-component analysis are well approximated by observable factor structures including sector-specific ETFs. Similar results have been found by Pelger (2019). Our Bayesian factor state space approach also differs from the factor HEAVY model in that the latter is an observation-driven GARCH-type framework statistically analyzed by a two-step quasi Maximum Likelihood (ML) procedure. This two-step approach is required as the joint estimation of all the parameters (including the degrees of freedom of the Wishart distribution) becomes computationally difficult in high-dimensional applications. Last but not least, the factor HEAVY model as implemented by Sheppard and Xu (2019) uses low-order GARCH$(1, 1)$ recursions so that it ignores potential long-memory type dynamics in the realized (co)variances of asset returns which is often found in empirical applications (see, e.g. Corsi et al., 2012; Bekierman and Manner, 2018, and the literature cited therein). In order to accommodate potential long-memory we endow the state variables directing the variances in the WFSS model with heterogeneous autoregressive (HAR) processes which are known to provide effective approximations of long-memory dynamics (Corsi, 2009).

An alternative to the observed factor approach with explicit exogenously supplied factors as adopted in this paper is to rely on implicit factors extracted from realized

covariance data as proposed by Tao et al. (2011) and Asai and McAleer (2015). They first construct from the realized covariance matrices of stock returns via an eigenanalysis the covariance matrices of common factors for which they build in a second step dynamic time-series models. In contrast to our approach, this two-step procedure is based on the assumption that both the factor loadings as well as the residual covariance matrix are time-invariant which restricts the flexibility to account for nontrivial contemporaneous and dynamic interactions in the (co)variances of asset returns. The same applies to the factor approach of Jin et al. (2019) which imposes an eigenvalue decomposition directly in the dynamic model for the realized covariance matrix of stock returns.

As a preview of our main empirical results, we apply the WFSS model to daily covariance matrices for the returns of 60 NYSE traded stocks and find by using Bayesian model comparisons that the single factor approach based on the CAPM market factor is empirically rejected in favor of a factor structure including the three Fama-French factors and the sector-specific ETF factors. Most importantly, this extended factor structure combined with a WFSS allowing for dynamically varying factor loadings and long-memory type dependence is extremely useful in explaining the observed dynamic variation in the covariance matrix of the stock returns. We also run out-of-sample forecasts and illustrate the predictive performance of our approach relative to competing models along several dimensions: Accuracy of the (co)variance predictions, reliability of density forecast and the ability to produce predictions for the Value-at-Risk and the global-minimum variance portfolio. Our results show that our WFSS model performs favorable in nearly all dimensions relative to its competitors.

The rest of the paper is organized as follows: Section 2.2 introduces the baseline model. The proposed MCMC procedure for the Bayesian posterior analysis and model comparisons as well as the construction of forecasts are discussed in Section 2.3. Section 2.4 presents the empirical application to NYSE data and Section 2.5 concludes. Additional material is provided in Appendix A.

2.2. The Model

2.2.1. Factor structure

Consider the $m \times m$ realized covariance matrix C_t used to approximate the joint period-t integrated covariance matrix Σ_t for a vector of log-prices for p individual assets together with q observable risk factors with $m = p + q$. Here, p is assumed to be substantially larger than q. Let Σ_t and C_t be partitioned as

$$\Sigma_t = \begin{pmatrix} \Sigma_t^f & \Sigma_t^{rf} \\ \Sigma_t^{fr} & \Sigma_t^r \end{pmatrix}, \qquad C_t = \begin{pmatrix} C_t^f & C_t^{rf} \\ C_t^{fr} & C_t^r \end{pmatrix}, \tag{2.1}$$

where Σ_t^r denotes the $p \times p$ integrated covariance matrix for the assets, Σ_t^f the $q \times q$ integrated covariance matrix for the factors and Σ_t^{rf} the $q \times p$ matrix of the integrated covariances between the factors and assets. C_t in Eq. (2.1) is partitioned conformably with Σ_t so that C_t^r is the realized covariance matrix of the assets, C_t^f that of the factors and C_t^{rf} the matrix of the realized covariances between the factors and the assets.

Our aim is to predict the potentially large-dimensional realized covariance matrix C_t^r. For this purpose, we propose a joint dynamic model for the realized (co)variances of the factors and assets in C_t. This model is based on a decomposition of the assets' integrated covariance matrix Σ_t^r into a low-rank component driven by the integrated covariance matrix of the observed risk factors Σ_t^f and a residual component. Such a decomposition obtains by imposing a continuous-time factor model for the asset prices relating them to the factor prices and assuming that the factor prices and the residual components of the asset prices are uncorrelated (Fan et al., 2016; Aït-Sahalia and Xiu, 2017). The resulting integrated covariance matrix for the assets obtains as

$$\Sigma_t^r = \mathcal{B}_t \Sigma_t^f \mathcal{B}_t' + \Sigma_t^e, \tag{2.2}$$

where $\mathcal{B}_t$ denotes the integrated factor loading matrix of size $p \times q$ and Σ_t^e is the

integrated covariance matrix of the residual components of the factor model. Here, we assume a strict factor model such that Σ_t^e is a diagonal matrix. This ensures that the number of parameters increases only linearly in the number of assets which is obviously desirable in (very) high-dimensional applications. Clearly, assuming Σ_t^e to be diagonal appears to be fairly restrictive, especially, for a small number of observed factors like in the one-factor CAPM and the three-factor model of Fama and French (1993), where we cannot expect the residual correlations to be negligible. However, when augmenting those small factor models by including sector specific Exchange Traded Funds (ETFs) as observed additional factors the sparsity of the residual correlation significantly improves (see Fan et al., 2016, and the empirical results in Section 2.4.2).

In our approach, we allow the factor risk premia in Σ^f and the idiosyncratic risks in Σ^e as well as the betas given by the factor loadings in $\mathcal{B}$ to be time-varying. While time-variation in the factor risk premia and the idiosyncratic risk now is well accepted, it is less so for the betas. However, the recent empirical literature reports increasing evidence of dynamically varying betas in a one-factor CAPM model (Andersen et al., 2005; Ghysels and Jacquier, 2006; Ang and Chen, 2007; Kalnina, 2015; Sheppard and Xu, 2019) as well in a Fama-French three-factor model (Bollerslev and Zhang, 2003; Engle, 2016). A further justification for time-varying betas, is that they obtain as population regression coefficients $\mathcal{B}_t = \Sigma_t^{fr}(\Sigma_t^f)^{-1}$ so that their time-invariance would require to impose severe restrictions on the dynamics of the joint integrated covariance matrix Σ_t (Engle, 2016).

2.2.2. Wishart factor state-space model

Taking the integrated covariance matrix Σ_t with its components as a latent state variable observed through the noisy realized covariance matrix C_t it is reasonable to model C_t by a state-space approach with a measurement density $f(C_t|\Sigma_t)$ relating the measurements C_t to the states Σ_t and a transition density $f(\Sigma_t|\Sigma_{1:t-1})$ for the time-varying Σ_t to be designed to approximate the observed dynamics of C_t. The notation $A_{s:\tau}$ is used to denote the collection $\{A_s, \ldots, A_\tau\}$.

A natural selection for the measurement density $f(C_t|\Sigma_t)$ of the positive definite matrix valued realized covariance measure C_t is that of a central m-dimensional Wishart distribution, $C_t|\Sigma_t \sim \mathcal{W}_m(n, \Sigma_t/n)$, where $n \geq m$ is the scalar degree of freedom and Σ_t/n is the scale matrix (see Philipov and Glickman, 2006; Golosnoy et al., 2012 and Noureldin et al., 2012, for applications of the Wishart distribution to realized covariance matrices). The scale matrix is normalized by n so that the conditional expectation of C_t is given by $\mathrm{E}(C_t|\Sigma_t) = \Sigma_t$. The density function of this Wishart distribution is

$$f_{\mathcal{W}}(C_t|n, \Sigma_t/n) = \frac{|C_t|^{(n-m-1)/2}|\Sigma_t/n|^{-n/2}}{2^{nm/2}\pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma([n+1-i]/2)} \exp\left\{-\frac{n}{2}\mathrm{tr}\left(\Sigma_t^{-1}C_t\right)\right\}, \quad (2.3)$$

where $\Gamma(\cdot)$ denotes the Gamma function.

An important advantage of using a Wishart measurement density for C_t is that it admits a parametrization reflecting the factor decomposition of the integrated covariance of the assets Σ_t^r in Eq. (2.2), which when combined with a diagonal form of the integrated residual covariance Σ_t^e translates into a convenient factorization of the measurement density. This greatly simplifies the statistical inference in high-dimensional applications. In particular, using the partitioning of C_t and Σ_t in Eq. (2.1) and taking

$$C_t^e = C_t^r - C_t^{fr}(C_t^f)^{-1}C_t^{rf}, \quad \Sigma_t^e = \Sigma_t^r - \Sigma_t^{fr}(\Sigma_t^f)^{-1}\Sigma_t^{rf}, \quad \mathcal{B}_t = \Sigma_t^{fr}(\Sigma_t^f)^{-1}, \quad (2.4)$$

the Wishart density function for C_t in Eq. (2.3) factorizes into the product of a Wishart density for C_t^f, a conditional Gaussian density for C_t^{fr} given C_t^f and an independent Wishart density for C_t^e (Muirhead, 2005, Theorem 3.2.10),

$$\begin{aligned} f_{\mathcal{W}}(C_t|n, \Sigma_t/n) = {} & f_{\mathcal{W}}\left(C_t^f \mid n,\, \Sigma_t^f/n\right) \\ & \times f_{\mathcal{MN}}\left(C_t^{fr} \mid \mathcal{B}_t C_t^f,\, (\Sigma_t^e/n) \otimes C_t^f\right) f_{\mathcal{W}}\left(C_t^e \mid n-q,\, \Sigma_t^e/n\right). \end{aligned} \quad (2.5)$$

Here the function $f_{\mathcal{MN}}$ is the density of a matrix-variate normal distribution for C_t^{fr} with mean and covariance for $\mathrm{vec}(C_t^{rf})$ given by $\mathrm{vec}(C_t^f\mathcal{B}_t')$ and $(\Sigma_t^e/n) \otimes C_t^f$,

respectively. In order to highlight the implications of this factorization for the observed factor structure assumed for the covariance matrix of the asset returns, we take C_t^e defined in Eq. (2.4) to write the assets's realized covariance matrix as

$$C_t^r = B_t C_t^f B_t' + C_t^e, \qquad B_t = C_f^{fr}(C_t^f)^{-1}. \tag{2.6}$$

Since B_t and C_t^e respectively define realized measures for the integrated factor loadings $\mathcal{B}_t$ and the integrated residual covariance matrix Σ_t^e, Eq. (2.6) represents the realized equivalent to the factor decomposition in Eq. (2.2) assumed for the integrated covariance for the assets Σ_t^r. According to the Wishart factorization in Eq. (2.5), the individual realized factor components in Eq. (2.6) then have the following properties: First, the realized residual covariance matrix C_t^e is stochastically independent of the components B_t and C_t^f defining the systematic factor part, and follows a Wishart distribution with mean $\mathrm{E}(C_t^e|\Sigma_t) = (n-q)\Sigma_t^e/n$. As a consequence, the conditional distribution of the covariance of the asset C_t^r given (B_t, C_t^f) is a shifted Wishart. Second, since the Wishart factorization implies that C_t^{fr} given C_t^f is matrix-variate normally distributed, it follows that the realized factor loadings $B_t = C_f^{fr}(C_t^f)^{-1}$ are conditionally on C_t^f also matrix-variate normally distributed with mean $\mathrm{E}(\mathrm{vec}(B_t)|C_t^f, \Sigma_t) = \mathrm{vec}(\mathcal{B}_t)$ and covariance matrix $\mathrm{Cov}(\mathrm{vec}(B_t)|C_t^f, \Sigma_t) = (\Sigma_t^e/n) \otimes (C_t^f)^{-1}$. Finally, the Wishart factorization implies that the marginal distribution of the realized factor covariance matrix C_t^f is a Wishart with mean $\mathrm{E}(C_t^f|\Sigma_t) = \Sigma_t^f$.

If we now combine the Wishart factorization in Eq. (2.5) with the assumption that the integrated residual covariance is diagonal with $\Sigma_t^e = \mathrm{diag}(\sigma_{1t}^e, \ldots, \sigma_{pt}^e)$, then the joint Wishart density for C_t as a function of the latent states $\Sigma_t = (\Sigma_t^f, \mathcal{B}_t, \Sigma_t^e)$ obtains as

$$\begin{aligned} f_{\mathcal{W}}(C_t|n, \Sigma_t/n) \propto & \left[|\Sigma_t^f|^{-n/2} \exp\left\{-\frac{n}{2}\mathrm{tr}[(\Sigma_t^f)^{-1}C_t^f]\right\}\right] \\ & \times \left[\prod_{i=1}^{p} (\sigma_{it}^e)^{-n/2} \exp\left\{-\frac{n}{2\sigma_{it}^e}(\beta_{it}' C_t^f \beta_{it} - 2\beta_{it}' c_{it}^{rf} + c_{it}^r)\right\}\right], \end{aligned} \tag{2.7}$$

where $\beta_{it} = (\beta_{i1t}, \ldots, \beta_{iqt})'$ denotes the vector of loadings of asset i on the q factors such that $\mathcal{B}_t = (\beta_{1t}, \ldots, \beta_{pt})'$ and $c_{it}^{rf} = (c_{i1t}^{rf}, \ldots, c_{iqt}^{rf})'$ is the vector of realized covariances between the q factors and the ith asset with $C_t^{rf} = (c_{1t}^{rf}, \ldots, c_{pt}^{rf})$. The scalar c_{it}^{r} is the ith diagonal element of C_t^r representing the realized variance of asset i. The component in the first bracket in Eq. (2.7) is the kernel of the Wishart density for the realized factor covariance C_t^f as given in Eq. (2.5), while the component in the second bracket is obtained from the multiplication of the Wishart density kernel for the realized residual covariance C_t^e by the conditional Gaussian density kernel for C_t^{fr} given C_t^f. As a result of the diagonal form of Σ_t^e, this second component factorizes into p (functionally) independent factors, one for each asset.

In order to complete the factor state-space model, we specify the transition densities for the latent time-varying integrated covariance matrices (Σ_t^f, Σ_t^e) and the factor loadings in $\mathcal{B}_t$. To accommodate the observed dynamics in the realized covariance matrix with its typically strong persistence, especially in the realized variances, we combine simple AR(1) processes for the factor loadings with heterogeneous autoregressive (HAR) processes for the integrated variances. Those HAR processes as introduced by Corsi (2009) provide simple yet effective approximations of long-memory type persistence and are well-suited for modelling realized variances (see, e.g., Bekierman and Manner, 2018, and the literature cited therein).

For the logs of the p elements in the diagonal idiosyncratic covariance matrix Σ_t^e denoted by $x_{it}^e = \log \sigma_{it}^e$ we assume mutually independent Gaussian HAR processes of the form

$$x_{it}^e - \gamma_i^e = \phi_{i1}^e \bar{x}_{i[t-1:t-1]}^e + \phi_{i2}^e \bar{x}_{i[t-1:t-5]}^e + \phi_{i3}^e \bar{x}_{i[t-1:t-22]}^e + \nu_i^e \eta_{it}^e, \tag{2.8}$$

with $\eta_{it}^e \sim \mathcal{N}(0,1)$, $i = 1, \ldots, p$, where $\bar{x}_{i[t-1:t-h]}^e = \sum_{\tau=1}^{h} (x_{it-\tau}^e - \gamma_i^e)/h$ for $h = 1, 5, 22$ represents daily, weekly and monthly lags, respectively. The parameters are $\theta_i^e = (\gamma_i^e, \phi_{i1}^e, \phi_{i2}^e, \phi_{i3}^e, \nu_i^e)$, and for the restriction $\phi_{i2}^e = \phi_{i3}^e = 0$ the HAR process in Eq. (2.8) reduces to a standard AR(1). For the pq factor loadings in $\mathcal{B}_t$ we assume the following

independent standard Gaussian AR(1) processes:

$$\beta_{ikt} - \gamma_{ik}^{\beta} = \phi_{ik}^{\beta}(\beta_{ikt} - \gamma_{ik}^{\beta}) + \nu_{ik}^{\beta}\eta_{ikt}^{\beta} \tag{2.9}$$

with $\eta_{ikt}^{\beta} \sim \mathcal{N}(0,1)$, $i = 1,\ldots,p,\ k = 1,\ldots,q$, parameterized by $\theta_{ik}^{\beta} = (\gamma_{ik}^{\beta}, \phi_{ik}^{\beta}, \nu_{ik}^{\beta})$.

The factor covariance matrix Σ_t^f is Cholesky-decomposed into

$$\Sigma_t^f = L_t^{-1} D_t L_t^{-1\prime}, \tag{2.10}$$

where D_t is a diagonal matrix and L_t is a lower-triangular matrix with unit diagonal elements, say,

$$\begin{aligned} D_t &= \operatorname{diag}(\sigma_{1t}^f, \ldots, \sigma_{qt}^f), \\ L_t &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \ell_{21t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \ell_{q1t} & \ell_{q2t} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \ell_{1t}' \\ \ell_{2t}' \\ \vdots \\ \ell_{qt}' \end{pmatrix}. \end{aligned} \tag{2.11}$$

In order to allow for sufficient flexibility in accounting for the observed dynamics in the factor covariance we assume for the logs of the q-elements in D_t, denoted by $x_{kt}^f = \log(\sigma_{kt}^f)$ independent Gaussian HAR processes, that is

$$x_{kt}^f - \gamma_k^f = \phi_{k1}^f \bar{x}_{k[t-1:t-1]}^f + \phi_{k2}^f \bar{x}_{k[t-1:t-5]}^f + \phi_{k3}^f \bar{x}_{k[t-1:t-22]}^f + \nu_k^f \eta_{kt}^f, \tag{2.12}$$

where $\eta_{kt}^f \sim \mathcal{N}(0,1)$, $k = 1,\ldots,q$, with parameters $\theta_k^f = (\gamma_k^f, \phi_{k1}^f, \phi_{k2}^f, \phi_{k3}^f, \nu_k^f)$, and for the $q(q-1)/2$ free elements in the matrix of pseudo-loadings L_t independent Gaussian AR(1) processes,

$$\ell_{kjt} - \gamma_{kj}^{\ell} = \phi_{kj}^{\ell}(\ell_{kjt-1} - \gamma_{kj}^{\ell}) + \nu_{kj}^{\ell}\eta_{kjt}^{\ell}, \tag{2.13}$$

where $\eta_{kjt}^{\ell} \sim \mathcal{N}(0,1)$, $k > j = 1,\ldots,q-1$, and $\theta_{kj}^{\ell} = (\gamma_{kj}^{\ell}, \phi_{kj}^{\ell}, \nu_{kj}^{\ell})$.

For later reference we note that a HAR process of the form as given in Eqs. (2.8) and (2.12) can be written as the following restricted AR(22):

$$x_t - \gamma = \left[\phi_1 + \frac{\phi_2}{5} + \frac{\phi_3}{22}\right](x_{t-1} - \gamma) + \left[\frac{\phi_2}{5} + \frac{\phi_3}{22}\right](x_{t-2} - \gamma) + \cdots \tag{2.14}$$
$$+ \left[\frac{\phi_2}{5} + \frac{\phi_3}{22}\right](x_{t-5} - \gamma) + \frac{\phi_3}{22}(x_{t-6} - \gamma) + \cdots + \frac{\phi_3}{22}(x_{t-22} - \gamma) + \nu\eta_t,$$

where we have omitted the indices for the assets and factors.

The vector of parameters in this Wishart factor state space (WFSS) model as defined by Eqs. (2.7)–(2.13) consists of $5[p+q] + 3[pq + q(q-1)/2] + 1$ parameters, which are the set of HAR-parameters $(\gamma, \phi_1, \phi_2, \phi_3, \nu)$ for the $[p+q]$ state processes $\{x^e_{it}\}$ and $\{x^f_{kt}\}$, the set of AR(1) parameters (γ, ϕ, ν) for the $[pq + q(q-1)/2]$ state processes $\{\beta_{ikt}\}$ and $\{\ell_{kjt}\}$, and the degree of freedom n of the Wishart measurement density. Due to the factor structure, both the number of state processes as well as the number of parameters are (for a given set of factors) linear in the number of assets p. Still, for a large number of assets, the actual amount of parameters appears to be fairly large. For instance, with $p = 60$ assets and $q = 9$ observable factors as in our application below we have for the (unrestricted) WFSS model 645 state processes and 2074 parameters to be estimated. However, as we use a data set covering $T = 1510$ trading days with $(p+q)(p+q+1)T/2 = 3,646,650$ (co)variance observations we have 1758 observations per parameter which can be expected to provide enough information for a reliable statistical inference.

In order to efficiently handle the large number of parameters in the Bayesian MCMC posterior analysis we take full advantage of the factorization of the measurement density in Eq. (2.7). The fact that its factor component as a function in the factor-specific states Σ^f_t and its p asset components as functions in the asset-specific states $\{\sigma^e_{it}, \beta_{it}\}$ are mutually functionally independent together with the independent priors for the state processes given in Eqs. (2.8)–(2.13) allows us to use a simple MCMC approach and to exploit for its implementation computational parallelization techniques. This makes our approach fully scalable w.r.t. the number assets p.

Using the WFSS model for the joint covariance matrix of the assets and the observed risk factors we can perform forecasting of the assets' integrated covariance matrix Σ_t^r as further detailed in Section 2.3.3 below. Our WFSS framework also allows us to test the hypothesis that the factor loadings in $\mathcal{B}_t$ are time-invariant, which is the case if in the AR specification for the betas β_{ikt} in Eq. (2.9) $\phi_{ik}^{\beta} = 0$ and $\nu_{ik}^{\beta} \to 0$ so that $\beta_{ikt} = \gamma_{ik}^{\beta}\ \forall t$. Likewise, the framework enables us to analyze by means of corresponding model comparisons the relative importance of the different factors under consideration in explaining and predicting the observed variation of the asset returns. Instrumental for such an analysis is the conditional density of C_t^r given $(C_t^{fr}, C_t^f, \Sigma_t)$ for the respective set of included factors. As discussed further below in Section 3.2 this density is required to obtain the conditional likelihood for the observed realized covariance matrices of the assets. According to Eqs. (2.4) and (2.5) the realized covariance for the assets can be represented as $C_t^r = C_t^{fr}(C_t^f)^{-1}C_t^{rf} + C_t^e$ with $C_t^e \sim \mathcal{W}_p(n-q, \Sigma_t^e/n)$ and Jacobian $\mathrm{d}C_t^r = \mathrm{d}C_t^e$, so that the conditional density for C_t^r is given by

$$f(C_t^r \mid C_t^{fr}, C_t^f; \Sigma_t) = f(C_t^e \mid \Sigma_t^e) = f_{\mathcal{W}}\left(C_t^e \mid n-q,\, \Sigma_t^e/n\right). \tag{2.15}$$

2.2.3. A note on the Wishart assumption for the measurement density

Asset returns are typically subject to the presence of outliers generating fat tails in the corresponding covariance measures (Opschoor et al., 2017). Under the assumed conditional Wishart distribution, with its fairly thin tails, significant fat-tail behavior can only originate from the unconditional variation of the latent states (Σ_t^f, Σ_t^e, $\mathcal{B}_t$), but this may not suffice to fully capture the tail behavior of realized covariance data. In such cases, the Wishart can be usefully replaced by a distribution allowing for conditional heavy tails such as the Matrix-F distribution which obtains from a Wishart-inverted-Wishart mixture (Konno, 1991). A successful application of the matrix-F distribution to realized covariances is found in Opschoor et al. (2017) who combine it with an observation-driven generalized autoregressive score (GAS)

approach. The Matrix-F nests the Wishart distribution as a special case and admits a parametrization based on the strict factor decomposition in Eq. (2.2) leading (when conditioned on the mixing variable) to a factorization of the density for the realized covariance matrix C_t which is of the same form as under the Wishart (see Eqs. 2.5 and 2.7). This makes the Matrix-F distribution easily applicable for a straightforward fat-tailed generalization of the WFSS model.[2] However, the results we obtained for an initial posterior analysis of this generalization show that there is no evidence against the Wishart in favor of the fat-tailed Matrix-F distribution (results are not presented here). Actually, the posterior estimates for all the parameters of the Matrix-F generalization are virtually equal to their values obtained for the fitted WFSS model. This indicates that the marginalization of the conditional Wishart distribution for C_t w.r.t. the latent state variables suffices to capture the tail behavior of the realized covariance data. This is illustrated in Figure 2.1 where we plot the histogram of the observed realized variances of the Citigroup and Caterpillar stock together with their unconditional distribution predicted under the fitted WFSS model defined by Eqs. (2.7)–(2.13) for 60 assets using 9 factors (see Section 2.4 below).

2.3. Bayesian Posterior Analysis and Forecasting

We utilize MCMC methods for a Bayesian posterior analysis of the WFSS model and use the Gibbs approach to simulate from the joint posterior of the parameter and states

$$\pi(\{\theta_i^e\}, \{\theta_{ik}^\beta\}, \{\theta_k^f\}, \{\theta_{kj}^\ell\}, n, \{x_{i,1:T}^e\}, \{\beta_{i,1:T}\}, \{x_{k,1:T}^f\}, \{\ell_{kj,1:T}\} | C_{1:T}). \tag{2.16}$$

The factorization of the measurement density (2.7) combined with the independent priors for the state processes as specified by the state-transitions (2.8)–(2.13) imply that the WFSS model can be devoted in $p + q$ conditionally independent state-space models, one for each of the q factors and p assets. This allows us to update within the

[2]The resulting Matrix-F mixture factor state-space model is presented in Appendix A.3.

Gibbs approach their respective state processes and their state specific parameters factor-by-factor and asset-by-asset.

2.3.1. MCMC algorithm

Our proposed MCMC implementation for the WFSS model consists of the following Gibbs sampling steps:[3]

1.) *Sampling* $\{\theta_k^f\}$, $\{x_{k,1:T}^f\}$: For $\{\theta_k^f\}$, the parameters of the Gaussian HAR-models for the factor state processes $\{x_{k,1:T}^f\}$ as given in Eq. (2.12), we select independent natural conjugate Normal-inverted-Gamma priors. Thus we can directly simulate from their full conditional posteriors $\pi(\theta_k^f|x_{k,1:T}^f)$, $k = 1, \ldots, q$.

The measurement density in Eq. (2.7) together with the state transitions in Eqs. (2.10) – (2.12) for Σ_t^f define conditionally independent nonlinear non-Gaussian state-space models for the q factor state processes $\{x_{k,1:T}^f\}$ given the pseudo loadings in L_t. The resulting q full conditional posteriors for $\{x_{k,1:T}^f\}$ are

$$\pi(x_{k,1:T}^f|\ell_{k,1:T}, \theta_k^f, n, C_{1:T}) \propto \prod_{t=1}^{T} \exp\left\{-\frac{n}{2}\left[x_{kt}^f + (\ell_{kt}' C_t^f \ell_{kt}) \exp(-x_{kt}^f)\right]\right\} \quad (2.17)$$
$$\times f_{\mathcal{N}}\left(x_{kt}^f \,\middle|\, \gamma_k^f + \phi_{k1}^f \bar{x}_{k[t-1:t-1]}^f + \phi_{k2}^f \bar{x}_{k[t-1:t-5]}^f + \phi_{k3}^f \bar{x}_{k[t-1:t-22]}^f,\, [\nu_k^f]^2\right),$$

with $k = 1, \ldots, q$, where $f_{\mathcal{N}}(\cdot|\mu, \sigma^2)$ denotes a Gaussian density with mean μ and variance σ^2. To sample a full trajectory $x_{k,1:T}^f$ from its posterior in one block we use the Particle Gibbs (PG) procedure based on the Bootstrap Particle Filter (BPF) combined with Ancestor Sampling (AS). In a nutshell, the PG as proposed by Andrieu et al. (2010) is a standard Gibbs sampler where we can use the BPF (Gordon et al., 1993) inside the Gibbs procedure in order to propose approximate samples from the posterior in such a way that the 'ideal' but infeasible Gibbs sampler is approximated. In order to improve the mixing of the resulting PG algorithm we combine the BPF with AS as recently proposed

[3]Further details on the proposed MCMC scheme are given in Appendix A.1.

by Lindsten et al. (2014). A detailed description of this PG-AS procedure is also found in Grothe et al. (2019).[4] Since the q state processes with their specific parameters $\{(x^f_{k,1:T}, \theta^f_k)\}$ are conditionally independent they can be updated by first drawing in parallel the q parameter vectors and then running in parallel the PG-AS for the q state processes.

2.) *Sampling* $\{\theta^\ell_{kj}\}$, $\{\ell_{kj,1:T}\}$: For the parameters $\{\theta^\ell_{kj}\}$ of the $q(q-1)/2$ Gaussian AR-processes of the pseudo loadings ℓ_{kjt} in Eq. (2.13) we select Normal-inverted-Gamma priors so that they can be directly sampled from their full conditional posteriors $\pi(\theta^\ell_{kj}|\ell_{kj,1:T})$, $k > j = 1, \ldots, q-1$.

Let $\tilde{\ell}_{kt} = (\ell_{k1t}, \ldots, \ell_{kk-1t})'$ denote the $(k-1)$-dimensional vector consisting of the unrestricted elements in the vector of pseudo loadings ℓ_{kt} in Eq. (2.11) such that $\ell_{kt} = (\tilde{\ell}'_{kt}, 1, 0, \ldots, 0)'$, $k = 2, \ldots, q$. Then the measurement density (2.7) together with the state transitions for the ℓ_{kjt}'s in Eq. (2.13) define $q-1$ conditionally independent linear Gaussian state-space models for the $(k-1)$-dimensional processes $\{\tilde{\ell}_{k,1:T}\}$ given $\{x^f_{k,1:T}\}$. The corresponding full conditional posterior of $\tilde{\ell}_{k,1:T}$ has the following particular linear Gaussian form in $\tilde{\ell}_{kt}$ given $\tilde{\ell}_{kt-1}$:

$$\pi(\tilde{\ell}_{k,1:T}|x^f_{k,1:T}, \{\theta^\ell_{kj}\}, n, C_{1:T}) \propto \prod_{t=1}^{T} \exp\left\{-\frac{n}{2}\left[\tilde{\ell}'_{kt}\tilde{C}^f_{kt}\tilde{\ell}_{kt} - 2(\tilde{c}^f_{kt})'\tilde{\ell}_{kt}\right]\exp(-x^f_{kt})\right\} \times f_{\mathcal{N}}(\tilde{\ell}_{kt}\,|\,\gamma^\ell_k + \Phi^\ell_k(\tilde{\ell}_{kt-1} - \gamma^\ell_k)\,,\, \Sigma^\ell_k), \qquad (2.18)$$

for $k = 2, \ldots, q$, with $\gamma^\ell_k = (\gamma^\ell_{k1}, \ldots, \gamma^\ell_{kk-1})'$, $\Phi^\ell_k = \text{diag}(\phi^\ell_{k1}, \ldots, \phi^\ell_{kk-1})$ and $\Sigma^\ell_k = \text{diag}([\nu^\ell_{k1}]^2, \ldots, [\nu^\ell_{kk-1}]^2)$. The matrix $\tilde{C}^f_{kt}$ denotes the upper left block of C^f_t consisting of its first $k-1$ rows and columns, and $\tilde{c}^f_{kt}$ is the column vector consisting of the first $k-1$ elements of k'th row of C^f_t. To simulate the $k-1$ trajectories in $\tilde{\ell}_{k,1:T}$ in one block we can straightforwardly apply the forward-filtering backward-sampling (FFBS) procedure of de Jong and Shephard (1995)

[4] An alternative to the PG-AS procedure for sampling $x^f_{k,1:T}$ is the auxiliary mixture sampler based upon the Kalman filter as proposed by Kim et al. (1998). We also experimented with this alternative and the MCMC results we obtained are virtually the same as those based upon the PG-AS.

for linear Gaussian state-space models. The conditional independence of the state processes $\{\tilde{\ell}_{k,1:T}\}$ allows us to simulate them together with their specific parameters $\{(\theta^{\ell}_{k1}, \ldots, \theta^{\ell}_{kk-1})\}$ in parallel.

3.) *Sampling* $\{\theta^e_i\}$, $\{x^e_{i,1:T}\}$: For $\{\theta^e_i\}$, the parameters of the p HAR-processes for the log of the idiosyncratic variances in Eq. (2.8), we use Normal-inverted-Gamma priors, so that we can directly simulate from their full conditional posteriors $\pi(\theta^e_i|x^e_{i,1:T})$, $i = 1, \ldots, p$.

The measurement density in Eq. (2.7) together with the state transitions in Eq. (2.8) define conditionally independent nonlinear non-Gaussian state-space models for the p state processes for the logs of the idiosyncratic variances $\{x^e_{i,1:T}\}$ given the loadings $\{\beta_{i,1:T}\}$. The corresponding full conditional posteriors for $\{x^e_{i,1:T}\}$ are

$$\begin{aligned}
&\pi(x^e_{i,1:T}|\beta_{i,1:T}, \theta^e_i, n, C_{1:T}) \\
&\qquad \propto \prod_{t=1}^{T} \exp\left\{-\frac{n}{2}\left[x^e_{it} + (\beta'_{it} C^f_t \beta_{it} - 2\beta'_{it} c^{rf}_{it} + c^r_{it}) \exp(-x^e_{it})\right]\right\} \qquad (2.19) \\
&\qquad\qquad \times f_{\mathcal{N}}\left(x^e_{it} \,\middle|\, \gamma^e_i + \phi^e_{i1}\bar{x}^e_{i[t-1:t-1]} + \phi^e_{i2}\bar{x}^e_{i[t-1:t-5]} + \phi^e_{i3}\bar{x}^e_{i[t-1:t-22]}, \, [\nu^e_i]^2\right),
\end{aligned}$$

with $i = 1, \ldots, p$. To sample the full trajectory $x^e_{i,1:T}$ from its posterior in one block we use the PG-AS procedure. Here again we can exploit the inherent parallel structure in the simulation step for $\{(x^e_{i,1:T}, \theta^e_i)\}$.

4.) *Sampling* $\{\theta^{\beta}_{ik}\}$, $\{\beta_{i,1:T}\}$: As for all the other AR parameters we select for those of the pq time-varying factor loadings in Eq. (2.9) Normal-inverted-Gamma priors so that we can simulate directly from their full conditional posteriors $\pi(\theta^{\beta}_{ik}|\beta_{ik,1:T})$, $i = 1, \ldots, p$, $k = 1, \ldots, q$.

Since the measurement density in Eq. (2.7) together with the state transitions of the factor loadings β_{ikt} define p conditionally independent linear Gaussian state space models for $\{\beta_{i,1:T}\}$ given $\{x^e_{i,1:T}\}$, we can easily simulate by the FFBS

procedure each $\beta_{i,1:T}$ from its full conditional posterior

$$\pi(\beta_{i,1:T}|x^e_{i,1:T},\{\theta^\beta_{ik}\},n,C_{1:T}) \propto \prod_{t=1}^{T}\exp\left\{-\frac{n}{2}\left(\beta'_{it}C^f_t\beta_{it}-2\beta'_{it}c^{rf}_{it}\right)\exp(-x^e_{it})\right\} \times f_{\mathcal{N}}(\beta_{it}\,|\,\gamma^\beta_i+\Phi^\beta_i(\beta_{it}-\gamma^\beta_i),\,\Sigma^\beta_i), \tag{2.20}$$

with $i=1,\ldots,p$, where $\gamma^\beta_i=(\gamma^\beta_{i1},\ldots,\gamma^\beta_{iq})'$, $\Phi^\beta_i=\text{diag}(\phi^\beta_{i1},\ldots,\phi^\beta_{iq})$ and $\Sigma^\beta_i=\text{diag}([\nu^\beta_{i1}]^2,\ldots,[\nu^\beta_{iq}]^2)$. As in the previous steps the simulation of the beta states and corresponding parameters $\{(\beta_{i,1:T},\theta^\beta_{i1},\ldots,\theta^\beta_{iq})\}$ can be parallelized.

5.) *Sampling n*: In the last step we simulate the Wishart degrees of freedom parameter. We select for $n>m$ an uniform prior, $\text{pr}(n)$, defined on a discrete grid. Thus we can directly simulate from its full conditional posterior which is given by a multinomial distribution, i.e.,

$$\pi(n|\{x^e_{i,1:T}\},\{\beta_{i,1:T}\},\{x^f_{k,1:T}\},\{\ell_{k,1:T}\},C_{1:T}) \propto \left[\prod_{t=1}^{T}f_{\mathcal{W}}(C_t|n,\Sigma_t/n)\right]\text{pr}(n). \tag{2.21}$$

Our MCMC algorithm repeatedly cycles through Step 1.) to 5.). After dropping the draws from the first cycles as burn-in we use the draws from the next S cycles for the purpose of approximating the joint posterior in Eq. (2.16). Bayesian point estimates (posterior means) of the model parameters and latent state variables are then obtained as sample averages over the corresponding Gibbs draws.

2.3.2. Model comparison

For the purpose of comparing alternative WFSS model specifications obtained under different sets of risk factors included in C^f_t and/or different restrictions imposed on the state transitions densities in Eqs. (2.8)–(2.13) we rely upon the Deviance Information Criterion (DIC) based on the likelihood function (Spiegelhalter et al., 2002). For models with alternative sets of factors the joint realized covariances for the assets

and the factors C_t consist of different set of variables which prevents the use of the likelihood function for $C_{1:T}$ for model comparisons. Hence, we rely on the DIC based on the conditional likelihood function of the assets' realized covariances $C^r_{1:T}$ given $C^{fr}_{1:T}$ and $C^f_{1:T}$.

Let $\theta = (\{\theta^e_i\}, \{\theta^\beta_{ik}\}, \{\theta^f_k\}, \{\theta^\ell_{kj}\}, n)$ denote the list of all parameters. Then the conditional DIC is given by

$$\text{DIC} = -2 \log p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \hat{\theta}) + 2p_D, \tag{2.22}$$

with small values of the criterion preferred. The term $p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \hat{\theta})$ represents the conditional likelihood function evaluated at the posterior estimates for the parameters rewarding good fits, and p_D is the effective sample size, penalizing good fits achieved by means of excessively rich parameterizations. The effective sample size is defined as

$$p_D = -2 \left[\mathrm{E}_{post}[\log p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \theta)] - \log p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \hat{\theta}) \right], \tag{2.23}$$

where $\mathrm{E}_{post}[\log p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \theta)]$ is the mean of the conditional log-likelihood function taken w.r.t. the posterior distribution of θ.

Based on the result for the conditional density of C^r_t given (C^{fr}_t, C^f_t) provided in Eq. (2.15) the conditional likelihood function in Eqs. (2.22) and (2.23) obtains as

$$p(C^r_{1:T} | C^{fr}_{1:T}, C^f_{1:T}; \theta) = \int \prod_{t=1}^{T} f(C^e_t \mid \Sigma^e_t; \theta) f(\Sigma^e_t \mid \Sigma^e_{1:t-1}; \theta) \, \mathrm{d}\Sigma^e_{1:T}. \tag{2.24}$$

Under the Wishart density for $f(C^e_t \mid \Sigma^e_t; \theta)$ together with the diagonal assumption $\Sigma^e_t = \mathrm{diag}(\exp\{x^e_{1t}\}, \ldots, \exp\{x^e_{pt}\})$ and the independent priors for the state processes $\{x^e_{i,1:T}\}$ the conditional likelihood in Eq. (2.24) as a function in $\{x^e_{i,1:T}\}$ factorizes into

p asset-specific components which are functionally independent, so that

$$p(C^r_{1:T}|C^{fr}_{1:T}, C^f_{1:T}; \theta) \propto \prod_{i=1}^{p} \int \left[\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} \left[(n-q)x^e_{it} + nc^e_{it} \exp(-x^e_{it}) \right] \right\} \right. \\ \left. \times f_{\mathcal{N}}\left(x^e_{it} \,\middle|\, \gamma^e_i + \phi^e_{i1} \bar{x}^e_{i[t-1:t-1]} + \phi^e_{i2} \bar{x}^e_{i[t-1:t-5]} + \phi^e_{i3} \bar{x}^e_{i[t-1:t-22]},\ [\nu^e_i]^2 \right) \right] \mathrm{d}x^e_{i,1:T}, \tag{2.25}$$

where c^e_{it} denotes the ith diagonal element of the realized idiosyncratic covariance matrix C^e_t. For a given value of θ the p integrals w.r.t. the $x^e_{i,1:T}$'s can be taken as likelihood functions of independent univariate nonlinear non-Gaussian state-space models so that they can be easily evaluated in parallel using the standard BPF.[5] Using this BPF for the likelihood evaluation we can estimate the posterior mean of the conditional log-likelihood function in Eq. (2.23) by the arithmetic mean over the Gibbs draws of the parameters $\{\theta^{(i)}\}_{i=1}^{S}$, that is

$$\widehat{\mathrm{E}}_{post}\left[\log p(C^r_{1:T}|C^{fr}_{1:T}, C^f_{1:T}; \theta) \right] = \frac{1}{S} \sum_{i=1}^{S} \log p(C^r_{1:T}|C^{fr}_{1:T}, C^f_{1:T}; \theta^{(i)}). \tag{2.26}$$

2.3.3. Forecasting

Using the Gibbs sampler outlined in Section 2.3.1 for fixed values of the parameters θ, we can perform out-of-sample point- and density forecasting for the realized covariance matrix of asset returns C^r_{t+1}. A density forecast for C^r_{t+1} obtains as

$$p(C^r_{t+1}|C_{1:t}; \theta) = \int f(C^r_{t+1}|\Sigma_{t+1}; \theta)\ f(\Sigma_{t+1}|\Sigma_{1:t}; \theta)\ \pi(\Sigma_{1:t}|C_{1:t}; \theta)\ \mathrm{d}\Sigma_{1:t+1}, \tag{2.27}$$

where $\pi(\Sigma_{1:t}|C_{1:t}; \theta)$ denotes the posterior density of the state variables in $\Sigma_{1:t}$ for the observed data up to period t, and $f(C^r_{t+1}|\Sigma_{t+1}; \theta)$ is the marginal density for the realized covariance of the assets. Under the Wishart assumption for C_t, this marginal density for C^r_{t+1} is itself a Wishart density given by $f_{\mathcal{W}}(C^r_{t+1}|n, \Sigma^r_{t+1}/n)$ (Muirhead, 2005, Corollary 3.2.6). The forecasting density (2.27) evaluated at the ex-post observed value for C^r_{t+1} defines the period $t+1$ predictive likelihood. For its

[5] Implementation details for the BPF approximation of Eq. (2.24) are given in Appendix A.2.

computation we set the parameters θ equal to their posterior mean estimates based on the data observed until period t.

Note that the density forecast in Eq. (2.27) and its moments do not account for parameter estimation uncertainty. An alternative which would account for this uncertainty is to use the standard Bayesian predictive density (Geweke, 2005). While it can be straightforwardly implemented for the WFSS model using Gibbs draws from the joint posterior distribution of the state variables and the parameters, we decided to perform forecasting based on the density forecast in Eq. (2.27), because we shall compare the predictive performance of the WFSS model with that of alternative forecasting approaches for which it is not clear how to account for estimation uncertainty.

The predictive likelihood according to Eq. (2.27) can be approximated via MC integration, i.e.,

$$p(C^r_{t+1}|C_{1:t};\theta) \simeq \frac{1}{S}\sum_{i=1}^{S} f(C^r_{t+1}|\Sigma^{(i)}_{t+1};\theta), \tag{2.28}$$

where $\{\Sigma^{(i)}_{t+1}\}$ are simulated draws from the convolution

$$f(\Sigma_{t+1}|\Sigma_{1:t};\theta)\ \pi(\Sigma_{1:t}|C_{1:t};\theta)$$

based on Gibbs simulations from $\pi(\Sigma_{1:t}|C_{1:t};\theta)$. Using the simulated draws $\{\Sigma^{(i)}_{t+1}\}$ the point forecast of C^r_{t+1} given by $\mathrm{E}(C^r_{t+1}|C_{1:t};\theta) = \mathrm{E}(\Sigma^r_{t+1}|C_{1:t};\theta)$ can be approximated by

$$\mathrm{E}(C^r_{t+1}|C_{1:t};\theta) \simeq \frac{1}{S}\sum_{i=1}^{S} \Sigma^{r(i)}_{t+1}. \tag{2.29}$$

While the simulated draws $\{\Sigma^{(i)}_{t+1}\}$ required for the computation of the predictive likelihood in Eq. (2.28) can be straightforwardly obtained, the evaluation of the $(p \times p)$-dimensional measurement density $f(C^r_{t+1}|\Sigma^{(i)}_{t+1};\theta) = f_{\mathcal{W}}(C^r_{t+1}|n, \Sigma^{r(i)}_{t+1}/n)$ in high-dimensional applications turns out to be numerically very unstable. In fact,

when the Wishart measurement density is high-dimensional as in our application for $p = 60$ assets its evaluation suffers from frequent floating-point underflows (see Kastner, 2019, for a discussion of similar computational problems). Thus, instead of using the predictive likelihood for the covariance matrix C^r_{t+1} itself, we exploit the property of Wishart distributions of being closed under linear transformations and rely on the predictive likelihood for $\iota' C^r_{t+1} \iota$ where ι denotes a vector full of ones. This transformation can be interpreted as the return variance of an equally weighted portfolio. Its predictive likelihood obtains by replacing in Eqs. (2.27) and (2.28) the multivariate Wishart $f(C^r_{t+1}|\Sigma_{t+1}; \theta) = f_{\mathcal{W}}(C^r_{t+1}|n, \Sigma^r_{t+1}/n)$ by the corresponding univariate density $f(\iota' C^r_{t+1} \iota|\Sigma_{t+1}; \theta)$ which is a one-dimensional Wishart given by $f_{\mathcal{W}}(\iota' C^r_{t+1} \iota|n, \iota' \Sigma^r_{t+1} \iota/n)$ (Muirhead, 2005, Theorem 3.2.5).

2.4. Empirical Application

2.4.1. Data

We use the WFSS model to analyze the dynamics of the daily realized covariance matrix for 60 stocks traded at the New York Stock Exchange. The stocks are selected by liquidity from the S&P 500 index and sorted by their sector and industry classification according to the Global Industrial Classification Standard (GICS). The list of stocks covering six industry sectors is provided in Table 2.1. For observed risk factors we use the market, high-minus-low price-earnings ratio (HML) and small-minus-big market capitalization (SMB) factor in the Fama-French 3-factor model. Following Fan et al. (2016) and Aït-Sahalia and Xiu (2017), we additionally consider the sector-specific Spyder Exchange-Traded Funds (SPDR ETFs) for the six sectors covered by the 60 stocks: XLI (Industrials), XLY (Consumer Discretionary), XLP (Consumer Staples), XLV (Health Care), XLF (Financials), and XLK (Information Technologies). These SPDR ETFs are traded at very high-frequency and therefore can serve in our approach for constructing realized covariance measures of factors which track the largest S&P sectors.

The daily realized covariance matrices C_t are computed using the composite realized kernel method of Lunde et al. (2016) based on 5-minute returns for the Fama-French factors and 1-minute returns for the assets as well as the sector-specific ETFs.[6] The data comprises 2415 time series of realized variances and covariances for the sample period from January 3, 2007 to December 31, 2012, covering 1510 trading days. See Figure 2.2 for time-series plots of the realized variance for two randomly selected stocks (Citygroup and Caterpillar), the ETFs for the sectors of those two stocks (XLF and XLI), and the three Fama-French factors. Figure 2.2 also provides the sample autocorrelation functions (ACF) of the realized variance for the two selected stocks which indicate a very strong serial correlation. Such a strong serial correlation we find for all assets and risk factors.

Given the joint realized covariance matrices for the assets and the factors C_t we can compute according to Eqs. (2.1) and (2.4) the realized residual covariance matrices C_t^e which represent estimates for integrated residual covariance matrices Σ_t^e. In order to assess the realized residual sparsity we compute from C_t^e for various sets of factors the time average of the corresponding residual correlation matrices. Figure 2.3 shows heat plots of this average residual correlation matrix obtained from using no factors, the CAPM factor (market), the three Fama-French factors (market, HML, SMB), and the three Fama-French factor plus the six ETFs. These plots illustrate dense asset correlations ranging from 0.18 to 0.52 (no factors). We also observe that these correlations are significantly reduced when including risk factors. Under the 9-factor case (Fama-French plus ETFs) the resulting residual correlation matrix is fairly close to diagonal with a maximum absolute average correlation of 0.22. These results are consistent with those reported by Fan et al. (2016) and justify the use of a strict factor approximation to the integrated covariance matrix of the assets Σ_t^r, especially, when combined with a Fama-French plus ETF factor structure.

[6] We are grateful to Dacheng Xiu for kindly providing us with the intraday Fama-French factor return data. The intraday return data for the 60 stocks and 6 ETFs has been obtained from QuantQuote.com.

2.4.2. Estimation results

Using the complete sample covering 1510 trading days we estimate the WFSS model for the following three factor structures: A 1-factor model with the market factor in the CAPM (1F), a 3-factor model based on the Fama-French factors (3F), and a 9-factor model including the Fama-French factors plus the six sector-specific EFTs (9F). For each of the factor structures we take the unrestricted WFSS specification defined in Eqs. (2.3)–(2.13) and compare it to model specifications obtained under the restrictions that the factor loadings are time-invariant ($\beta_{ikt} = \beta_{ik}$) and/or the log (pseudo) variances x^e_{it} and x^f_{kt} have simple short-memory AR(1) dynamics ($\phi^e_{i2} = \phi^e_{i3} = 0$ and $\phi^f_{k2} = \phi^f_{k3} = 0$). In total we compare for each of the factor structures 1F, 3F and 9F four WFSS specifications: The unrestricted one with time-varying loadings and HAR variance dynamics (HAR-vβ), the WFSS with constant loadings and HAR variance dynamics (HAR-cβ), time-varying loadings and AR(1) variance dynamics (AR1-vβ), and constant loadings and AR(1) variance dynamics (AR1-cβ).

The prior assumptions we use for the parameters are fairly uninformative. For the $(\gamma, \phi_1, \phi_2, \phi_3, \nu^2)$ parameters in each of the Gaussian HAR processes $\{x^e_{it}\}$ and $\{x^f_{kt}\}$ we assume independent conjugate Normal-inverted Gamma priors with hyper-parameters selected such that $\mathrm{E}(\nu^2) = 0.2$ and $\mathrm{Var}(\nu^2) = 0.0156$, $\mathrm{E}(\phi_1, \phi_2, \phi_3) = (0.3, 0.3, 0.3)$ and $\mathrm{Cov}(\phi_1, \phi_2, \phi_3) = \mathrm{diag}(0.1, 0.1, 0.1)$, $\mathrm{E}(\gamma^*) = 0$ and $\mathrm{Var}(\gamma^*) = 20$, where $\gamma^* = \gamma(1 - \phi_1 - \phi_2 - \phi_3)$. Likewise we use for the (γ, ϕ, ν^2) parameters in each of the Gaussian AR(1) processes $\{\beta^e_{ikt}\}$ and $\{\ell^f_{kjt}\}$ independent conjugate Normal-inverted Gamma priors with $\mathrm{E}(\nu^2) = 0.2$ and $\mathrm{Var}(\nu^2) = 0.0156$, $\mathrm{E}(\phi) = 0.86$ and $\mathrm{Var}(\phi) = 0.1$, $\mathrm{E}(\gamma^*) = 0$ and $\mathrm{Var}(\gamma^*) = 20$, where $\gamma^* = \gamma(1 - \phi)$. This prior we also assume for the parameters of the Gaussian HAR processes $\{x^e_{it}\}$ and $\{x^f_{kt}\}$ under the AR(1) restriction. For the Wishart degrees of freedom n we select a discrete uniform prior on the interval $(p + q, 350]$ with 1000 equally distant grid points.

For parameter estimation we run the MCMC algorithm proposed in Section 2.3.1 for 15,000 iterations, where the first 5,000 are discarded. The PG-AS procedure in the MCMC update step for the nonlinear state-trajectories $\{x^f_{k,1:T}\}$ and $\{x^e_{i,1:T}\}$ is

implemented using 50 particles (see steps 1.) and 3.) of the MCMC algorithm). For the computation of the DIC criterion in Section 2.3.2 we run the BPF using 25,000 particles. The MCMC algorithm is implemented in MATLAB. The average CPU computing time per MCMC iteration ranges from 2.4 (1F-AR1-cβ) to 5.9 sec (9F-HAR-vβ) on an Intel i7 3.4GHz processor with 4 cores. In order to evaluate the sampling efficiency of the proposed MCMC procedure for estimating the parameters of the WFSS model we use the inefficiency factor. Following Kim et al. (1998), it is computed as IF $= 1 + 2[B/(B-1)] \sum_{j=1}^{B} K(j/B)\, \hat{\rho}_j$, where $\hat{\rho}_j$ denotes the lag-j sample autocorrelation of the MCMC draws of a parameter, $K(\cdot)$ is the Parzen kernel function, and B the bandwidth which we set equal to $B = 100$. The interpretation is that an MCMC sample with $S^* \times$ IF parameter draws is needed in order to obtain the same precision as a (hypothetical) iid sample from the posterior of size S^*, so that small IF-values are to be preferred. For the unrestricted full WFSS model (9F-HAR-vβ) the IF values across all of its parameters range from 1.5 to 17.36 with an average value of 3.44 indicating a high sampling efficiency with a fast mixing rate of our high-dimensional MCMC algorithm. In fact, it even achieves the efficiency level reported by Kim et al. (1998) for their MCMC auxiliary mixture sampling approach which is a well-established efficiency benchmark for simple univariate stochastic volatility models.

Table 2.2 contains the DIC values for the WFSS specifications. They show that of all factor structures the 9-factor one including the Fama-French factors and the ETFs yields the best trade-off between goodness of fit for the realized covariance matrix of the assets and parametric simplicity. This applies to all model specifications and is consistent with the results of our initial analysis showing that under the 9-factor structure the time average of the realized residual correlation matrices is closest to diagonal (see Section 2.4.2). Hence, these results indicate strong evidence against the 1-factor CAPM structure as used by Sheppard and Xu (2019) in their factor HEAVY model. We also find that the gain in terms of DIC values obtained from adding to a 3-factor Fama-French model the six sector-specific ETFs is substantially larger than

the gain we obtain when moving from a 1-factor CAPM model to a 3-factor structure by adding the two Fama-French factors HML and SMB. Hence, the sector-specific risk factors when considered together with the Fama-French factors appear to be extremely useful in explaining the daily variation of the assets' covariance matrix. As for the dynamic structure, we observe that WFSS models with dynamically varying factor loadings uniformly outperform those with constant loadings, and that WFSS models with long-memory type HAR dynamics for the variances strictly dominate their short-memory AR(1) counterparts. This evidence in favor of time-varying loadings is fully consistent with the results reported in Bollerslev and Zhang (2003), Kalnina (2015), Engle (2016) and Sheppard and Xu (2019). Our DIC results for the loadings are also particularly remarkable in that they reveal that the improvement in model fit achieved by allowing the loadings to be time-varying does not appear to be compromised by the involved significant inflation of parameters (e.g. 994 for the 9F-HAR-cβ versus 2,074 in the 9F-HAR-vβ). Of all 12 considered WFSS models, the DIC-preferred specification for the assets' covariance matrix is the unrestricted full 9F-HAR-vβ WFSS.

Figure 2.4 summarizes the parameter estimates for the DIC-preferred 9F-HAR-vβ model.[7] Figure 2.4 shows box-plots of the posterior mean values for all the autoregressive coefficients of the AR(1)-processes (ϕ) and HAR-processes (ϕ_1, ϕ_2, ϕ_3) directing the integrated (co)variances in (Σ_t^f, Σ_t^e) and factor loadings in $\mathcal{B}_t$ (see Eqs. 2.8–2.13). It also provides the posterior mean values for the maximal root of the restricted AR(22) representation of the HAR processes (z), as well as the posterior mean values of the stationary variance (V) of the AR(1) and HAR processes.[8] As for the maximal estimated roots of the HAR processes for the assets' residual variances (z^e) and the diagonals of the factor covariance matrix (z^f), we observe that they are close to unity implying stationarity though a very strong persistence. The estimates of the stationary variances V^β and AR(1) roots ϕ^β reveal that the assets' loadings on the risk factors typically exhibit a relatively small, yet significant variation and a mod-

[7] The corresponding posterior standard deviations are provided in Figure 2.5.

[8] For the AR(1) processes the stationary variance obtains as $V = \nu^2/(1-\phi^2)$ and for the HAR processes analogously from their restricted AR(22) representation in Eq. (2.14) (see Hamilton, 1994, p. 59).

erate persistence. This finding is in line with the DIC values in Table 2.2 preferring dynamically varying factor loadings over constant ones. In Figure 2.2 we plot the ACFs for the realized variance of Citigroup and Caterpillar predicted under the fitted 9F-HAR-vβ WFSS together with their observed sample counterparts. They illustrate that this model combining long-memory type HAR dynamics for the variances and short-memory AR(1) dynamics for the loadings is well able to account for the observed strongly persistent movements of the assets' realized variances. Similar results we obtain for all assets. Figure 2.6 displays the smoothed estimates of the total integrated variance (σ_t^r), the residual variance (σ_t^e) and the loadings on the Fama-French and sector-specific risk factor (β_t) for the Citigroup stock under the 9F-HAR-vβ WFSS together with the respective observed realized variances and loadings. The smoothed estimates are obtained as posterior mean values of the integrated variances and loadings computed from the MCMC posterior draws of the state variables. Unsurprisingly, we see that the smoothed estimates follow closely the variation in their realized counterparts.

In order to quantify the relative importance of the risk factors we provide in Figure 2.7 their estimated contributions to the assets' total variances under the HAR-vβ WFSS with 1 (market), 3 (market+HML+SMB) and 9 factors (market+HML+SMB+ ETFs), respectively. For asset i the estimated factor contribution under a given factor structure is computed as $1 - (1/T)\sum_{t=1}^{T} \mathrm{E}(\sigma_{it}^e|C_{1:T})/c_{it}^r$, where $\mathrm{E}(\sigma_{it}^e|C_{1:T})$ is the posterior mean of the integrated residual variance and c_{it}^r the total realized variance. The explained contribution of the market risk under the 1-factor CAPM is quite substantial and varies between 18% and 35%. Combining the market risk with the HML and SMB factor only moderately increases the explained contribution to a range between 23% and 37%, while a significant increase in that contribution up to a level of 51% is obtained by adding the ETFs to the market, HML and SMB. This corroborates our previous results on the relative importance of the HML, SMB and sector-specific risk factors.

2.4.3. Out-of-sample forecasting results

We now analyze the out-of-sample forecasting performance of the WFSS for the realized covariance matrix of asset returns C_t^r and compare it to alternative forecasting approaches. As alternative approaches we consider the Exponentially Weighted Moving Average (EWMA) model (Morgan, 1996) and the scalar Realized consistent Dynamic Conditional Correlation (sRe-cDCC) model of Bauwens et al. (2016). The EWMA is popular in industry practice while the sRe-cDCC model is found by Bauwens et al. (2016) to be highly effective in predicting large covariance matrices of asset returns outperforming alternative popular high-dimensional forecasting models.

The EWMA is given by

$$\mathrm{E}(C_{t+1}^r \mid C_{1:t}^r) = (1-\lambda)C_t^r + \lambda \mathrm{E}(C_t^r \mid C_{1:t-1}^r), \tag{2.30}$$

where we set the smoothing parameter λ to its typically selected value of 0.96 (Callot et al., 2017). The sRe-cDCC model assumes for C_t^r given $C_{1:t-1}^r$ a Wishart distribution, $C_t^r \mid C_{1:t-1}^r \sim \mathcal{W}_p(n, S_t/n)$, where the scale matrix is factorized as

$$S_t = AV_tR_tV_tA', \tag{2.31}$$

with A being the lower triangular Cholesky factor of $\mathrm{E}(C_t^r) = \mathrm{E}(S_t)$. The matrix V_t is defined as $V_t = \mathrm{diag}(\sqrt{v}_{1t}, \ldots, \sqrt{v}_{pt})$, where v_{it} is the conditional variance of asset i, and R_t is the conditional correlation matrix obtained as $R_t = (Q_t \odot I_p)^{-1/2} Q_t (Q_t \odot I_p)^{-1/2}$, where $\odot$ denotes the Hadamard element-by-element product. Both the v_{it}'s and Q_t are endowed with GARCH-type recursions of the form

$$v_{it} = (1-\kappa_i-\delta_i) + \kappa_i c_{it-1}^* + \delta_i v_{it-1}, \quad i = 1, \ldots, p, \tag{2.32}$$

$$Q_t = (1-\alpha-\varphi)I_p + \alpha C_{t-1}^Q + \varphi Q_{t-1}, \tag{2.33}$$

where c_{it}^* is the i'th diagonal element of $C_t^* = A^{-1}C_t^r A^{-1\prime}$ and $C_t^Q = (Q_t \odot I_p)^{1/2} V_t^{-1} C_t^* V_t^{-1} (Q_t \odot I_p)^{1/2}$, with I_p denoting the p-dimensional identity matrix. The model

parameters are given by the scalars $\{\kappa_i\}$, $\{\delta_i\}$, α, φ and are estimated by the three-step Quasi-ML (QML) procedure proposed by Bauwens et al. (2016).

In our forecast experiments we focus on 1-day-ahead predictions. They are obtained by re-estimating the model parameters every 10 trading days on a rolling 4-year window with 1008 daily observations and then producing a sequence of new 1-day-ahead forecasts based on the updated parameter estimates. We consider two out-of-sample forecasting periods each covering one year with 251 trading days (see the shaded areas in the time series plots in Figure 2.2). The first period is the year 2011 where the volatility ranges from a small to a relatively high level triggered by the August 11 stock market fall after the credit rating downgrade of US sovereign debt. The second period covers the year 2012 with a constantly fairly low volatility.

Statistical forecast evaluation

For statistical forecast evaluation we rely on the predictive likelihood and the accuracy of the point forecasts for the covariance matrix (see Section 3.3).

To assess the point forecast accuracy we follow Ledoit et al. (2003) and use the root-mean-squared-error (RMSE) based on the Frobenius norm comparing the covariance matrix forecast $\widehat{C}_t^r$ and the ex-post observed value for C_t^r. This RMSE is given by

$$\text{RMSE} = \frac{1}{T^\star}\sum_t ||C_t^r - \widehat{C}_t^r|| = \frac{1}{T^\star}\sum_t \left[\sum_i (c_{it}^r - \hat{c}_{it}^r)^2 + 2\sum_{i<j}(c_{ijt}^r - \hat{c}_{ijt}^r)^2\right]^{\frac{1}{2}}, \quad (2.34)$$

where c_{it}^r and c_{ijt}^r denote the realized variance of asset i and the realized covariance between asset i and j, respectively, and $\hat{c}_{it}^r$ and $\hat{c}_{ijt}^r$ their forecasts.[9] $T^\star$ is the number of forecast periods. In order to disentangle the forecast performance w.r.t. the different elements in the covariance matrix we also compute the RMSE separately for the

[9] Alternative frequently applied loss functions are the Stein loss, the QLIKE and the von Neumann divergence (Bauwens et al., 2014). However, for parameter-driven state-space models as the WFSS these loss functions are not guaranteed to provide consistent rankings of the forecasting performance (Laurent et al., 2013).

variances and covariances, i.e.,

$$\text{RMSE}^{\text{v}} = \frac{1}{T^\star}\sum_t \left[\sum_i (c_{it}^r - \hat{c}_{it}^r)^2\right]^{\frac{1}{2}}, \quad \text{RMSE}^{\text{c}} = \frac{1}{T^\star}\sum_t \left[\sum_{i<j} (c_{ijt}^r - \hat{c}_{ijt}^r)^2\right]^{\frac{1}{2}}. \quad (2.35)$$

For assessing the significance of differences in the RMSE, RMSE^{v} and RMSE^{c} losses across models we rely on the model confidence set (MCS) approach of Hansen et al. (2011). The MCS identifies the model or set of models having the best forecasting performance at a given confidence level. The MSC is computed for confidence levels of 75% and 90% using a block bootstrap with block length $\lfloor (T^\star)^{1/3} \rfloor$ and 10,000 bootstrap replications.

The RMSE results for the out-of-sample periods 2011 and 2012 as well as for the two periods aggregated together are summarized in Table 2.3. They reveal that for forecasting the assets' covariance matrix based on the WFSS approach it is important to make use of all the 9 available risk factors and to account for time-variation in the factor loadings. In fact, for the fairly volatile period 2011 it is the 9-factor WFSS with time-varying loadings and AR(1) variance dynamics (9F-AR1-vβ) which exhibits among all WFSS models the smallest RMSE loss and for the low-volatility period 2012, the 9-factor WFSS with time-varying loadings and HAR variance dynamics (9F-HAR-vβ). The latter is also the best performing WFSS for the aggregated period 2011–2012. Comparing the RMSE losses with RMSE^{v} and RMSE^{c} values indicates that the comparatively good predictive performance of those two 9-factor WFSS models with time-varying loadings is due to their ability to produce relatively precise forecasts, especially for the covariances. However, if it is the forecast performance only for the variances, the best WFSS is the 9-factor model with constant loadings and HAR dynamics (9F-HAR-cβ). As for the comparison of the WFSS approach with the competing alternatives, we find that for the periods 2011 and 2011–2012 the RMSE loss of the best-performing WFSS is at the 75% MCS-confidence level significantly smaller than that of the EWMA but is on par with the loss of the sRe-cDCC model. For the 2012 period both, the EWMA and sRe-cDCC are at the 75% level significantly

outperformed by the best WFSS. Recall that for the sake of comparison with the (non-Bayesian) EWMA and sRe-cDCC, the point forecasts $\mathrm{E}(C_t^r|C_{1:t-1};\theta)$ used in this experiment for the (Bayesian) WFSS approach do not account for the uncertainty about the parameters. In the Technical details Section A.4 we provide RMSE results for WFSS forecasts including this uncertainty, for which we use the mean under the Bayesian predictive density $\mathrm{E}(C_t^r|C_{1:t-1})$. They show that parameter uncertainty does not materially affect forecast accuracy which is fully consistent with the small posterior standard deviations we observe for the parameters (see Figure 2.5).

Table 2.4 reports the sum of the 1-day-ahead log-predictive likelihoods for the variance of an equally weighted portfolio $\sum_{t=1}^{T^\star} \log p(\iota' C_{t+1}^r \iota|C_{1:t},\theta)$ for the two out-of-sample periods, and Figure 2.8 shows the time-series plots of the period-wise accumulated log-predictive likelihoods for the entire range of the two out-of-sample periods. As the EWMA in Eq. (2.30) lacks a distributional assumption its predictive likelihoods are computed under the assumption of a conditional Wishart distribution for the realized covariance matrix given its prediction with degrees of freedom estimated by ML. The results show that the out-of-sample fit of all WFSS specifications is substantially better than that of the EWMA and sRe-cDCC. The biggest gain in predictive fit of the WFSS models relative to the EWMA and sRe-cDCC we observe for the turbulent trading days in the aftermath of the August 2011 stock market fall (see Figure 2.8). The WFSS models with the largest log-predictive likelihoods are the same as those with the lowest RMSE loss for the point forecasts: For the volatile period 2011 it is the 9F-AR1-vβ and for the low-volatility period 2012 the 9F-HAR-vβ (see Table 2.4). Hence, overall our WFSS approach with 9 factors and dynamically varying loadings performs favorably in relation to the competing models.

Value-at-risk forecasts

For an economic evaluation of predictive performance we consider Value-at-Risk (VaR) forecasts for portfolios constructed from the p assets. For simplicity, we focus on an equally-weighted portfolio with return $\omega' r_t$, where r_t denotes the p-dimensional

vector of the period-t asset returns and ω is the vector of equal portfolio weights. Assuming for the 1-period-ahead predictive distribution of the portfolio returns a normal $\mathcal{N}(0, \omega' \widehat{C}_t^r \omega)$, the predicted period-$t$ portfolio VaR at level α obtains as

$$\mathrm{VaR}_t(\alpha) = z_\alpha \sqrt{\omega' \widehat{C}_t^r \omega}, \tag{2.36}$$

where z_α is the $\alpha\%$ quantile of a standard normal distribution. The normal distribution for the portfolio returns can also be replaced by a more flexible student-t. However, the results we obtained under the student-t (not reported here) do not qualitatively differ from those for the normal distribution.

For assessing the accuracy of the predicted VaR we follow Chib et al. (2006) and test for unconditional and conditional coverage based on the 'hit-indicator' variable $I_t = \mathbb{1}[\omega' r_t \leq \mathrm{VaR}_t(\alpha)]$ for $t = 1, \ldots, T^\star$, signaling that the realized portfolio return is lower or equal than the predicted VaR. The hypothesis of correct unconditional coverage can be tested with the likelihood-ratio statistic (Kupiec, 1995)

$$\mathrm{LR}_{UC} = 2\left\{\log[\hat{\alpha}^{\hat{\alpha}T^\star}(1-\hat{\alpha})^{(1-\hat{\alpha})T^*}] - \log[\alpha^{\hat{\alpha}T^\star}(1-\alpha)^{(1-\hat{\alpha})T^*}]\right\}, \tag{2.37}$$

where $\hat{\alpha}$ is the hit-rate defined as $\hat{\alpha} = \sum_{t=1}^{T^\star} I_t / T^\star$. Under the hypothesis that the observed hit-rate $\hat{\alpha}$ is equal to the nominal level α (correct unconditional coverage) the statistic LR_{UC} is distributed asymptotically as a $\chi^2_{(1)}$. Conditional coverage can be tested by jointly testing for unconditional coverage and serial independence of the hit-indicator sequence $\{I_t\}$ with the statistic (Christoffersen, 1998)

$$\mathrm{LR}_{CC} = 2\left\{\log[(1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}] - \log[\alpha^{\hat{\alpha}T^\star}(1-\alpha)^{(1-\hat{\alpha})T^*}]\right\}, \tag{2.38}$$

where T_{ij} denotes the number of cases for which we observe $I_t = j$ and $I_{t-1} = i$ for $i, j \in \{0, 1\}$ while $\pi_{01} = T_{01}/(T_{01} + T_{00})$ and $\pi_{11} = T_{11}/(T_{10} + T_{11})$. Under the joint hypothesis of correct unconditional coverage and serial independence LR_{CC} is distributed asymptotically as a $\chi^2_{(2)}$.

The p-values of the coverage tests for the predicted 1% and 5% VaR are shown in Table 2.5. They reinforce our earlier results on the point-forecast accuracy and predictive fit for the covariance matrix of the assets. The best performing WFSS models for both the 1% and 5% VaR are those with all 9 risk factors and time-varying loadings (9F-AR1-vβ and 9F-HAR-vβ). For the year 2011 as well as the year 2012 they pass the tests for unconditional and conditional coverage at the 1% significance level. However, for the aggregated period 2011-2012 both models fail to pass the unconditional coverage test for the 1% VaR as they significantly underestimate the VaR, but note that this applies to all models. Overall, the performance of the 9-factor WFSS with time varying loadings is better than that of the EWMA and not worse than that of sRe-cDCC model.

Global-minimum-variance-portfolio forecasts

As a further economic experiment designed to evaluate the forecasting performance we use the predicted covariance matrices of the assets to construct optimal investment portfolios (Bauwens et al., 2016; Callot et al., 2017). For portfolio allocation we consider a strategy based on the global-minimum-variance portfolio (GMVP) which has the advantage relative to minimum-variance portfolios that its ex-ante portfolio weights only depend on the covariance matrix. When constructing the GMVP we exclude short-selling by imposing the portfolio weights to be non-negative. This is rather typical in high-dimensional portfolio allocation problems and acts like a regularization device for reducing the impact of errors in covariance matrix forecasts on the allocation (Frost and Savarino, 1986; Jagannathan and Ma, 2003; and Bauwens et al., 2016).

For a given covariance matrix forecast $\widehat{C}_t^r$ computed in period $t-1$, the GMVP weights $\widehat{w}_t$ obtain by solving the minimization problem

$$\widehat{w}_t = \arg\min_{w_t} w_t' \widehat{C}_t^r w_t, \quad \text{subject to} \quad \sum_{i=1}^{p} w_{it} = 1, \quad w_{it} > 0 \ \forall i, \tag{2.39}$$

where $w_t = (w_{1t}, \ldots, w_{pt})'$ is the vector of period-t portfolio weights to be selected in period $t-1$.

For assessing the relative capabilities of the competing models for optimal portfolio allocation we use the following four measures:

(i) The average out-of-sample portfolio return defined as $\hat{\mu}_p = \sum_{t=1}^{T^\star} r_{pt}/T^\star$ with $r_{pt} = \hat{w}_t' r_t$;

(ii) The accumulated portfolio return over the out-of-sample window $\hat{\mu}_p^{acc} = \left[\prod_{t=1}^{T^\star}(1 + r_{pt})\right] - 1$;

(iii) The portfolio return standard deviation over the out-of-sample window $\hat{\sigma}_p = \sqrt{\sum_{t=1}^{T^\star}(r_{pt} - \hat{\mu}_p)^2/T^\star}$;

(iv) The Sharpe ratio $\mathrm{SR}_p = \hat{\mu}_p/\hat{\sigma}_p$.

The results are summarized in Table 2.6. As expected, for the low-volatility period 2012 the portfolio return standard deviations are uniformly smaller across all forecasting models than for the more turbulent year 2011. On the other hand, all models generate for 2012 also a lower average and accumulated return than for 2011. As for the comparison of the competing models, we find that the EWMA and sRe-cDCC are outperformed in terms of all measures by the respective best WFSS. In terms of minimal portfolio risk, the 9-factor WFSS HAR model with constant factor loadings (9F-HAR-cβ) shows across all periods the best performance, though the differences to the other models are typically not very large. The highest average and accumulated return as well as the largest Sharpe ratio are obtained by portfolio allocations based on WFSS models with time-varying loadings: For period 2011 it is the 3-factor WFSS HAR and for 2012 the 1-factor WFSS AR(1) model.

2.5. Conclusion

In this paper we propose a Wishart factor state-space (WFSS) model for high–dimensional realized covariance matrices of asset returns. The model assumes a strict

factor structure for the latent integrated covariance matrix of the assets and relies on observed risk factors such as the 3 Fama-French factors. The WFSS is based on a Wishart measurement density for the joint realized covariance measure of the asset- and factor returns and assumes latent autoregressive processes for the factor- and idiosyncratic (co)variances as well as the factor loadings. An important advantage of the Wishart density is that it admits a convenient likelihood factorization, which results in fast, scalable and numerically precise one-step inference using simple Bayesian MCMC techniques.

We apply the factor model to a data set of daily realized covariance matrices for 60 NYSE traded stocks. The observed risk factors we consider are the CAPM market factor, the Fama-French HML and SMB factors as well as sector-specific ETFs. It turns out that for justifying a strict factor structure with uncorrelated residual components it is critical to include all those factors. The fitted factor models feature up to 2,074 parameters, which are estimated with high numerical precision. A Bayesian model comparison based on the Deviance Information Criterion (DIC) furthermore indicates the importance of accounting for time-varying loadings and persistent HAR dynamics for the integrated variances.

The in-sample findings are confirmed by an out-of-sample forecasting experiment including both statistical and economic evaluation criteria: Our results show that for practical forecasting applications it is important to make use of the complete set of risk factors and to account for time-variation in the factor loadings. The WFSS model shows good forecasting performance and outperforms its competitors in nearly all dimensions. We conclude that the WFSS approach is a valuable tool for modeling and forecasting time-series of high-dimensional realized covariance matrices.

2.6. Tables and Figures

Table 2.1.: List of the stocks included in the data set.

Symbol	Company	Sector	Symbol	Company	Sector
BA	Boeing Company	I	MDT	Medtronic Inc.	H
CAT	Caterpillar, Inc.	I	BAC	Bank of America Corporation	F
DE	Deere & Company	I	C	Citigroup Inc.	F
EMR	Emerson Electric Company	I	FITB	Fifth Third Bankcorp	F
GE	General Electric Company	I	JPM	J P Morgan Chase & Co.	F
GLW	Corning Incorporated	I	AXP	American Express Company	F
HON	Honeywell International Inc.	I	GS	Goldman Sachs Group, Inc.	F
LB	la Barge Inc.	I	MS	Morgan Stanley	F
LMT	Lockheed Martin Corporation	I	ADBE	Adobe Systems Incorporated	T
AMZN	Amazon.com, Inc.	D	ADP	Automatic Data Processing	T
BBBY	Bed Bath & Beyond Inc.	D	CTXS	Citrix Systems, Inc.	T
BBY	Best Buy Co., Inc.	D	ES	Electronic Arts, Inc.	T
GPS	Gap, Inc.	D	EBAY	Ebay, Inc.	T
HD	Home Depot, Inc.	D	EMC	EMC Corporation MA	T
JCP	J.C. Penney Company	D	IBM	International Business Machines	T
KSS	Kohl's Corporation	D	INTU	Intuit, Inc.	T
LOW	Lowe's Companies, Inc.	D	MSFT	Microsoft Corporation	T
COST	Costco Wholesale Corporation	S	NTAP	NetApp, Inc.	T
CVS	CVS Caremark Group	S	ORCL	Oracle Corporation	T
KR	Kroger Company	S	ADI	Analog Devices, Inc.	T
CAG	ConAgra, Inc.	S	ALTR	Altera Corporation	T
KO	Coca-Cola Company	S	AMAT	Applied Materials, Inc.	T
MO	Altria Group	S	AMD	Advanced Micro Devices, Inc.	T
PEP	Pepsico, Inc.	S	BRCM	Broadcom Corporation	T
A	Agilent Technologies Inc.	H	INTC	Intel Corporation	T
ABT	Abbott Laboratories	H	KLAC	KLA-Tencor Corporation	T
BAX	Baxter International Inc.	H	LLTC	Linear Technology Corporation	T
BMY	Bristol-Myers Squibb Company	H	MU	Micron Technology, Inc.	T
CAH	Cardinal Health Inc.	H	NVDA	NVIDIA Corporation	T
JNJ	Johnson & Johnson	H	QCOM	QUALCOMM Incorporated	T

Note: Stocks are selected by liquidity from the S&P 500 index and sorted by their sector and industry classification according to the Global Industrial Classification Standard (GICS). The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies.

Table 2.2.: Deviance information criteria (DIC) for the WFSS model specifications.

WFSS model	# params	# obs./param	DIC
1F-AR1-cβ	244	11,703	-2,860,015
3F-AR1-cβ	379	8,032	-3,414,454
9F-AR1-cβ	856	4,260	-6,030,216
1F-AR1-vβ	364	7,845	-2,868,559
3F-AR1-vβ	739	4,119	-3,469,892
9F-AR1-vβ	1,936	1,884	-6,266,725
1F-HAR-cβ	366	7,802	-2,874,627
3F-HAR-cβ	505	6,028	-3,428,088
9F-HAR-cβ	994	3,669	-6,041,832
1F-HAR-vβ	486	5,875	-2,882,791
3F-HAR-vβ	865	3,519	-3,483,658
9F-HAR-vβ	2,074	1,758	-6,279,064

Note: The table reports the Deviance information criteria (DIC) for the WFSS model specifications. The likelihood for the DICs given in Eq. (2.24) is computed using a bootstrap particle filter with $25,000$ particles. # params is the number of parameters and # obs/param the number of observations per parameter.

Table 2.3.: Evaluation of point-forecast accuracy.

	Period 2011			Period 2012			Period 2011 – 2012		
Model	RMSE	RMSE^{v}	RMSE^{c}	RMSE	RMSE^{v}	RMSE^{c}	RMSE	RMSE^{v}	RMSE^{c}
WFSS									
1F-AR1-cβ	56.98	18.28	37.02	25.89	12.62	15.41	41.50	15.46	26.26
3F-AR1-cβ	58.57	18.23	38.15	26.06	12.54	15.55	42.38	15.40	26.90
9F-AR1-cβ	52.33	17.82	33.93	25.08	12.11	15.06	38.76	14.97	24.53
1F-AR1-vβ	52.41	19.10	33.42	25.17	13.03	14.65	38.84	16.08	24.08
3F-AR1-vβ	53.11	18.67	34.01	25.23	12.83	14.76	39.22	15.76	24.42
9F-AR1-vβ	**50.58**	18.39	**32.51**	25.14	12.55	14.95	37.91	15.48	23.77
1F-HAR-cβ	58.35	16.77	38.69	25.37	10.67	15.80	41.14	13.59	26.74
3F-HAR-cβ	59.18	16.73	39.30	25.52	10.66	15.91	41.62	13.56	27.09
9F-HAR-cβ	53.46	**16.68**	35.22	23.84	**10.54**	14.71	38.00	**13.48**	24.52
1F-HAR-vβ	53.62	17.14	35.09	24.09	10.82	14.75	38.21	13.84	24.47
3F-HAR-vβ	54.36	16.97	35.65	24.61	10.75	15.17	38.83	13.72	24.96
9F-HAR-vβ	51.24	16.87	33.54	**23.02**	10.63	**14.03**	**36.51**	13.61	**23.36**
EWMA	57.78	18.64	38.02	23.15	10.81	14.14	41.62	14.88	26.93
sRe-cDCC	53.13	17.92	34.60	24.94	11.27	15.43	39.12	14.60	25.07

Note: The table reports the RMSE, RMSE^{v}, and RMSE^{c} losses as given in Eqs. (2.34) and (2.35). Bold figures indicate the smallest loss; Light grey-shaded cells indicate that the respective model is in the 90% model confidence set, dark grey-shaded cells indicate that the respective model belongs to the 75% model confidence set.

Table 2.4.: Evaluation of density forecasts.

Model	Period 2011	Period 2012	Period 2011 – 2012
WFSS			
1F-AR1-cβ	-2291.7	-2092.5	-4384.2
3F-AR1-cβ	-2306.3	-2096.7	-4403.1
9F-AR1-cβ	-2203.3	-2034.8	-4238.1
1F-AR1-vβ	-2247.7	-2066.7	-4314.4
3F-AR1-vβ	-2262.8	-2084.4	-4347.2
9F-AR1-vβ	**-2185.8**	-2026.4	-4212.2
1F-HAR-cβ	-2325.6	-2126.5	-4452.2
3F-HAR-cβ	-2333.0	-2128.7	-4461.7
9F-HAR-cβ	-2214.5	-2035.5	-4250.0
1F-HAR-vβ	-2269.4	-2083.8	-4353.2
3F-HAR-vβ	-2296.2	-2122.7	-4419.0
9F-HAR-vβ	-2191.9	**-2018.8**	**-4210.7**
EWMA	-5348.9	-3785.6	-9134.5
sRe-cDCC	-4297.3	-3778.0	-8075.2

Note: The table reports the sum of the log-predictive likelihoods for the return variance of an equally weighted portfolio $\iota' C^r_{t+1} \iota$. Bold figures indicate the largest value for the sum of the log-predictive likelihoods.

Table 2.5.: VaR forecasting results.

	Period 2011				Period 2012				Period 2011 – 2012			
	LR_{UC}		LR_{CC}		LR_{UC}		LR_{CC}		LR_{UC}		LR_{CC}	
Model	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
WFSS												
1F-AR1-cβ	0.00	**0.05**	0.00	**0.02**	0.00	**0.01**	0.00	**0.01**	0.00	0.00	0.00	0.00
3F-AR1-cβ	0.00	0.00	0.00	**0.01**	0.00	0.00	0.00	**0.01**	0.00	0.00	0.00	0.00
9F-AR1-cβ	**0.01**	**0.14**	**0.02**	**0.08**	**0.02**	**0.21**	**0.05**	**0.13**	0.00	**0.03**	0.00	**0.05**
1F-AR1-vβ	0.00	**0.14**	0.00	**0.08**	0.00	**0.01**	0.00	**0.03**	0.00	0.00	0.00	**0.02**
3F-AR1-vβ	0.00	**0.05**	0.00	**0.02**	0.00	0.00	0.00	**0.01**	0.00	0.00	0.00	0.00
9F-AR1-vβ	**0.02**	**0.49**	**0.05**	**0.30**	**0.06**	**0.32**	**0.14**	**0.61**	0.00	**0.24**	**0.01**	**0.37**
1F-HAR-cβ	0.00	**0.01**	0.00	**0.02**	0.00	**0.01**	0.00	**0.01**	0.00	0.00	0.00	0.00
3F-HAR-cβ	0.00	**0.01**	0.00	**0.01**	0.00	**0.01**	0.00	**0.01**	0.00	0.00	0.00	0.00
9F-HAR-cβ	**0.01**	**0.02**	**0.02**	**0.08**	**0.01**	**0.02**	**0.02**	**0.02**	0.00	0.00	0.00	0.00
1F-HAR-vβ	0.00	**0.02**	0.00	**0.08**	**0.01**	**0.02**	**0.02**	**0.02**	0.00	0.00	0.00	0.00
3F-HAR-vβ	0.00	**0.01**	0.00	**0.02**	0.00	**0.01**	0.00	**0.01**	0.00	0.00	0.00	0.00
9F-HAR-vβ	**0.01**	**0.14**	**0.02**	**0.27**	**0.02**	**0.08**	**0.05**	**0.03**	0.00	**0.01**	0.00	**0.01**
EWMA	0.00	**0.22**	0.00	**0.03**	**0.02**	**0.32**	**0.05**	**0.13**	0.00	**0.12**	0.00	0.00
sRe-cDCC	**0.06**	**0.49**	**0.15**	**0.78**	**0.02**	**0.67**	**0.05**	**0.55**	0.00	**0.55**	0.00	**0.46**

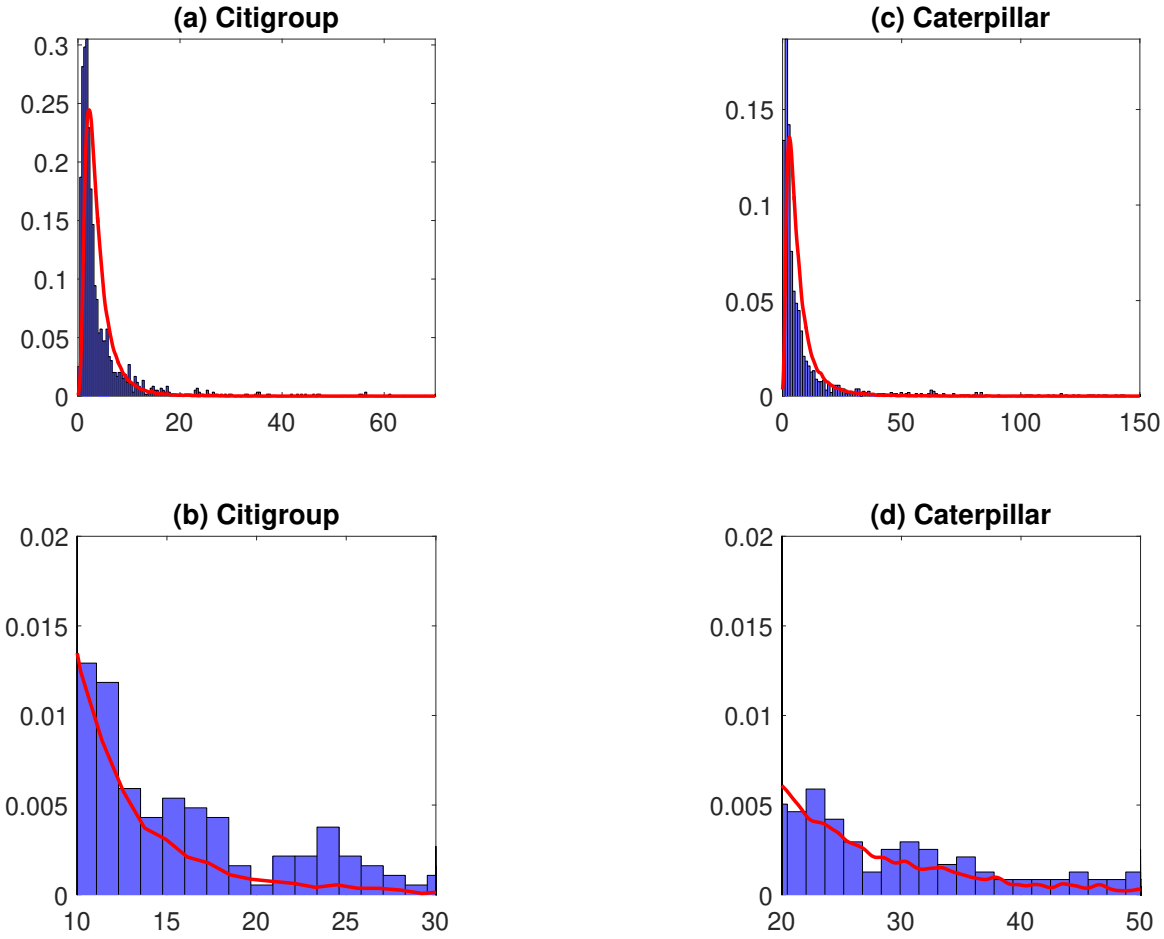
Note: The table reports the p-values for the unconditional coverage test LR_{UC} and conditional coverage test LR_{CC} given in Eqs. (2.37) and (2.38) for VaR predictions at the 1% and 5% level. Bold figures indicate that the null hypothesis cannot be rejected at the 1% significance level.

Table 2.6.: GMVP forecasting results.

	Period 2011				Period 2012				Period 2011 – 2012			
Model	$\hat{\sigma}_p$ $\times 10^3$	$\hat{\mu}_p$ $\times 10^2$	SR_p $\times 10^2$	$\hat{\mu}_p^{acc}$	$\hat{\sigma}_p$ $\times 10^3$	$\hat{\mu}_p$ $\times 10^2$	SR_p $\times 10^2$	$\hat{\mu}_p^{acc}$	$\hat{\sigma}_p$ $\times 10^3$	$\hat{\mu}_p$ $\times 10^2$	SR_p $\times 10^2$	$\hat{\mu}_p^{acc}$
WFSS-												
1F-AR1-cβ	6.78	4.17	6.15	9.91	4.22	2.79	6.62	7.34	5.59	3.45	6.17	18.0
3F-AR1-cβ	6.75	4.45	6.59	10.7	4.21	2.76	6.54	7.24	5.57	3.57	6.41	18.7
9F-AR1-cβ	6.56	4.58	6.99	11.1	4.20	2.62	6.24	6.85	5.45	3.56	6.53	18.7
1F-AR1-vβ	6.66	4.48	6.72	10.8	4.29	**3.51**	**8.18**	**9.37**	5.55	3.97	**7.16**	21.1
3F-AR1-vβ	6.67	4.44	6.65	10.6	4.29	3.42	7.97	9.12	5.55	3.91	7.04	20.7
9F-AR1-vβ	6.79	4.51	6.64	10.8	4.32	3.50	8.11	9.34	5.63	**3.98**	7.07	**21.2**
1F-HAR-cβ	6.77	3.89	5.75	9.19	4.22	2.81	6.66	7.39	5.58	3.33	5.96	17.3
3F-HAR-cβ	6.73	4.16	6.19	9.91	4.20	2.71	6.43	7.10	5.55	3.40	6.13	17.7
9F-HAR-cβ	**6.55**	4.15	6.33	9.90	**4.19**	2.60	6.20	6.80	**5.44**	3.34	6.14	17.4
1F-HAR-vβ	6.56	4.33	6.61	10.4	4.25	3.19	7.51	8.46	5.47	3.74	6.83	19.7
3F-HAR-vβ	6.61	**4.65**	**7.02**	**11.2**	4.27	3.16	7.41	8.38	5.51	3.87	7.03	20.5
9F-HAR-vβ	6.63	4.56	6.88	11.0	4.25	2.86	6.74	7.54	5.51	3.67	6.67	19.3
EWMA	6.72	3.32	4.94	7.72	4.25	2.10	4.93	5.40	5.57	2.68	4.82	13.5
sRe-cDCC	6.70	4.33	6.47	10.4	4.22	2.18	5.16	5.63	5.54	3.21	5.79	16.6

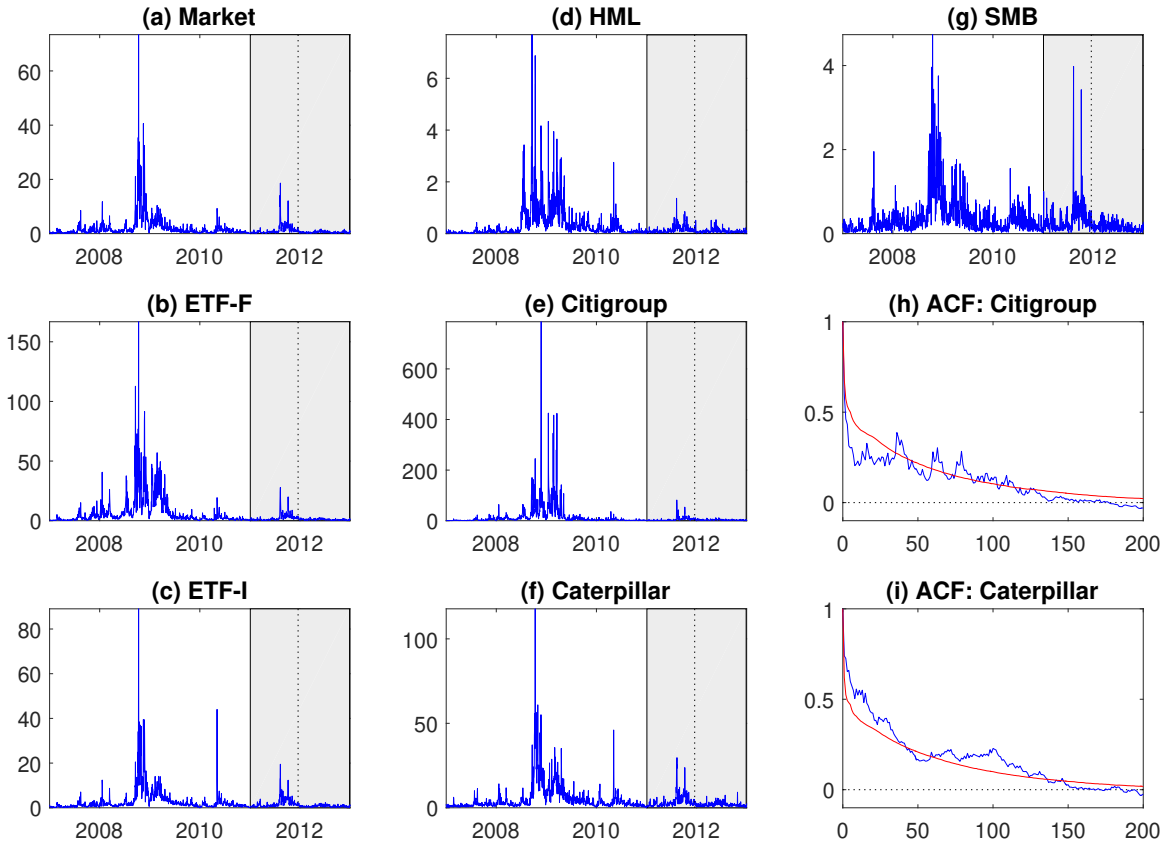
Note: The table reports GMVP summary statistics for the three out-of-sample periods under consideration. $\hat{\sigma}_p$ and $\hat{\mu}_p$ denote the portfolio standard deviation and mean return, SR_p is the Sharpe ratio and μ_p^{acc} denotes the accumulated portfolio return. Bold figures indicate the best models.

Figure 2.1.: Unconditional distribution of two realized asset variances predicted under the fitted WFSS model.



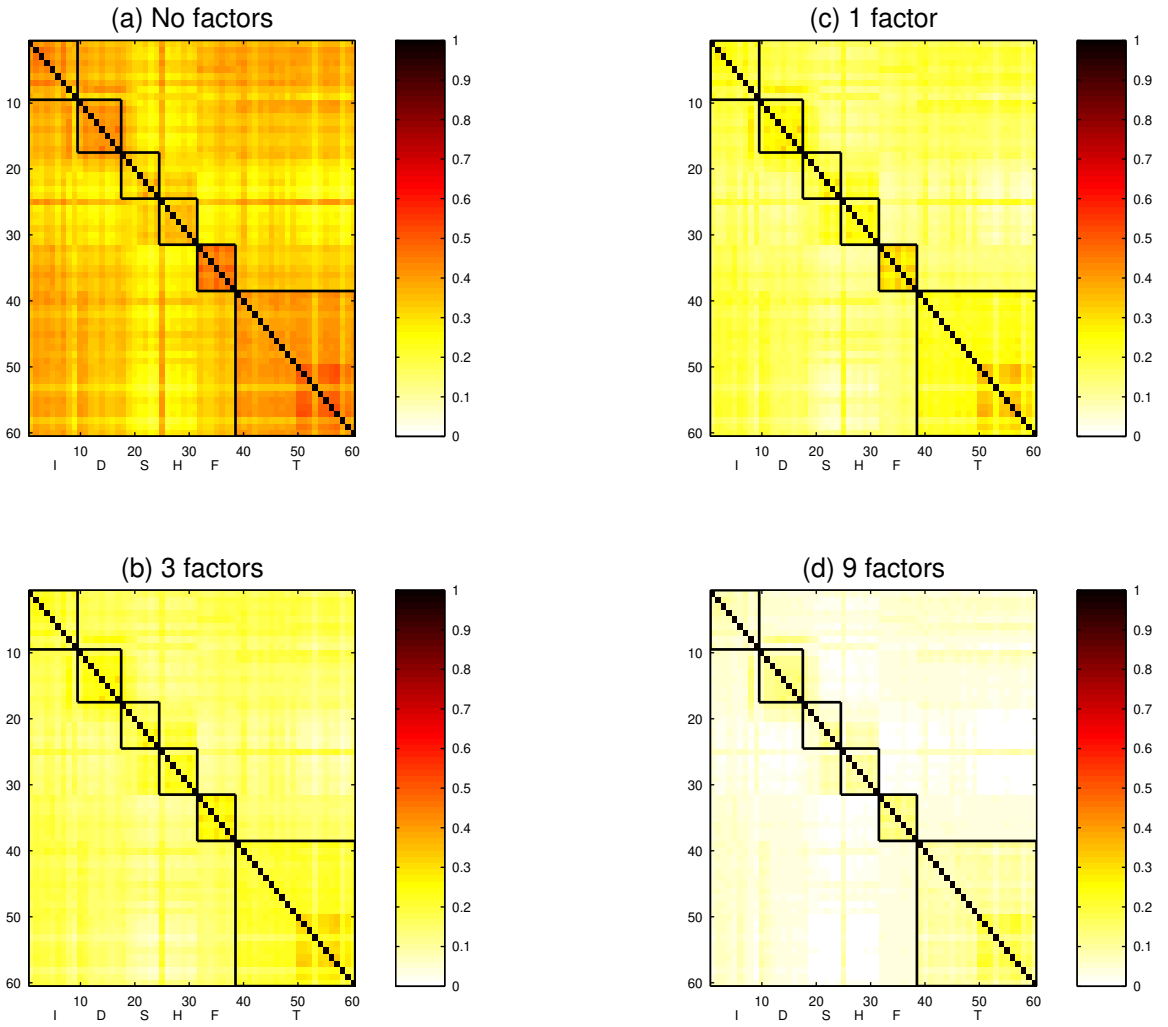
Panels (a) and (c): Histograms of the realized variances for the Citigroup and Caterpillar stocks and their unconditional distribution predicted under the fitted WFSS model (9F-HAR-vβ) defined by Eqs. (2.7)–(2.13) for 60 assets using 9 factors (red solid line). The predicted distributions are obtained from kernel density estimates based on simulated data from the fitted WFSS model for $T = 50,000$ time periods. Panels (b) and (d): Section of the histograms in panels (a) and (c) for the ranges $[10, 30]$ and $[20, 50]$, respectively.

Figure 2.2.: Time-series plots of selected realized factor and asset variances.



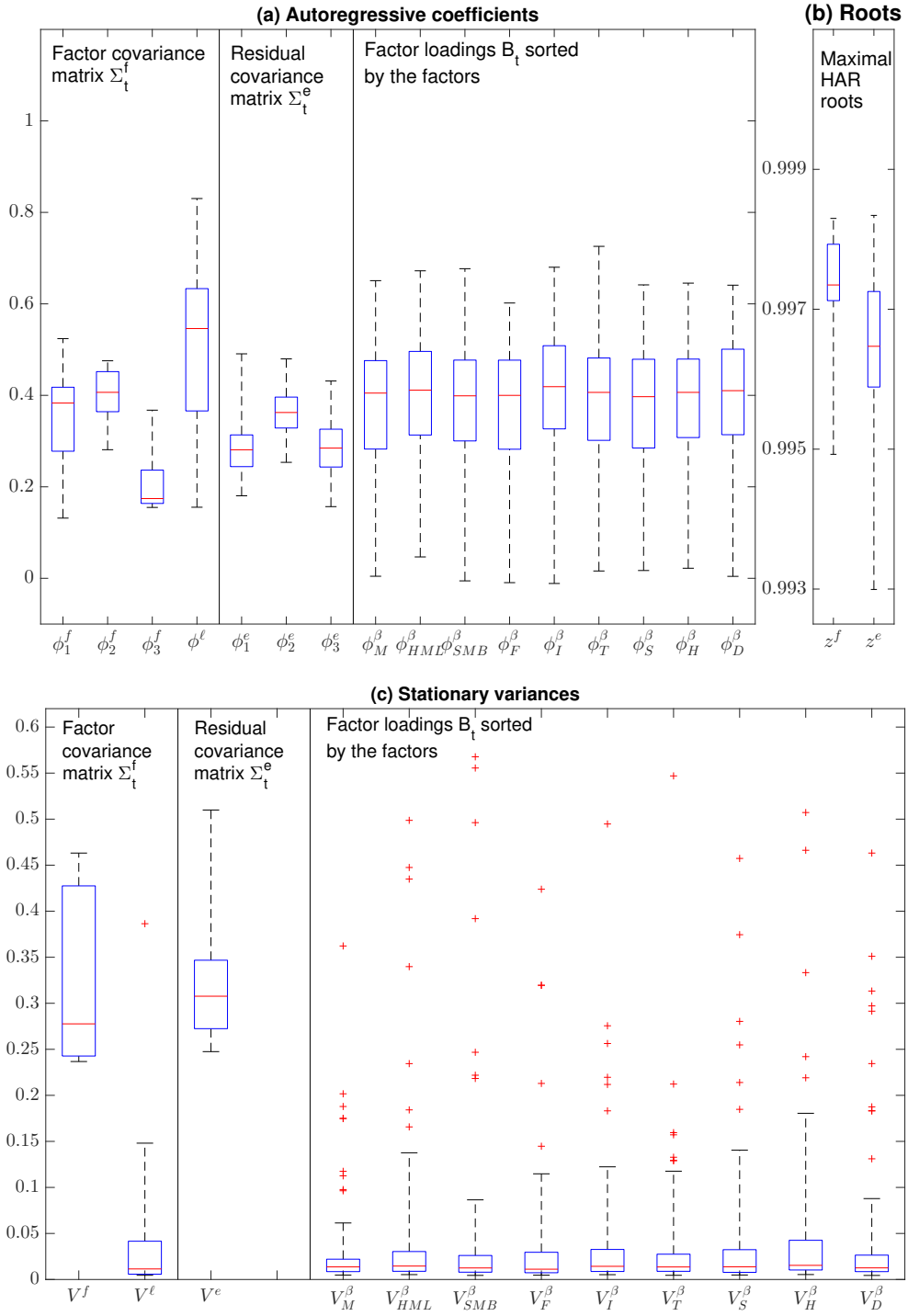
Panels (a) – (g): Time series plots of the realized variances; The gray shaded areas mark the two out-of-sample windows used in the forecasting experiments in Section 2.4.3. Panels (h) and (i): Sample ACF of the realized variance (blue line) and predicted ACF under the fitted 9F-HAR-vβ WFSS model (red line). Predicted ACF is obtained from simulated data from the fitted WFSS model for $T = 50,000$ time periods.

Figure 2.3.: Heat plots of the time-average of daily realized (residual) correlation matrices.



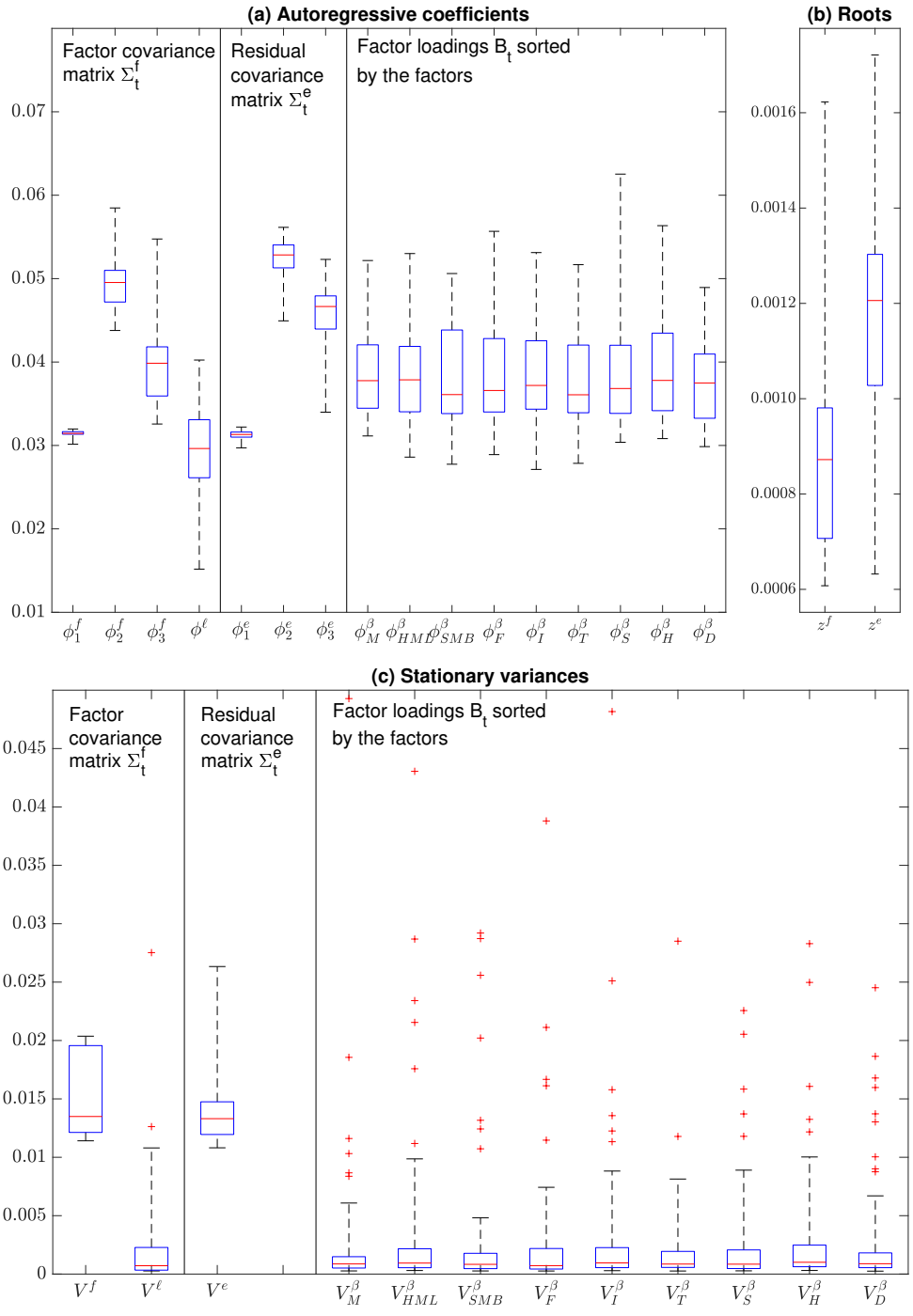
Panel (a): unconditional without factors; Panel (c): conditional on the market factor; Panel (b): conditional on the 3 Fama-French factors; Panel (d): conditional on the 3 Fama-French factors and 6 sector-specific ETFs. The black squares along the diagonal indicate sector blocks. The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies.

Figure 2.4.: Parameter estimates of the 9F-HAR-vβ WFSS model.



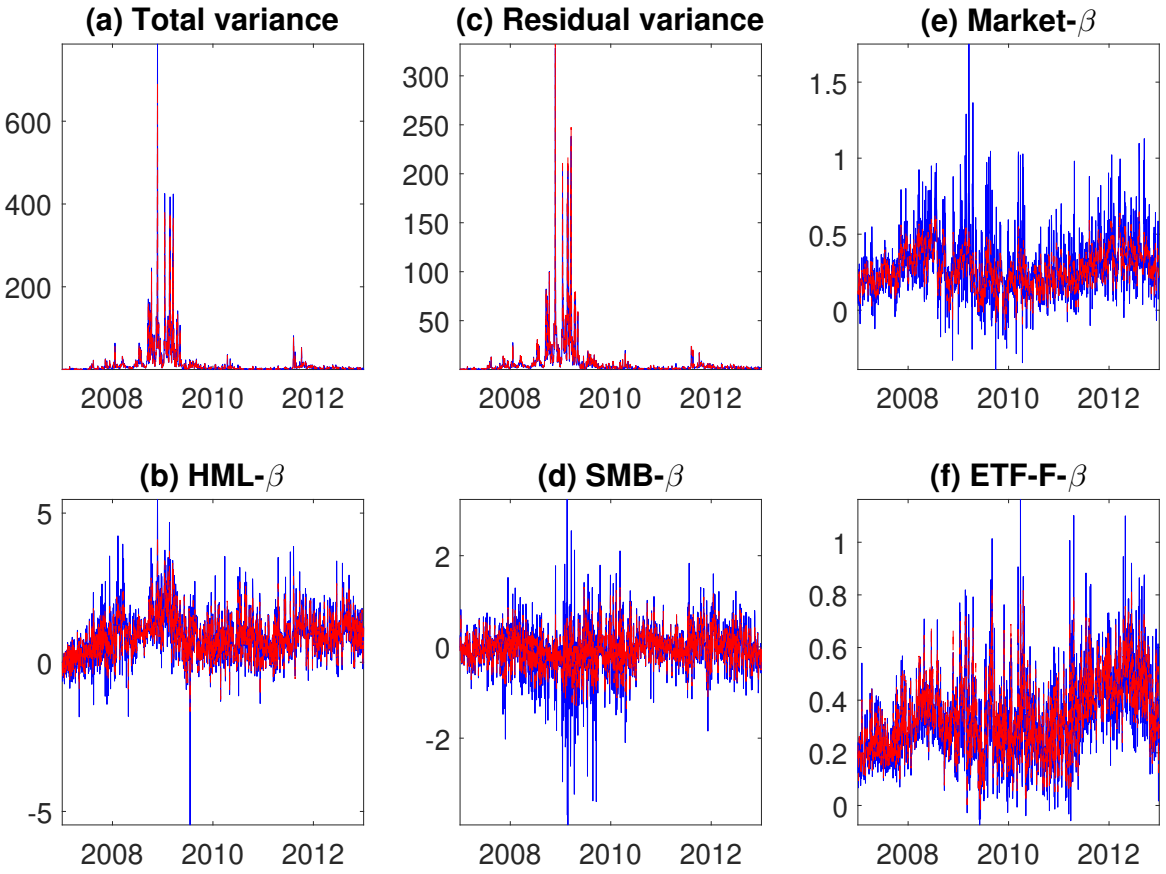
Box-plots of the posterior mean values for the autoregressive coefficients (panel a), maximal HAR roots in modulus (panel b) and stationary variances (panel c) for the AR(1) and HAR processes of the 9F-HAR-vβ WFSS model directing the integrated (co)variances in (Σ_t^f, Σ_t^e) and factor loadings in $\mathcal{B}_t$ (see Eqs. 2.8 – 2.13). The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies.

Figure 2.5.: Posterior standard deviation values of the parameters of the 9F-HAR-vβ WFSS model.



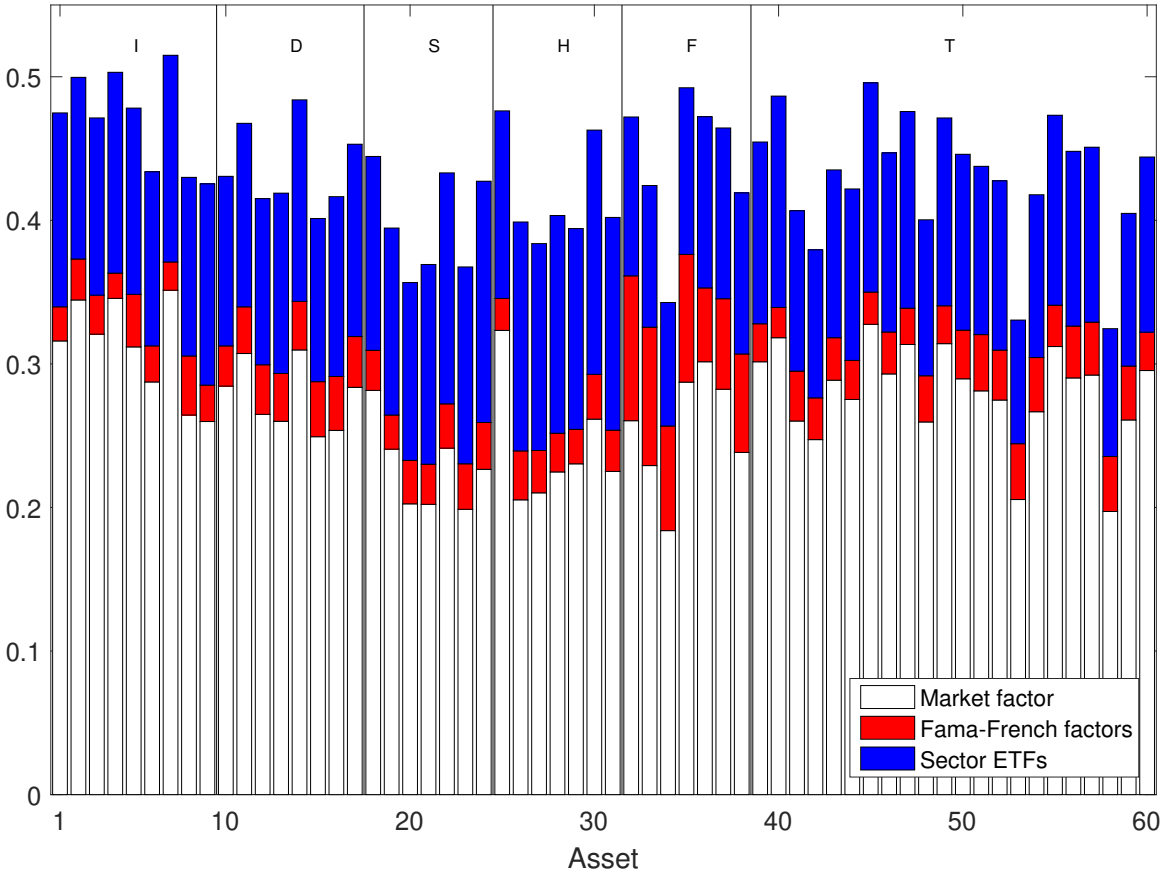
Box-plots of posterior standard deviation values for the autoregressive coefficients (panel a), maximal HAR roots in modulus (panel b) and stationary variances (panel c) for the AR(1) and HAR processes of the 9F-HAR-vβ WFSS model directing the integrated (co)variances in (Σ_t^f, Σ_t^e) and factor loadings in $\mathcal{B}_t$ (see Eqs. 2.8 – 2.13). The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies.

Figure 2.6.: Smoothed estimates for the integrated variance and loadings.



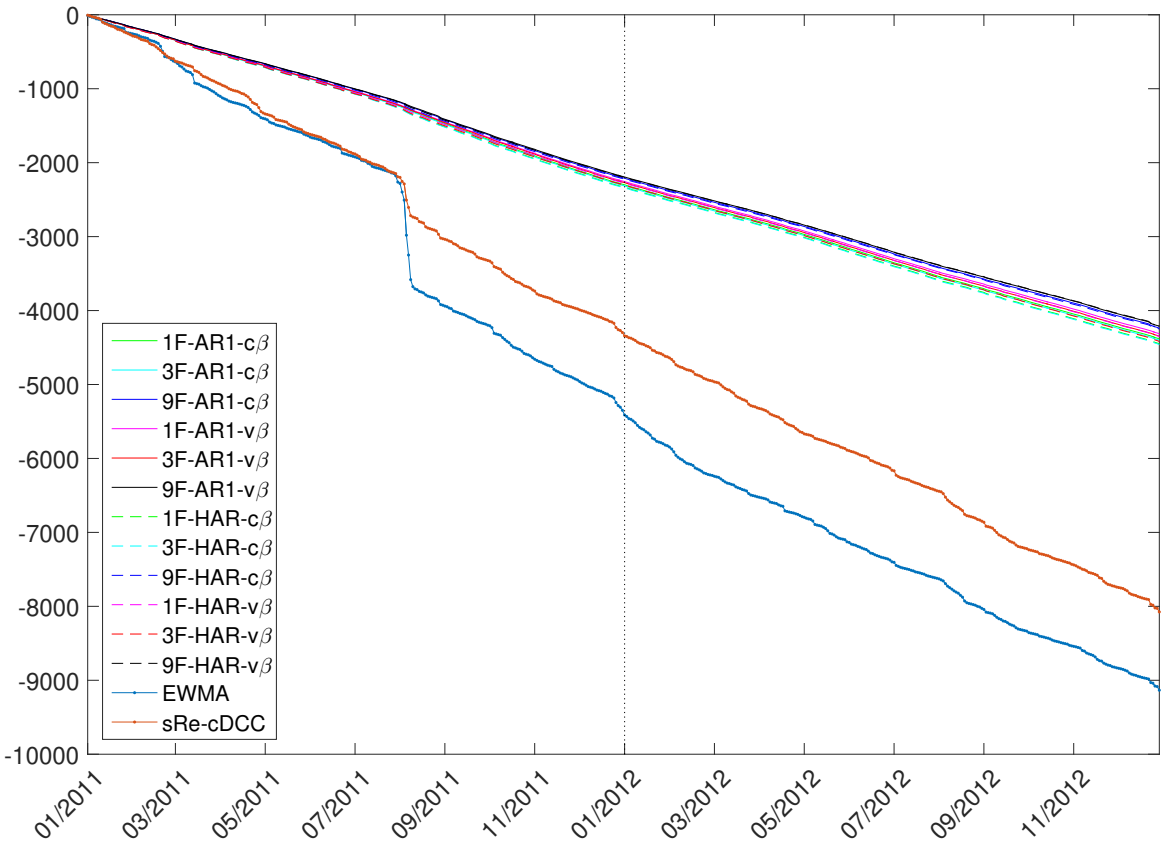
Time series plots of the smoothed estimates for the total integrated variance σ_t^r, integrated residual variance σ_t^e, and integrated loadings β_t on the market, HML, SMB and financial sector specific risk factor for the Citigroup stock under the 9-factor WFSS-tvβ-HAR model (red line), together with their respective observed realized counterparts (blue line).

Figure 2.7.: Fractions of realized return variation explained by the risk factors.



Bar plots of the fractions of realized stock return variation of the 60 assets explained by the risk factors, computed as $1 - (1/T)\sum_{t=1}^{T} \mathrm{E}(\sigma_{it}^e|C_{1:T})/c_{it}^r$, where $\mathrm{E}(\sigma_{it}^e|C_{1:T})$ is the posterior mean of the integrated residual obtained under the fitted HAR-vβ WFSS model with the 1 (market), 3 (market+HML+SMB) and 9 factors (market+HML+SMB+ETFs), respectively. White bar: Explained variation by the market; White and red bar: Explained variation by the market+HML+SMB; white, red and blue bar: explained variation by the market+HML+SMB+ETFs. The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies.

Figure 2.8.: Accumulated log-predictive likelihoods.



Time-series plots of the period-wise accumulated log-predictive likelihoods for the return variance of an equally weighted portfolio $\iota' C^r_{t+1} \iota$.

Chapter 3.

Composite Forecasting of Vast-Dimensional Realized Covariance Matrices using Factor State-Space Models

3.1. Introduction

Modeling and forecasting covariance matrices of financial assets is essential in various fields like option pricing, risk management and portfolio allocation. Numerous approaches have been studied in the past decades, mostly based on daily asset return data. Due to the increasing availability of intraday asset return data it is nowadays possible to compute accurate non-parametric ex-post estimates of daily integrated covariance matrices of asset returns, called realized covariance, which are known to incorporate much more information about daily conditional covariances than daily return series, and come with the advantage that time-series models can directly be applied.[1]

However, this involves two key challenges. Firstly, the need to assure positive definiteness of the covariance matrix forecasts, and secondly, to cope with the parameter proliferation arising through the fact that the number of objects to be modeled is proportional to the square of the number of assets. These aspects become even more pronounced in vast-dimensional applications, that are considered in this paper.

Nevertheless, the direct time-series modeling of realized covariance matrices has arisen as an important new strand of literature, pioneered by the studies of Gourieroux et al. (2009), Noureldin et al. (2012), Golosnoy et al. (2012), and Jin and Maheu

[1] For a description of the concept of realized covariance matrices see, for example, Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004), Park and Linton (2012), and Lunde et al. (2016).

(2013), who focused on capturing the time-series behavior of realized covariance matrices through matrix-variate distributions like the Wishart and inverse-Wishart. In addition Bauer and Vorkink (2011) and Chiriac and Voev (2011) focused on decompositions of realized covariances to apply standard time-series models for the prediction of future realizations. More recent contributions examine additional aspects in modeling realized covariances by using more complex distributions like mixtures of inverse-Wisharts or the Matrix-F distribution that allow for fat tails in the covariance (Jin and Maheu, 2016; Opschoor et al., 2017; Opschoor and Lucas, 2019; Vassallo et al., 2019). As further example, the Riesz distribution allows for incorporating liquidity differences in the intraday returns (Gribisch and Hartkopf, 2020, see also Chapter 4). The benefits of LASSO (least absolute shrinkage and selection operator) techniques on the predictive accuracy are examined in Callot et al. (2017). However, most of the proposed models are limited to low-dimensional applications, say with 30 assets or less.

In this paper we contribute to another branch in realized covariance modeling, i.e., the application of factor structures for the assets' covariance matrices to deal with empirically more realistic scenarios in increasing dimensions. Prominent examples of the factor approach are the models proposed by Tao et al. (2011), Asai and McAleer (2015), Jin et al. (2019), Sheppard and Xu (2019), Gribisch et al. (2020, see also Chapter 2) and Brito et al. (2018). While the first three approaches extract implicit factors from realized covariance data via an eigenanalysis and therefore build dynamic time-series models, the remaining three rely on economically motivated observed risk factors, like in the Fama and French (1993) three-factor model.

The use of observed risk factors is particularly appealing as it allows for time-variation in both the factor loadings and the residual components which yields flexibility gains when modeling the complex structures driving the evolution of realized covariances. We follow this route and propose a flexible factor state-space model for the prediction of realized covariance matrices of asset returns based on observed risk factors which is capable of handling dynamic systems of vast dimensions (say 200 and

more).

Our method relies on a factor decomposition of the realized asset covariance matrix from a joint construction of the realized measure for the factors and assets. After observing the covariances of both the assets and factors, we construct realized factor loadings and realized residual components from standard matrix decompositions resulting in a time-series for each of the individual parts. We then propose independent state-space models applied to the time-series of the factors, factor loadings and residual components akin to an approximate factor model taking the 'true' integrated factor and residual covariance matrices and integrated loadings as a latent state variables which are observed through their noisy realized counterparts.

For the factor and residual covariance matrices we combine independent Wishart measurement densities with parsimoniously parameterized matrix-variate Beta transitions in the vein of Windle and Carvalho (2014). Besides yielding positive definite covariance forecasts by construction, this specification has the advantage of being tractable in a sense that the predictive distribution of the data, and hence, the likelihood function is available in closed-form. As a further contribution, we derive additional useful properties of the covariance model at hand. For the factor loadings we rely on well established models from the TVP-VAR literature with stochastic volatility estimated by Maximum Likelihood (ML) techniques based on Kalman-like filtering.

Although a separate modeling approach has certain drawbacks as it ignores data-imposed dependencies in the measurements, it comes with the huge advantage of reducing model complexity allowing to efficiently handle vast-dimensional portfolios. In contrast to other state-space approaches that require high computational effort, our approach enables parameter estimation and prediction on a minute time scale. Based on the individual models, predictions for each component are readily made completely in parallel and are combined afterwards to obtain a composite forecast of the full asset covariance matrix. This makes the model particularly appealing for practitioners.

In order to investigate whether the loss of information arising from the composite

nature of our proposed model has any influence on its predictive accuracy we later include the Factor HEAVY (high-frequency based volatility) model of Sheppard and Xu (2019) in the set of competing models. The Factor HEAVY relies on a GARCH-type modeling of the observed factors' and assets' covariances based on a joint Wishart distribution combined with a strict factor structure (i.e., a sparse diagonal residual component). The diagonal residual assumption is crucial for their model as it enables a neat factorization into (conditionally) independent low-dimensional components allowing for straightforward estimation by Quasi Maximum Likelihood (QML). However, it imposes severe restrictions on the residual component which can be unrealistic in vast-dimensional applications. Our modeling approach allows to relax this assumption without effort.

The proposed composite factor state-space (CFSS)[2] model relies on the same factor decomposition as used in Brito et al. (2018). They combine the LASSO approach of Callot et al. (2017) for the factor and residual covariances with linear heterogeneous autoregressive (HAR, Corsi, 2009) processes for the factor loadings. Their approach has two major drawbacks in comparison to the CFSS model. First, since their model specification is prone to over-parameterization, they have to restrict the dynamics of the residual component to be driven by past residual variances only, neglecting possibly important information when forecasting the assets' covariance matrices. Second, and maybe more important, they operate with transformed time-series for the covariance parts based on the matrix-logarithm and make use of the matrix exponential function to ensure positive definiteness of the predictions which induces a bias to the forecasts. This bias might be neglectable in short-term forecasts, however, it gets more pronounced over longer horizons.

In the empirical part, we apply the CFSS model to daily covariance matrices for the returns of 225 NYSE traded stocks. In an extensive out-of-sample analysis we compare several model specifications based on varying sets of observed risk factors and different residual structures, as well as different restrictions imposed to the dynamics

[2]The name composite factor state-space model arises from the composite nature of how the predictions for the full asset covariance matrix are obtained.

of the realized covariance components. We illustrate the predictive performance of our approach relative to several competing models. As performance measures we consider the accuracy of the (co)variance predictions and the ability to produce predictions for the global-minimum variance portfolio under different side restrictions as well as the ability to produce predictions for the mean-variance portfolio using momentum signals. Our out-of-sample results show that the CFSS model performs favorable in almost all dimensions relative to its competitors.

The remainder of the paper is organized as follows: Section 3.2 introduces the proposed CFSS model. The Maximum Likelihood estimation and parameter restrictions and the construction of composite forecasts are discussed in Section 3.3. Section 3.4 describes the data set. Section 3.5 discusses implementation details and presents the out-of-sample results. Finally, Section 3.6 concludes. Additional material and derivations for the model (parts) at hand are provided in Appendix B.

3.2. The Model

3.2.1. Factor decomposition of realized covariances

Consider a positive definite, symmetric consistent estimate C_t of the joint latent integrated covariance matrix Σ_t, observed for a panel of p logarithmic asset prices and q observed factor prices at the trading days $t = 1, \ldots, T$. Let C_t be partitioned as

$$C_t = \begin{pmatrix} C_t^f & C_t^{rf} \\ C_t^{fr} & C_t^r \end{pmatrix}, \tag{3.1}$$

where C_t^f constitutes the $q \times q$ realized covariance matrix for the factors, C_t^r denotes the $p \times p$ realized covariance matrix for the assets and $C_t^{fr} = (C_t^{rf})'$ denotes the $p \times q$ matrix of the realized covariances between the factors and assets.

In the present paper our particular interest is to predict the realized covariance matrix of the asset returns C_t^r. To this end, we exploit a factor decomposition relating

the assets' realized covariance to the risk factors via

$$C_t^r = B_t C_t^f B_t' + C_t^e, \tag{3.2}$$

where the matrices B_t $(p \times q)$ and C_t^e $(p \times p)$ are uniquely found from a block LDL factorization of the block matrix representation in Eq. (3.1) as the regression coefficient matrix and the resulting Schur complement of C_t^f in C_t, i.e.

$$B_t = C_t^{fr}(C_t^f)^{-1}, \qquad C_t^e = C_t^r - C_t^{fr}(C_t^f)^{-1}C_t^{rf}. \tag{3.3}$$

The matrices B_t and C_t^e can be interpreted as realized factor loadings and as realized residual covariance matrix approximating the integrated loadings $\mathcal{B}_t$ and residual components Σ_t^e, respectively.

Since we observe the complete covariance C_t at each point in time, we can use Eq. (3.3) to determine distinct time-series for C_t^f, C_t^e and B_t, such that these can be modeled and predicted separately. Afterwards they can be recomposed to obtain a forecast for the major quantity of interest, the full covariance of the assets C_t^r. Specifications for the dynamics of the individual parts are presented in the following section.

3.2.2. Model equations

We take the integrated components Σ_t^f, Σ_t^e and $\mathcal{B}_t$ as latent state variables observed through their noisy realized counterparts and model them by a state-space approach with independent measurement densities $f(C_t^f \mid \Sigma_t^f)$, $f(C_t^e \mid \Sigma_t^e)$, $f(B_t \mid \mathcal{B}_t)$, and corresponding transition equations.

Before specifying the respective measurements for the individual parts, we make two additional assumptions on B_t and C_t^e. First, we assume row-wise independence of the betas, i.e., $\mathrm{Cov}(b_{it}, b_{jt}) = 0 \ \forall i \neq j$, and $i, j = 1, \ldots, p$, where b_{it} is a vector containing the ith row's entries of B_t. This assumption is comparably mild as, e.g., Brito et al. (2018) assume element-wise independence for the entries in B_t. Second,

we impose the following block structure for the residual covariance

$$C_t^e = \begin{pmatrix} C_{1t}^e & * & * \\ * & \ddots & * \\ * & * & C_{St}^e \end{pmatrix}, \tag{3.4}$$

where the S diagonal blocks are of size $p_i \times p_i$, $i = 1, \ldots, S$. The asterisk '$*$' indicates residual covariances which are economically insignificant and thus are ignored.[3] The structure in Eq. (3.4) is supported by empirical evidence in many studies when the number of factors is chosen large enough, see, e.g., Fan et al. (2016); Aït-Sahalia and Xiu (2017); Brito et al. (2018); Gribisch et al. (2020). Expecting the significant entries to cluster in blocks around the diagonal might appear arbitrary. However, by rearranging the rows and columns in C_t^r such a block structure can almost always be achieved (at least approximately). As it is now common in the literature to model Σ_t^e as a sparse matrix (e.g., Sheppard and Xu, 2019 assume Σ_t^e to be strict diagonal and only use residual variance information in a one-factor setting), we impose a block-diagonal structure and assume that the integrated residual covariance Σ_t^e is sufficiently measured by the marginals $f(C_{1t}^e \mid \Sigma_{1t}^e), \ldots, f(C_{St}^e \mid \Sigma_{St}^e)$.

Next, we specify the measurement distributions for the different components C_t^f, $\{C_{it}^e\}_{i=1}^S$, and $\{b_{it}\}_{i=1}^p$, which read as follows

$$C_t^f \mid \Sigma_t^f \sim \mathcal{W}_q(n^f, \Sigma_t^f/n^f), \quad C_{it}^e \mid \Sigma_{it}^e \sim \mathcal{W}_{p_i}(n_i^e, \Sigma_{it}^e/n_i^e), \quad b_{it} \mid \beta_{it}, \Sigma_{it}^b \sim \mathcal{N}_q(\beta_{it}, \Sigma_{it}^b), \tag{3.5}$$

where $\mathcal{W}_m(d, s)$ denotes a m-dimensional Wishart distribution with $d > m - 1$ degrees of freedom and scale matrix s, and $\mathcal{N}_m(a, b)$ denotes a m-dimensional Normal distribution with mean a and covariance b. The latter choice is motivated by the fact that similar (marginal) distributions result from the assumption of a joint conditional Wishart distribution for C_t in (3.1), which can be seen as a natural candidate for the measurement (c.f. Gribisch et al., 2020, and see Philipov and Glickman, 2006;

[3] Economic significance is explained in more detail in Section 3.4.

Golosnoy et al., 2012; Noureldin et al., 2012; Bauwens et al., 2016, for applications of the Wishart to realized covariances).

In order to complete the composite factor state space (CFSS) model, we specify the transition densities for the latent time-varying integrated factor and residual covariances, Σ_t^f and $\{\Sigma_{it}^e\}$, as well as the integrated factor loadings $\{\beta_{it}\}$ and their time-varying covariance matrices $\{\Sigma_{it}^b\}$. For C_t^f and the distinct blocks $\{C_{it}^e\}$ we adopt the model of Windle and Carvalho (2014), which combines the measurements in Eq. (3.5) with a generalized Matrix-Beta type-I transition for the integrated precision, i.e., the inverse of the integrated covariances Σ_t^f and $\{\Sigma_{it}^e\}$, respectively, implying a shifted Matrix-F transition for Σ_t^f and $\{\Sigma_{it}^e\}$ themselves (see Appendix B.1.2 for a derivation). In the vein of Windle and Carvalho (2014) we refer to this model as Uhlig Extension (UE) as it depicts a generalization of the process originally proposed in Uhlig (1994, 1997).

The UE model comes with three major advantages. First, irrespective of the dimension of the underlying covariance parts, it is parameterized parsimoniously with only three parameters. Parsimony has proven to be particularly advantageous in predicting high-dimensional covariance matrices (see, e.g., Bauwens et al., 2016). Second, it yields positive definite (p.d.) forecasts for C_t^f and C_t^e by construction. Hence, there is no need to resort to any transformation like the Cholesky decomposition, that depends on the ordering of assets in C_t, or the matrix logarithmic transformation (Bauer and Vorkink, 2011), which may induce severe biases of the predictions $\widehat{C}_{t+1}^r$ through the exponential re-transformation. Third, and most important, for the UE model we can exploit the conjugacy between the Wishart and the Matrix-Beta distribution to obtain closed-form filtering formulas for tracking the latent states of the system and allowing the model parameters to be estimated by Maximum Likelihood in one step (details are given in the following section).

In the UE model for the factor covariance matrix, the $q \times q$ symmetric, p.d. inte-

grated precision $\Omega_t^f = (\Sigma_t^f)^{-1}$ evolves through

$$\Omega_t^f = \mathcal{U}(\Omega_{t-1}^f)'\Psi_t^f\mathcal{U}(\Omega_{t-1}^f)\Big/\lambda^f, \qquad \Psi_t^f \sim \mathcal{B}_q^I\left(\frac{k^f}{2}, \frac{n^f}{2}\right), \tag{3.6}$$

with conditional expectation given by $\mathrm{E}[\Omega_t^f|\Omega_{t-1}^f] = \Omega_{t-1}^f k^f/[\lambda^f(k^f+n^f)]$ and initial condition $\Omega_1^f \sim \mathcal{W}_q(k^f, (n^f S_0^f)^{-1}/\lambda^f)$. $\mathcal{U}(\cdot)$ denotes the upper Cholesky factor, Ψ_t^f are $q \times q$ symmetric, p.d. iid Matrix-Beta type-I shocks to Ω_t^f.[4] Besides the symmetric scale matrix S_0^f of the initial distribution, the UE model has three parameters, the degrees of freedom (d.o.f.) $n^f, k^f > q-1$ and the smoothing parameter $\lambda^f > 0$, controlling the dynamics of the Ω_t^f process.[5]

Similarly to the integrated factor precision, we define $\Omega_{it}^e = (\Sigma_{it}^e)^{-1}$ for the $i = 1, \ldots, S$ distinct $p_i \times p_i$ symmetric, p.d. blocks of Σ_t^e, and assume

$$\Omega_{it}^e = \mathcal{U}(\Omega_{it-1}^e)'\Psi_{it}^e\mathcal{U}(\Omega_{it-1}^e)\big/\lambda_i^e, \qquad \Psi_{it}^e \sim \mathcal{B}_{p_i}^I\left(\frac{k_i^e}{2}, \frac{n_i^e}{2}\right), \tag{3.7}$$

with initial conditions $\Omega_{i1}^e \sim \mathcal{W}_{p_i}(k_i^e, (n_i^e S_{i0}^e)^{-1}/\lambda_i^e)$, where $n_i^e, k_i^e > p_i - 1$, $\lambda_i^e > 0$ and S_{i0} is the p.d. $p_i \times p_i$ initial scale matrix.

For the factor loadings $\{b_{it}\}$ we use a local-level version of the time-varying parameter VAR with stochastic volatility proposed in Moura and Noriller (2019). This model assumes the latent βs to follow heteroscedastic random walks and uses a Wishart process for modeling the volatility of the system. The latent $q \times 1$ vectors of factor loadings b_{it}, $i = 1, \ldots, p$, evolve as driftless random walk processes of the form

$$\beta_{it} = \beta_{it-1} + \eta_{it}, \qquad \eta_{it} \sim \mathcal{N}(0, \Sigma_{it}^b/\sigma_i), \tag{3.8}$$

where the transition is allowed to be affected by multivariate stochastic volatility shocks via Σ_{it}^b, scaled by $1/\sigma_i$. Similar to the UE model, the evolution of $\Omega_{it}^b = (\Sigma_{it}^b)^{-1}$

[4] See e.g. Gupta and Nagar (2000) for useful theorems about the matrix-variate Beta distribution.

[5] In the literature the smoothing parameter λ^f is typically not estimated, but tied to the d.o.f. n^f and k^f through different restrictions which are directly imposed during estimation of the latter. We discuss these restrictions in Section 3.3.2 below.

is specified by the following multiplicative transition

$$\Omega_{it}^b = \mathcal{U}(\Omega_{it-1}^b)'\Psi_{it}^b\mathcal{U}(\Omega_{it-1}^b)/\,\lambda_i^b, \qquad \Psi_{it}^b \sim \mathcal{B}_q^I\left(\frac{k_i^b+1}{2}, \frac{1}{2}\right), \tag{3.9}$$

with fixed initial condition. Here, Ψ_{it}^b are $q \times q$ iid singular Beta type-I shocks to Ω_{it}^b, with single d.o.f. parameter k_i^b governing the time-variation in the precision, given a predetermined value of $0 < \lambda_i^b < 1$. Again, exploiting the conjugacy between the Normal, Wishart and Matrix-Beta distributions, analytical expressions for the filtering distribution of the loadings and the likelihood function can be found, enabling fast and straightforward estimation of the unknown model parameters. It is worth noting that for $k_i^b \to \infty$ the nonlinear state-space model in Eqs. (3.5), (3.8) and (3.9) collapses to a homoscedastic version (Kim, 2014; Moura and Noriller, 2019). For $1/\sigma_i \to 0$ the latent factor loadings are time-invariant.

3.3. Estimation and Forecasting

3.3.1. Predictive distributions and likelihood estimation

This section describes the ML estimation of the CFSS model. The unknown model parameters are estimated by independently maximizing the distinct elements in the conditional log-likelihood

$$\log L^r = \log L^f(\theta^f; C_{1:T}^f) + \sum_{i=1}^{S} \log L_i^e(\theta_i^e; C_{i1:T}^e) + \sum_{j=1}^{p} \log L_j^b(\theta_j^b; b_{j1:T}), \tag{3.10}$$

where L^f, L_i^e and L_j^b denote the likelihood contributions of the factors, residual blocks and factor loadings, with parameters $\theta^f = (n^f, k^f, \lambda^f)'$, $\theta_i^e = (n_i^e, k_i^e, \lambda_i^e)'$ and $\theta_j^b = (k_j^b, \lambda_j^b, \sigma_j)'$, respectively. The notation $A_{s:\tau}$ is used to denote the collection $\{A_s, \ldots, A_\tau\}$.

The measurement distributions in Eq. (3.5) combined with the corresponding transition equations for Ω_t^f, $\{\Omega_{it}^e\}$, $\{\beta_{it}\}$ and $\{\Omega_{it}^b\}$ in Eqs. (3.6) to (3.9) constitute independent nonlinear state-space models for the realized quantities C_t^f, $\{C_{it}^e\}$ and $\{b_{jt}\}$. By

exploiting the conjugacy between Normal, Wishart and Matrix-Beta distributions, Windle and Carvalho (2014) and Moura and Noriller (2019) derived several useful propositions for the UE model and the TVP-VAR, including the forward filtering distributions for the latent precision matrices and factor loadings, which turn out to be Wishart and multivariate-t, respectively. In addition, Windle and Carvalho (2014) propose a backward sampling scheme to efficiently draw the latent sequences $\{\Omega_t^f\}$ ($\{\Omega_{it}^e\}, i = 1, \ldots, S$) in a single sweep within a Bayesian MCMC sampler. In the following we focus on their results enabling fast and easy to implement ML estimation of the model parameters, and the prediction of future realizations of C_t^r in Eq. (3.2).

Let $C_{1:t}^f$ capture the information on the realized factor covariance matrices from period 1 up to t. With the transition in Eq. (3.6) and the corresponding initial condition, the predictive distribution for C_{t+1}^f given past information is of Matrix-F type, i.e. $C_{t+1}^f | C_{1:t}^f \sim \mathcal{F}_q(n^f, k^f, \lambda^f S_t^f)$, with density given by

$$f(C_{t+1}^f \mid C_{1:t}^f) = \frac{\Gamma_q(\frac{n^f+k^f}{2})}{\Gamma_q(\frac{n^f}{2})\Gamma_q(\frac{k^f}{2})} \frac{|C_{t+1}^f|^{(n^f-q-1)/2}|\lambda^f S_t^f|^{k^f/2}}{|\lambda^f S_t^f + C_{t+1}^f|^{(n^f+k^f)/2}}, \tag{3.11}$$

with d.o.f. n^f, k^f and scale $\lambda^f S_t^f$, where $S_t^f = \lambda^f S_{t-1}^f + C_t^f$ (see Proposition 3 in Windle and Carvalho, 2014). The d.o.f. k^f are known as fat-tail parameters with low (high) values indicating fatter (thinner) tails. For $k^f \to \infty$ the predictive distribution degenerates to the Wishart with n^f d.o.f. (see Opschoor et al., 2017). The conditional first moment that is used to generate one-step forecasts for C_{t+1}^f is given by[6]

$$\mathrm{E}[C_{t+1}^f \mid C_{1:t}^f] = \frac{\lambda^f n^f}{k^f - q - 1} S_t^f, \tag{3.12}$$

where solving the recursion for S_t^f yields

$$S_t^f = \sum_{i=0}^{t-1} (\lambda^f)^i C_{t-i}^f + (\lambda^f)^t S_0^f, \tag{3.13}$$

implying that the forecast of C_{t+1}^f will be a scaled, geometrically weighted sum of the

[6] A derivation for the conditional second moment, i.e. $\mathrm{Cov}[\mathrm{vec}(C_{t+1}^f)]$, is given in Appendix B.1.1.

previous observations. From Eq. (3.13) one can see that λ^f controls the degree of smoothing of the observations. This is crucial when forming estimates and one-step ahead predictions (cf. Windle and Carvalho, 2014).[7]

Based on Eq. (3.11) the likelihood function for the factor part obtains as $L^f = f(C_1^f; S_0^f) \prod_{t=2}^{T} f(C_t^f | C_{1:t-1}^f)$.[8] Analogous results hold for the predictive distributions $f(C_{it+1}^e \mid C_{i1:t}^e)$ and likelihoods L_i^e of the distinct residual covariance blocks, and left out here for space saving reasons.

Now, let $b_{i1:t}$ capture the information on the realized factor loadings from period 1 up to t, for $i = 1, \ldots, p$. The predictive density of b_{it+1} is that of a scaled multivariate t distribution (see Moura and Noriller, 2019, Corollary 1)[9]

$$f(b_{it+1} \mid b_{i1:t}) = \frac{\Gamma\left(\frac{k_i^b+1}{2}\right)}{\pi^{q/2}\Gamma\left(\frac{k_i^b-q+1}{2}\right)} |P_{it+1}|^{1/2} \left[1 + (b_{it+1} - \bar{\beta}_{it+1})' P_{it+1} (b_{it+1} - \bar{\beta}_{it+1})\right]^{-\frac{k_i^b+1}{2}}, \tag{3.14}$$

with $k_i^b - q + 1$ d.o.f., mean $\bar{\beta}_{it+1}$, and scale $(k_i^b - q + 1)^{-1} P_{it+1}^{-1}$, where

$$\bar{\beta}_{it+1} = \frac{N_{it}}{(N_{it}+1)} (\bar{\beta}_{it} + b_{it} N_{it}^{-1}), \tag{3.15}$$

$$P_{it+1} = \frac{N_{it}}{k_i^b (N_{it}+1)} (S_{it}^b)^{-1}, \tag{3.16}$$

$$S_{it+1}^b = \lambda_i^b S_{it}^b + \frac{\lambda_i^b N_{it}}{(k_i^b+1)(N_{it}+1)} (b_{it} - \bar{\beta}_{it})(b_{it} - \bar{\beta}_{it})', \tag{3.17}$$

$$N_{it+1} = \frac{\sigma_i \lambda_i^b (N_{it}+1)}{\sigma_i + \lambda_i^b (N_{it}+1)}. \tag{3.18}$$

The initial conditions are fixed parameters.[10]

[7] In Section 3.3.2 we propose several restrictions on the smoothing parameter, directly relating it to the shape of the predictive distribution.

[8] The likelihood function is based on S_0^f, which could already be challenging to estimate in small to medium dimensional applications. To circumvent this problem, we apply the approach proposed by Windle and Carvalho (2014), i.e., we set aside the first T' observations and use the set $\{S_{T'}^f, C_{T':T}^f\}$, with $S_{T'}^f = \sum_{i=0}^{T'} (\lambda^f)^i C_{T'-i}^f$ to estimate the remaining model parameters.

[9] Note a typo in Moura and Noriller (2019), where they define the scale of the multivariate t as $\frac{k_i^b}{k_i^b-q+1} S_{it}^b$ by mistake. The error has been fixed here.

[10] We set $N_{i1} = 1$ and $S_{i1}^b = I_q$.

The predictive expectation of the factor loadings is given by

$$\mathrm{E}[b_{it+1} \mid b_{i1:t}] = \bar{\beta}_{it+1}. \tag{3.19}$$

Again, the likelihood function L_i^b is obtained as the product of the predictive densities in Eq. (3.14).

3.3.2. Parameter restrictions

Our proposed CFSS model as defined by Eqs. (3.5) to (3.9) includes $3 \times (1 + S + p)$ parameters, whereof the majority relates to the covariance dynamics of the factor loadings. In order to further increase the parsimony of our approach we utilize commonly imposed restrictions to the d.o.f. parameters of the loading processes and the smoothing parameters.

In initial investigations of the realized factor loadings in scenarios with different risk factors, we found the estimates for k_i^b to be similar throughout the asset dimension p. Moreover, we found that minor changes in the d.o.f. parameters do not have noteworthy effect on the predictive ability regarding mean forecasts for b_{it+1}. Hence, we preset k_i^b in the ongoing of this paper. Similar restrictions are indeed imposed by Uhlig (1997) and Kim (2014) in the context of constant coefficient VARs (see also Moura and Noriller, 2019). Kim (2014) suggests to choose values $k_i^b \in [10, 25]$ which mirrors our findings for $\hat{k}_i^b$. Consequently, we adopt this in the empirical analysis.[11] For given k_i^b the process Ω_{it}^b is asymptotically degenerate if λ_i^b is too large, and explosive if λ_i^b is too small. Hence, it seems reasonable to further tie the smoothing parameter to the d.o.f. via $\lambda_i^b = k_i^b/(k_i^b + 1)$, implying $0.90 < \lambda_i^b < 0.97$, for $k_i^b \in [10, 25]$.

In addition we consider three different restrictions on the smoothing parameters $(\lambda^f, \{\lambda_i^e\})$. They are all considered as distinct model variants and their predictive performance is analyzed in the empirical application below. Each of them has been found to be valuable on its own. Note that all restrictions imply $\lambda^f, \{\lambda_i^e\} \in (0, 1)$ and

[11] The exact restrictions for k_i^b are postponed to Section 3.5.

that they can straightforwardly be imposed during the estimation of the degree of freedom parameters. Here they are exemplified for the factor part. Similar restrictions are later imposed for the residual blocks as well.

The first restriction has been originally proposed by Windle and Carvalho (2014) and is given by

$$\lambda^f = \left(1 + \frac{n^f}{k^f - q - 1}\right)^{-1}, \tag{R1}$$

It implies that the one-step-ahead forecast for the realized factor (residual) covariance based on (3.12) is obtained as the exponentially weighted moving average (EWMA)

$$\mathrm{E}[C^f_{t+1} \mid C^f_{1:t}] = (1 - \lambda^f)C^f_t + \lambda^f \mathrm{E}[C^f_t \mid C^f_{1:t-1}],$$

which is known to deliver decent (short-term) predictive performance.[12] Notably, this restriction also results in a martingale for the evolution of latent integrated factor (residual) covariance matrices, i.e., $\mathrm{E}[\Sigma^f_t|\Sigma^f_{t-1}] = \Sigma^f_{t-1}$ ($\mathrm{E}[\Sigma^e_{it}|\Sigma^e_{it-1}] = \Sigma^e_{it-1}$). A derivation for this result is given in Appendix B.1.2.

The second restriction is chosen to induce a random walk behavior for the latent integrated precision matrices in Eq. (3.6), say $\mathrm{E}[\Omega^f_t|\Omega^f_{t-1}] = \Omega^f_{t-1}$, i.e.,

$$\lambda^f = \frac{k^f}{k^f + n^f}. \tag{R2}$$

A similar restriction has been imposed by Uhlig (1997) in the context of singular Matrix-Beta transitions for the latent precision. In the context of stochastic covariance modeling for daily return series, Moura et al. (2020) found (R2) to be performing well in out-of-sample portfolio allocations.

The third restriction traces back to the work of Shephard (1994), where in an univariate modeling framework akin to Uhlig (1997) the smoothing parameter is chosen

[12] Note that under restriction (R1) the prediction for C^f_{t+1} is similar to the one implied by the RiskMetrics methodology (Morgan, 1996). However, in the latter the smoothing parameter is set arbitrarily to 0.96, whilst in the former λ is determined by the distributional shape mirrored in the degree of freedom parameters n^f and k^f.

as the geometric mean of the latent beta shocks in the precisions' transition equations. Shephard (1994) argues that this restriction rules out the case of the precision collapsing to zero if $t \to \infty$. Here, we propose a multivariate extension of this restriction which takes the form

$$\lambda^f = \exp\{\mathrm{E}[\log|\Psi_t^f|]/q\}. \tag{R3}$$

In Appendix B.1.3 we show that $\mathrm{E}[\log|\Psi_t^f|] = \Psi_q^\star(k^f/2) - \Psi_q^\star((n^f+k^f)/2)$, where $\Psi_d^\star(a)$ denotes the d-variate digamma function (see e.g. Abramowitz and Stegun, 1972; Gupta and Nagar, 2000) and that (R3) results in a random walk for the log determinant process of the integrated precision matrices.

Figure 3.1 visualizes the restrictions on the smoothing parameter for an exemplary scenario of 25 assets, which corresponds to the average sector-size in the empirical application. Panels (a) – (c) show contour plots of (R1), (R2) and (R3) on the grid $[26, 300] \times [26, 300]$ for the d.o.f. parameters. The plots indicate that $\lambda \to 1$ ($\lambda \to 0$) for n fix at its lower bound and $k \to \infty$ (k fix at its lower bound and $n \to \infty$). Hence, the less fat tailed and noisier the observations are, the more the model smooths, and vice versa.

Panels (d) – (f) show contour plots of the differences between the ML estimate for λ (when fixing n and k) and the respective restrictions (R1), (R2) and (R3). The black circles mark the corresponding unrestricted ML estimates of the d.o.f. parameters, i.e., $\hat{n} \approx 89$, $\hat{k} \approx 175$. The unrestricted MLE for the smoothing parameter is $\hat{\lambda} = 0.63$ with a standard deviation of 0.017. We can see that (R1) and (R3) tend to be smaller than the MLE, whereas (R2) exceeds the MLE slightly, though all differences lie within one standard deviation. The performance of (R1) and (R3) deteriorates if both d.o.f. tend to their lower bound. However, it is common in practice to impose these restrictions when estimating the d.o.f. in order to reduce estimation uncertainty.

Imposing the restrictions stated above reduces the number of parameters to be estimated to $2 \times (1 + S) + p$. This further increases the parsimony but preserves the

flexibility of the model at hand and allows for very fast estimation of the unknown model parameters based on numerical optimization routines.[13]

3.3.3. Composite forecasting

Exploiting the decomposition of the realized asset covariance in Eq. (3.2), we write the forecasting equation for the realized asset covariance C^r_{t+1} by combining the distinct forecasts for C^f_{t+1}, $\{C^e_{it+1}\}$ and $\{b_{it+1}\}$, i.e.,

$$\widehat{C}^r_{t+1} = \widehat{B}_{t+1}\widehat{C}^f_{t+1}\widehat{B}'_{t+1} + \widehat{C}^e_{t+1}, \tag{3.20}$$

where we use $\widehat{C}^f_{t+1} = \mathrm{E}[C^f_{t+1} \mid C^f_{1:t}]$, $\widehat{C}^e_{it+1} = \mathrm{E}[C^e_{it+1} \mid C^e_{i1:t}]$ and $\widehat{b}_{it+1} = \mathrm{E}[b_{it+1} \mid b_{i1:t}]$ as obtained from the respective predictive moments in (3.12) and (3.19).[14] Since the forecasts of the factor and residual covariance are p.d., the forecast for the assets covariance will be p.d. by construction as well.

3.4. Data

The given data set includes and expands the 60 dimensional data of Gribisch et al. (2020). It consists of 1510 daily observations for 225 stocks traded at the New York Stock Exchange and 12 risk factors, observed between January 3, 2007 and December 31, 2012, comprising a total of 28,203 time series of realized variances and covariances. The stocks are selected by liquidity from the S&P 500 index and are sorted by their sector and industry classification according to the Global Industrial Classification Standard (GICS). For the observed factors we use the market, the high-minus-low price-earnings ratio (HML) and small-minus-big market capitalization (SMB) factors as in the Fama and French (1993) three-factor model. Additionally we consider the

[13] In the empirical application in Section 3.5 we, e.g., consider a model for 225 assets with block-diagonal residual assumption based on nine industry sectors resulting in 245 (705) parameters for the restricted (unrestricted) model. Parallel estimation of all parameters takes < 5 minutes on a standard PC with a 2.3 GHz Intel Core i7 processor in MATLAB.

[14] Since the predictions are of composite nature, they ignore non-linearities in the term $B_{t+1}C^f_{t+1}B'_{t+1}$ when predicting C^r_{t+1}. A simulation study, however, shows that these non-linearities are negligible (see Appendix B.3).

sector-specific Spyder Exchange-Traded Funds (SPDR ETFs) for the nine sectors covered by the 225 stocks (Fan et al., 2016; Aït-Sahalia and Xiu, 2017; Gribisch et al., 2020). A brief summary of the sector sizes along with the sector specific ETF tickers and some descriptive statistics are given in Table 3.1. A full list of all stocks included in the data set is given in Table B.3.

The daily realized covariance matrices C_t are computed by using the composite realized kernel method of Lunde et al. (2016) based on 5-minute returns for the Fama-French factors and 1-minute returns for the assets as well as the SPDR ETFs (see Barndorff-Nielsen et al., 2011; Lunde et al., 2016, for further details). Given the joint realized covariance matrices for the assets and factors, we compute according to Eq. (3.3) the realized residual covariance matrices C_t^e and the realized factor loadings B_t which represent estimates for integrated residual covariance matrices and loadings, for various sets of factors. Figure 3.2 shows time-series plots of the realized variances for the three Fama-French factors, as well as sector-wise averages of realized asset variances, realized factor loadings and realized residual variances for the Industrial, Health Care and Financial sector.

In Figure 3.3 we analyze the sparsity pattern for the realized residual covariance matrices. Following Aït-Sahalia and Xiu (2017) and Brito et al. (2018) we determine the economically significant entries of the residual correlation, i.e., we flag each entry with a minimum absolute correlation of 0.15 for at least 1/3 of the sample period. Panel (a) displays the results for the realized asset covariance matrices. Panels (b) – (d) display the results for the realized residual components after removing the common covariation driven by the market factor only, the three Fama-French factors and the Fama-French factors plus nine sector specific factors based on the ETFs, respectively.

The analysis shows that using only one or three factors still yields very dense residual correlation patterns, indicating not only a strong rejection of the strict diagonality assumption, but also of the block-diagonality based on industry sectors (highlighted as black rectangles). However, by inclusion of the sector ETFs almost all of the significant inter-sectoral and most of the intra-sectoral correlations vanish, leaving behind

a somewhat mixed pattern. This could indicate that a block-diagonal specification might model irrelevant correlations whereas a strict diagonal specification still ignores important information.

Both the diagonal and block-diagonal sparsity restrictions are imposed ex ante and are not changed during the analysis. In order to investigate whether the sparsity pattern is time-varying we plot the residual pattern for the 12 factor case at different points in time in panel (e). The results indicate that the pattern is rather constant over time.

3.5. Out-of-sample Forecasting Analysis

This section reports the out-of-sample forecasting results. Selected in-sample parameter estimation results for the proposed CFSS model are presented in Appendix B.2.

3.5.1. Implementation

We analyze the out-of-sample performance of three different factor structures: a 1-factor model with the market factor only (1F), a 3-factor model based on the Fama-French factors (3F), and a 12-factor model including the Fama-French factors plus the nine sector-specific ETFs (12F). For both the realized factor and residual covariance matrices we fit the unrestricted UE model (R0) and compare its performance to the restricted models (R1), (R2) and (R3) within the CFSS framework. The d.o.f. parameters k_i^b for the stochastic volatility process of the (latent) betas are fixed to their cross-sectional median values based on the in-sample estimates. For completeness, we consider diagonal (D) and block-diagonal residual assumptions based on the GICS classification (S) for each of the factor structures, although this is only (approximately) supported under the 12-factor specification. The diagonal specification sets $S = p$ and $p_i = 1 \forall i$, the block-diagonal specification sets $S = 9$ with sector sizes given in Table 3.1. Motivated by the findings of Figure 3.3, we also consider post-prediction shrinkage of the sector-blocks towards their diagonal. We use a linear shrinkage (LS)

approach akin to Ledoit and Wolf (2003, 2004). The LS prediction reads as

$$\widetilde{\widehat{C}^e_{it+1}} = \alpha_i \left\{ \widehat{C}^e_{it+1} \odot I_{p_i} \right\} + (1 - \alpha_i) \widehat{C}^e_{it+1}, \tag{3.21}$$

where $\odot$ denotes the Hadamard product (see Lütkepohl, 1996, p. 3). The optimal shrinkage intensity $\alpha_i^\star$ is found by minimizing the expected Frobenius distance between the in-sample LS prediction and the realized residual blocks.[15]

The first three years of the data set are used as in-sample and the remainder as out-of-sample periods. We employ a rolling-window size of three years to obtain one-day ($h = 1$), one-week ($h = 5$), two-weeks ($h = 10$) and one-month ($h = 22$) ahead predictions. The daily forecasts are obtained by re-estimating each model on a daily basis and forming predictions via Eq. (3.20). For the multi-step ahead forecasts we rely on a direct forecasting approach, i.e., we use a horizon-specific estimated model where the dependent variable is the multi-period ahead value being predicted (Marcellino et al., 2006).[16] Consequently, we predict the cumulative h-day ahead covariance matrix $C^r_{t+h} = \sum_{i=1}^h C^r_{t+i}$ by applying the CFSS models to the aggregated series $C^f_{t+h} = \sum_{i=1}^h C^f_{t+i}$, $C^e_{t+h} = \sum_{i=1}^h C^e_{t+i}$ and $B_{t+h} = \frac{1}{h} \sum_{i=1}^h B_{t+i}$ and again using Eq. (3.20). Following Chiriac and Voev (2011) we aggregate the realized components of C^r_t using non-overlapping h-day windows.

The sample period includes the financial crisis of 2008 as well as flash crashes in 2010 and 2011. These events lead to the presence of outliers in the (co)variance time-series which could distort parameter estimation results. In order to mitigate the effect of this comparably extreme events it is nowadays a conventional approach to perform a slight ex-post cleaning on the realized covariance matrices (Callot et al., 2017; Brito et al., 2018). As a means for cleaning the estimation sample we rely on the method proposed in Callot et al. (2017). Each day for which at least 25% of the unique elements of the realized covariance matrix exceed their sample mean up to then by

[15] That is $\alpha_i^\star = \arg\min_{\alpha_i} \mathrm{E}[\|\alpha_i \{\widehat{C}^e_{it} \odot I_{p_i}\} + (1 - \alpha_i)\widehat{C}^e_{it} - C^e_{it}\|_F^2]$.

[16] The direct forecasting approach is known to be more robust than the iterated approach as it is less prone to error propagation, which is especially relevant if there exists a model misspecification problem (Andersen et al., 2003; Chiriac and Voev, 2011; Bollerslev et al., 2018; Luo and Chen, 2020).

more than four standard errors are flagged for censoring. The flagged matrices are then replaced by the sample average of their ten nearest non-flagged neighbors. In total 22 covariance matrices are flagged, which corresponds to only 1.46% of the whole sample.

3.5.2. Competing models

In order to compare the out-of-sample forecasting performance of the proposed composite factor state-space model we consider five alternative state-of-the-art forecasting approaches as benchmarks:

1) The Factor LASSO approach of Brito et al. (2018),

2) the Factor HEAVY (FHEAVY) model of Sheppard and Xu (2019),[17]

3) the Realized consistent DCC (Re-cDCC) model of Bauwens et al. (2016),

4) the Exponentially Weighted Moving Average (EWMA; Morgan, 1996),

5) and the random walk (RW) model.

While the Factor LASSO and HEAVY models can be seen as natural competitors, the (non-factor) Re-cDCC is known to be a hard-to-beat benchmark when predicting realized asset covariance matrices. The EWMA and RW models can be seen as standard industry practice. Models 1) and 2) use the decomposition in Eq. (3.20) to predict C^r_{t+1}, models 3) to 5) are directly applied to the series of realized asset covariance matrices.

The LASSO (and also the EWMA and RW) approach can be implemented as described in Brito et al. (2018). To implement the HEAVY and Re-cDCC model, however, minor restrictions are required for the high-dimensional asset (and factor) setting as considered here. For the factor part in the multi-factor HEAVY model we implement a scalar version of the CAW model (see Golosnoy et al., 2012) and apply

[17] Since our focus lies on forecasting the realized measures, we ignore the HEAVY-P equations which link the realized covariance forecast to the conditional (latent) return covariance (see Sheppard and Xu, 2019).

a covariance targeting approach (see, e.g., Noureldin et al., 2012) to get rid of the constant part. For the Re-cDCC model we consider the scalar specification for the underlying correlation dynamics. Potential unreliability in the parameter estimation arising from numerically calculating determinants and inverting huge dimensional matrices is circumvented by further restricting the correlation part using the dynamic equicorrelation (DECO) approach. In the DECO model both the determinant and inverse of the correlation matrix are given by closed form expressions (see also Engle and Kelly, 2012). The resulting model (sRe-cDECO hereafter) is advocated by Bauwens et al. (2016) as valuable competitor when handling vast dimensional systems.

Following, e.g., Callot et al. (2017), Opschoor et al. (2017) and Gribisch et al. (2020), we set the EWMA smoothing parameter to 0.96.

3.5.3. Statistical forecast evaluation

We assess the accuracy of the direct h-step-ahead point forecast with two types of statistical loss functions. First, the Root Mean Squared Error (RMSE) based on the Frobenius Norm of the forecasting error that is calculated by comparing $\widehat{C}^r_{t+h}$ and the ex-post observed value C^r_{t+h}, is considered (cf. Ledoit et al., 2003). This RMSE is given by

$$\text{RMSE} = \frac{1}{T^\star} \sum_t \left[\sum_i (c^r_{it+h} - \hat{c}^r_{it+h})^2 + 2 \sum_{i<j} (c^r_{ijt+h} - \hat{c}^r_{ijt+h})^2 \right]^{\frac{1}{2}}, \tag{3.22}$$

where c_{it+h} and c_{ijt+h} denote the realized variance of asset i and the realized covariance between asset i and j, respectively. The corresponding forecasts are denoted by $\widehat{c}_{it+h}$ and $\widehat{c}_{ijt+h}$, and $T^\star$ denotes the number of forecasting periods. In order to see whether a model performs different w.r.t. the different elements in the covariance matrix we follow Gribisch et al. (2020) and disentangle the RMSE from above into the following

two loss functions

$$\text{RMSE}^{\text{v}} = \frac{1}{T^\star}\sum_t \left[\sum_i (c^r_{it+h} - \hat{c}^r_{it+h})^2\right]^{\frac{1}{2}}, \quad \text{RMSE}^{\text{c}} = \frac{1}{T^\star}\sum_t \left[\sum_{i<j} (c^r_{ijt+h} - \hat{c}^r_{ijt+h})^2\right]^{\frac{1}{2}}, \tag{3.23}$$

where RMSE^{v} uses the variances only and RMSE^{c} considers the covariances. In addition to the RMSE loss we consider the commonly used QLIKE loss function, i.e.,

$$\text{QLIKE} = \frac{1}{T^\star}\sum_t \log|\widehat{C}^r_{t+h}| + \text{tr}\left[(\widehat{C}^r_{t+h})^{-1} C^r_{t+h}\right]. \tag{3.24}$$

While the RMSE losses are symmetric, the QLIKE loss function is an asymmetric loss, penalizing under-predictions more heavily (cf. Luo and Chen, 2020). Both loss functions are known to be robust to noisy (co)variance proxies (Patton, 2011; Laurent et al., 2013; Sheppard and Xu, 2019).

In order to evaluate the statistical significance of differences in the RMSE, RMSE^{v}, RMSE^{c} and QLIKE losses across models, we employ the Model Confidence Set (MCS) approach of Hansen et al. (2011). The MCS is computed for the 75% confidence level using a block bootstrap with block length $\lfloor (T^\star)^{1/3} \rfloor$ and 10,000 bootstrap replications.

Table 3.2 reports the RMSE and QLIKE results for the short-, mid- and long-term out-of-sample forecasts. We first compare the performances of the different CFSS model specifications. For one-step and five-step ahead forecasts the 12-factor CFSS model clearly outperforms the single and three-factor settings across all loss functions, regardless of the residual specification considered (though the losses suggest that block-diagonal specifications S and LS do increase the forecast precision). The biweekly and monthly results reveal that the three-factor models with diagonal and linear-shrinkage-based residuals perform best in predicting long-term covariance matrices, especially for the covariance elements as indicated by RMSE^{c}. In terms of QLIKE loss, however, including sector ETF based factors still yields gains in forecast precision. Furthermore, we see that imposing restrictions on the smoothing parameter

helps to increase predictive accuracy – while restriction (R2) solely convinces for the QLIKE measure, restriction (R1) performs overall well. Now, we turn to a comparison with the nine competing models. Most striking is that for the 'full' RMSE loss all competitors are significantly outperformed by the CFSS model specifications. In particular, for daily and weekly horizons solely the UE-R1-LS specification belongs to the 75% MCS.[18] Similar results can be found for the QLIKE loss, where the UE-R2-LS specification significantly outperforms all competitors for daily forecasts. For multi-step ahead forecasts the sRe-cDECO model and further CFSS specifications, however, perform equally well. The EWMA and RW models are significantly outperformed in any case.

3.5.4. Global-minimum-variance-portfolio forecasts

In this section we turn to an economic application in out-of-sample portfolio construction. We analyze the forecasting performance for the CFSS model and its competitors based on several variations of the global minimum variance portfolio (GMVP).

For a given asset covariance forecast $\widehat{C}^r_{t+h}$ computed in period t, the (unrestricted) GMVP is the solution to the minimization problem

$$\widehat{w}_{t+h} = \arg\min_{w_{t+h}} w'_{t+h}\widehat{C}^r_{t+h}w_{t+h}, \qquad \text{subject to } w'_{t+h}\iota = 1, \tag{3.25}$$

where w_{t+h} is the $p \times 1$ vector of portfolio weights and ι is a $p \times 1$ vector of ones. If the asset covariance forecast is positive definite, the latter minimization problem has the unique solution $\widehat{w}_{t+h} = (\widehat{C}^r_{t+h})^{-1}\iota / (\iota'(\widehat{C}^r_{t+h})^{-1}\iota)$.

In large scale applications as considered here, it might occur that the covariance prediction is not well conditioned, possibly resulting in unrealistically volatile weights due to the inversion of $\widehat{C}^r_{t+h}$. For the factor model predictions based on Eq. (3.20) this problem is mitigated. Due to the (block)diagonality of the residual covariance prediction, the p dimensional inverse problem collapses to $\max\{q, p_1, \ldots, p_s\}$ dimen-

[18] The 75% and 90% MCS are congruent in this case, however, the 90% MCS results are not reported here.

sions by making use of the Woodbury matrix inversion lemma (see Lütkepohl, 1996, pp. 29–30).[19] Another popular approach to obtain a more stable portfolio is to impose further restrictions on the weights. Following Lunde et al. (2016) we consider short-selling constraints as well as upper and lower bounds for w_{it+h}, $i = 1, \ldots, p$. These additional restrictions are imposed to the minimization problem in Eq. (3.25) by including the following norm constraints

$$\begin{aligned} \|w_{t+h}\|_1 &\leq 1 + 2s, \\ \|w_{t+h}\|_\infty &\leq u, \end{aligned} \tag{3.26}$$

where $\|w_{t+h}\|_1 = \sum_{i=1}^{p} |w_{it+h}|$, and $\|w_{t+h}\|_\infty = \max_{1 \leq i \leq p} |w_{it+h}|$ denote the L_1 and L_∞ norm of the weight vector, respectively. Here, $s \in [0, 1]$ denotes the percentage that is allowed to be held short, and $u > 0$ denotes the boundary for every distinct position. The boundary constraint ensures well-diversified portfolios, since the number of assets included is at least $\lfloor 1/u \rfloor$ (see also Fan et al., 2012; Lunde et al., 2016, for more detailed discussions).[20]

In the present paper we follow Brito et al. (2018) and Lunde et al. (2016) and consider – besides the unrestricted GMVP – three restricted minimum variance problems, i.e., a *150/50 portfolio* ($s = 0.50$), a *130/30 portfolio* ($s = 0.30$) and a *long-only portfolio* ($s = 0.00$). We focus on daily portfolios and consider weekly horizons as robustness check. The latter portfolios are constructed using a boundary value of $u = 0.10$ for each weight, such that at least ten assets are included. For assessing the relative capabilities of the competing models for optimal portfolio allocation, we calculate their out-of-sample portfolio returns $r_t^p = \widehat{w}_t' r_t$ and report five measures, i.e., the portfolio standard deviation

$$\sigma_p = \sqrt{\frac{1}{T^\star} \sum_t (r_t^p - \mu_p)^2}, \qquad \text{with } \mu_p = \frac{1}{T^\star} \sum_t r_t^p, \tag{3.27}$$

[19] The inverse of C_t^r is then given by $(C_t^r)^{-1} = (C_t^e)^{-1} - (C_t^e)^{-1} B_t [(C_t^f)^{-1} + B_t' (C_t^e)^{-1} B_t]^{-1} B_t' (C_t^e)^{-1}$.

[20] Unfortunately, the minimization problem in Eqs. (3.25) and (3.26) has no closed-form solution. We therefore rely on the CVX package in Matlab of Grant and Boyd (2014) to solve for $\widehat{w}_{t+h}$.

the Sharpe ratio

$$SR = \mu_p/\sigma_p, \tag{3.28}$$

the average portfolio concentration

$$CO = \frac{1}{T^\star}\sum_t \left(\sum_{i=1}^{p} \widehat{w}_{it}^2\right)^{1/2}, \tag{3.29}$$

the average total short positions

$$SP = \frac{1}{T^\star}\sum_t \sum_{i=1}^{p} |w_{it}|\mathbb{I}_{w_{it}<0}, \tag{3.30}$$

and the average turnover rates

$$TO = \frac{1}{T^\star}\sum_t \left|\widehat{w}_{it} - \widehat{w}_{it-1}\frac{1+r_{it-1}}{1+r_{t-1}^p}\right|. \tag{3.31}$$

The reported portfolio standard deviations and Sharpe ratios are annualized by multiplication with $\sqrt{\lfloor 252/h \rfloor}$. For CO, SP and TO smaller values are favorable: whereas high concentration implies too little diversification of the portfolio possibly resulting from large prediction errors of C_{t+h}^r, verifying to which extent short sale constraints would be violated is of practical relevance, since many portfolio managers are prohibited from taking such positions (Hautsch et al., 2011). Furthermore, high turnover values can deteriorate the portfolio performance when facing transaction costs (Bauwens et al., 2016; Bollerslev et al., 2018). Statistical significance of the differences in portfolio standard deviations is assessed by the MCS approach using the squared demeaned portfolio returns as loss series.

The daily portfolio allocation results are collected in Table 3.3. For the unrestricted GMVP the EWMA has the lowest standard deviation. However, the 1F-S and 3F-S CFSS models cannot be outperformed significantly while yielding higher Sharpe ratios and more diversified portfolios with less extreme short positions. By impos-

ing 50% short-selling constraints the performance of the 12F and LS-based residual models improves with comparably lower turnover for the CFSS models. By imposing stronger (30% and 0%) short-selling constraints the portfolio standard deviations get hardly distinguishable. Though, comparing the CFSS model specifications to the LASSO and HEAVY competitors we find better performance regarding the SR and TO results for the CFSS. Overall, the lowest turnover is obtained with the EWMA. The highest turnover is obtained for the random walk, followed by the Factor LASSO specifications. The weekly portfolio results are summarized in Table 3.4. They indicate similar allocations. The lowest standard deviations are now obtained by the sRe-cDECO and EWMA. However the CFSS with block residual belongs to the MCS in every scenario, whereas the factor HEAVY is only included for long-only portfolios.

3.5.5. Mean-variance-portfolio forecasts

As a second out-of-sample portfolio allocation exercise we consider a mean-variance portfolio (MVP) based on an investor who aims at minimizing the portfolio risk subject to a target portfolio return. Hence, the GMVP problem in Eq. (3.25) has to be augmented as follows

$$\widehat{w}_{t+h} = \arg\min_{w_{t+h}} w'_{t+h}\widehat{C}^r_{t+h}w_{t+h}, \quad \text{subject to } \; w'_{t+h}\iota = 1, \; w'_{t+h}m = \mu_{target}, \tag{3.32}$$

where m is the signal variable and μ_{target} is the target return. For positive definite $\widehat{C}^r_{t+h}$ the problem has solution given by

$$\widehat{w}_{t+h} = (\widehat{C}^r_{t+h})^{-1}\frac{m(D\mu_{target} - E) - \iota(F - E\mu_{target})}{FD - E^2},$$

where $D = \iota'(\widehat{C}^r_{t+h})^{-1}\iota$, $E = m'(\widehat{C}^r_{t+h})^{-1}\iota$, $F = m'(\widehat{C}^r_{t+h})^{-1}m$.

Various approaches exist to construct the signal and to choose the target return. For example, Callot et al. (2017) compute m by a moving average of 100 days and fix μ_{target} arbitrarily. Chiriac and Voev (2011) consider varying target returns to construct the efficiency frontier. Here, we follow the more recent approach of Engle et al. (2019)

and Moura et al. (2020) and construct the signal m using the momentum factor of Jegadeesh and Titman (1993). For each of the p stocks the individual momentum m_i is computed as the geometric average of the previous 252 returns, but excluding the 22 most recent returns. Collecting all the momentums in a vector yields the signal m. The target return is computed as the arithmetic average of the momentums of those stocks that belong to the top-quintile stocks ranked according to momentum.

We combine the problem in Eq. (3.32) with the norm constraints in Eq. (3.26) to construct four similar portfolios as in the GMVP exercise. Again, we report daily allocation results, which are collected in Table 3.5. The results are qualitatively similar to those obtained in the GMVP exercise. For the unrestricted allocations the EWMA shows the lowest standard deviation. The second best models included in the MCS are the CFSS 1F-R1-S, 1F-R2-S and 3F-R2-S specifications, respectively. Compared to the EWMA all three have higher SR as well as lower CO and SP. Imposing short-selling constraints ameliorates the performance of the 12F CFSS model specifications in terms of lower standard deviations and lower turnover values than the 1F and 3F specifications. As robustness check, Table 3.6 shows weekly allocation results. The results do not differ substantially in comparison to the daily allocations.

3.6. Conclusion

We propose a new factor state-space approach for the composite prediction of vast-dimensional realized covariance matrices of asset returns. The composite factor state-space (CFSS) model exploits an observed factor structure of the realized covariances of asset returns. Its components are found by a block LDL decomposition of the joint realized covariance matrix of the observed risk factors and assets for each point in time, yielding individual time-series for the realized factor covariances, the realized residual covariances and the matrices of realized factor loadings.

A key idea of the proposed approach is to model and predict the individual components of the factor decomposition separately. This reduces model complexity and

allows for fast parameter estimation and prediction of future realizations of asset covariance matrices in scalable dimensions. A further advantage of the proposed approach is that the residual components can straightforwardly be modeled and predicted akin to an approximate factor model, which is empirically more realistic than imposing diagonality assumptions.

For the realized factors and residual components we adopt matrix-variate state-space models in the vein of Windle and Carvalho (2014), and combine independent Wishart measurement densities with Matrix-Beta type-I transition equations. This state-space framework is particularly appealing as it results in closed-form expressions for the predictive densities of the realized factor and residual covariances and yields positive-definite predictions without requiring any transformations like the Cholesky decomposition or the matrix-logarithm. For the realized factor loadings we rely on well-established models from the TVP-VAR literature, which incorporate Wishart stochastic volatility processes.

We apply the factor model to a data set of daily realized covariance matrices for 225 NYSE traded stocks. The observed risk factors we consider are the CAPM market factor, the Fama and French (1993) HML and SMB factors as well as sector-specific ETFs. It turns out that for justifying an approximate factor structure based on sector-blocks for the residual components it is critical to include all those factors. The fitted factor models are estimated within only a few minutes.

In an extensive out-of-sample application the CFSS model shows superior forecasting performance and outperforms its competitors in almost all forecast horizons. We conclude that the CFSS approach is a valuable tool for modeling and forecasting time-series of vast-dimensional realized covariance matrices.

3.7. Tables and Figures

Table 3.1.: Descriptive statistics for the data set.

Sector	ETF	Members	10%	Mean	Median	90%
Energy (E)	XLE	16 (–)	1.27	6.03	3.20	11.64
Materials (M)	XLB	9 (–)	0.90	4.80	2.51	9.97
Industrials (I)	XLI	26 (8)	0.75	3.95	2.10	8.16
Consumer Discretionary (D)	XLY	35 (9)	0.98	5.87	2.88	13.33
Consumer Staples (S)	XLP	20 (7)	0.39	2.07	1.04	4.06
Health Care (H)	XLV	25 (7)	0.55	2.98	1.50	6.18
Financials (F)	XLF	34 (7)	0.82	9.45	2.95	19.85
Information Technologies (T)	XLK	44 (22)	1.10	4.74	2.85	9.71
Utilities (U)	XLU	16 (–)	0.45	2.71	1.27	5.05

Note: The table lists the sectors covered by the data set. In column ETF the the symbols of the respective sector ETF is stated. Members denote the number of stocks per sector as classified through their GICS code. The number of assets in the 60D data set of Gribisch et al. (2020) is given in parentheses. The descriptive statistics are average mean and median asset volatility, as well as average lower and upper 10% asset volatility.

Table 3.2.: Evaluation of forecast accuracy.

	$h=1$				$h=5$				$h=10$				$h=22$			
Model	RMSE	RMSE^{v}	RMSE^{c}	QLIKE	RMSE	RMSE^{v}	RMSE^{c}	QLIKE	RMSE	RMSE^{v}	RMSE^{c}	QLIKE	RMSE	RMSE^{v}	RMSE^{c}	QLIKE
CFSS 1F																
UE-R0-D	138.90	31.92	94.24	294.21	117.90	24.36	80.60	133.96	120.34	23.39	82.51	83.18	138.43	25.51	95.30	46.00
UE-R0-S	138.07	31.85	93.79	283.42	118.47	23.96	81.34	131.41	123.69	22.78	85.43	81.70	145.59	24.20	101.11	45.22
UE-R0-LS	136.79	31.85	92.80	281.46	116.55	23.96	79.87	131.12	120.79	22.78	83.23	81.59	140.71	24.20	97.44	45.23
UE-R1-D	140.46	31.24	95.57	293.60	119.60	23.48	82.06	133.86	120.83	22.10	83.19	83.13	135.24	22.66	93.52	45.93
UE-R1-S	138.59	31.18	94.32	284.31	118.68	23.23	81.63	131.58	122.40	21.79	84.68	81.78	140.48	22.56	97.66	45.24
UE-R1-LS	137.85	31.18	93.75	282.02	117.52	23.23	80.74	131.22	120.57	21.79	83.28	81.62	137.02	22.56	95.04	45.19
UE-R2-D	137.20	32.60	92.82	295.43	116.03	25.53	78.96	134.20	120.90	25.29	82.47	83.34	145.56	28.77	99.82	46.16
UE-R2-S	137.91	32.60	93.53	282.74	119.09	25.13	81.56	131.27	127.32	24.58	87.75	81.63	155.76	27.33	107.94	45.25
UE-R2-LS	135.86	32.60	91.95	281.93	115.84	25.13	79.08	131.21	122.59	24.58	84.15	81.68	148.58	27.33	102.57	45.39
UE-R3-D	138.56	32.00	93.97	294.35	117.37	24.51	80.18	134.00	120.06	23.64	82.26	83.23	138.61	25.50	95.41	46.03
UE-R3-S	137.93	31.94	93.67	283.34	118.36	24.15	81.22	131.39	123.97	23.07	85.58	81.69	146.50	24.55	101.69	45.22
UE-R3-LS	136.57	31.94	92.62	281.46	116.26	24.15	79.61	131.12	120.80	23.07	83.17	81.60	141.19	24.55	97.69	45.25
CFSS 3F																
UE-R0-D	138.44	31.51	94.01	288.41	117.68	23.84	80.58	132.79	119.26	22.73	81.89	82.60	132.78	24.27	91.50	45.71
UE-R0-S	137.65	31.41	93.57	282.69	117.99	23.53	81.07	131.34	121.93	22.20	84.26	81.68	138.57	23.12	96.22	45.19
UE-R0-LS	136.72	31.41	92.86	280.63	116.61	23.53	80.01	131.02	119.74	22.20	82.59	81.55	134.83	23.12	93.41	45.17
UE-R1-D	140.28	30.95	95.51	287.99	119.58	23.13	82.14	132.73	120.34	21.64	82.95	82.58	132.05	21.83	91.45	45.66
UE-R1-S	138.53	30.90	94.33	283.80	118.55	22.95	81.58	131.55	121.20	21.38	83.87	81.79	135.28	21.74	94.04	45.23
UE-R1-LS	138.03	30.90	93.94	281.54	117.74	22.95	80.96	131.19	119.85	21.38	82.84	81.62	132.80	21.74	92.16	45.17
UE-R2-D	136.83	32.08	92.69	289.44	115.82	24.79	78.99	132.99	119.44	24.32	81.68	82.74	138.03	27.13	94.72	45.86
UE-R2-S	137.38	32.02	93.26	281.76	118.19	24.46	81.04	131.13	124.76	23.68	86.06	81.57	146.61	25.79	101.59	45.20
UE-R2-LS	135.86	32.02	92.09	280.58	115.82	24.46	79.22	130.98	121.15	23.68	83.32	81.56	140.98	25.79	97.37	45.29
UE-R3-D	138.21	31.56	93.82	288.54	117.30	23.94	80.28	132.83	118.90	22.87	81.59	82.64	132.66	24.18	91.38	45.74
UE-R3-S	137.54	31.47	93.48	282.58	117.86	23.65	80.95	131.30	121.90	22.38	84.19	81.66	138.87	23.32	96.38	45.19
UE-R3-LS	136.57	31.47	92.73	280.58	116.36	23.65	79.80	131.00	119.51	22.38	82.38	81.55	134.79	23.32	93.30	45.19
CFSS 12F																
UE-R0-D	132.52	30.98	89.99	280.83	115.17	23.29	79.10	131.36	124.98	22.04	86.56	81.81	150.81	23.27	105.15	45.28
UE-R0-S	132.53	30.82	90.06	279.58	115.31	23.11	79.24	130.89	125.42	21.66	86.95	81.47	151.83	22.37	106.00	45.08
UE-R0-LS	132.34	30.82	89.92	278.41	115.10	23.11	79.09	130.75	125.13	21.66	86.74	81.42	151.34	22.37	105.65	45.07
UE-R1-D	131.81	30.59	89.56	281.76	112.90	22.80	77.56	131.57	120.45	21.23	83.44	81.93	142.63	21.41	99.54	45.31
UE-R1-S	131.60	30.53	89.44	281.10	112.73	22.73	77.45	131.19	120.55	21.08	83.53	81.63	143.32	21.34	100.04	45.15
UE-R1-LS	131.51	30.53	89.37	279.89	112.63	22.73	77.37	131.04	120.40	21.08	83.42	81.56	143.03	21.34	99.84	45.12
UE-R2-D	134.51	31.40	91.37	280.13	119.30	23.97	81.97	131.13	132.68	23.23	91.94	81.69	164.29	25.74	114.43	45.28
UE-R2-S	134.88	31.22	91.71	278.06	120.02	23.73	82.57	130.52	133.78	22.73	92.84	81.26	165.91	24.52	115.80	45.03
UE-R2-LS	134.56	31.22	91.48	277.23	119.61	23.73	82.27	130.43	133.23	22.73	92.44	81.24	165.06	24.52	115.19	45.04
UE-R3-D	132.72	31.02	90.13	280.72	115.72	23.37	79.49	131.31	125.97	22.16	87.25	81.80	152.68	23.26	106.48	45.27
UE-R3-S	132.78	30.87	90.24	279.40	115.97	23.20	79.71	130.82	126.54	21.80	87.74	81.42	153.86	22.55	107.43	45.06
UE-R3-LS	132.58	30.87	90.09	278.25	115.74	23.20	79.54	130.68	126.21	21.80	87.50	81.37	153.32	22.55	107.04	45.05
Factor Lasso 1F	142.07	30.45	97.03	284.22	117.07	22.98	80.56	131.12	124.51	26.80	84.96	82.24	221.70	86.69	135.58	48.32
Factor Lasso 3F	150.61	30.81	103.09	284.12	123.06	22.80	84.87	131.05	134.02	25.68	92.16	82.13	206.17	57.43	136.20	48.15
Factor Lasso 12F	136.67	30.74	93.09	282.83	113.67	22.80	78.16	130.66	229.97	31.39	160.57	81.95	280.82	92.48	181.21	48.03
Factor HEAVY 1F	138.86	31.31	94.31	295.49	129.61	41.25	83.06	140.99	130.55	45.46	81.36	86.84	140.09	47.86	88.25	47.75
Factor HEAVY 3F	142.63	30.87	97.14	290.54	139.20	35.56	92.01	139.22	136.39	38.47	88.51	85.90	138.84	40.31	88.91	47.32
Factor HEAVY 12F	131.89	30.57	89.67	279.93	118.17	34.69	77.73	135.50	128.76	36.52	84.88	83.79	156.11	38.58	105.06	46.30
sRe-cDECO	174.45	33.39	120.49	279.06	157.91	26.57	109.82	130.48	166.63	26.05	116.24	81.23	189.63	28.83	132.45	45.14
EWMA	155.66	33.14	106.67	283.97	180.53	33.04	125.13	131.57	214.65	40.80	148.65	82.65	264.52	56.00	182.43	46.72
RW	154.12	37.66	104.48	328.31	124.78	26.86	85.41	135.56	131.79	24.09	91.05	83.14	158.97	23.68	110.91	45.58

Note: The table reports the RMSE, RMSE^{v}, RMSE^{c} and QLIKE losses as given in Eqs. (3.22) to (3.24) for the forecast horizon $h \in \{1, 5, 10, 22\}$. To make the results more comparable the losses have been divided by h. Grey-shaded cells indicate that the respective model belongs to the 75% model confidence set.

Table 3.3.: Daily GMVP forecasting results.

	Unrestricted GMVP					Restricted GMVP: 150/50					Restricted GMVP: 130/30					Restricted GMVP: Long-only				
Model	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO
CFSS 1F																				
UE-R0-D	7.73	1.91	0.04	0.48	0.30	7.62	1.90	0.04	0.44	0.30	7.35	1.89	0.04	0.30	0.28	7.75	1.27	0.05	0.00	0.20
UE-R0-S	6.62	2.54	0.06	0.68	0.60	6.66	2.47	0.06	0.48	0.53	6.74	2.33	0.05	0.30	0.45	7.78	1.54	0.05	0.00	0.30
UE-R0-LS	6.66	2.50	0.06	0.62	0.54	6.69	2.43	0.05	0.47	0.48	6.74	2.30	0.05	0.30	0.42	7.74	1.52	0.05	0.00	0.28
UE-R1-D	7.72	1.92	0.04	0.48	0.29	7.61	1.91	0.04	0.44	0.29	7.34	1.89	0.04	0.30	0.27	7.75	1.26	0.05	0.00	0.19
UE-R1-S	6.62	2.55	0.06	0.68	0.57	6.66	2.47	0.06	0.48	0.50	6.74	2.34	0.05	0.30	0.42	7.78	1.55	0.05	0.00	0.28
UE-R1-LS	6.65	2.51	0.06	0.64	0.52	6.68	2.45	0.05	0.48	0.47	6.74	2.31	0.05	0.30	0.40	7.75	1.53	0.05	0.00	0.27
UE-R2-D	7.75	1.91	0.04	0.48	0.29	7.63	1.90	0.04	0.44	0.28	7.36	1.88	0.04	0.30	0.27	7.75	1.26	0.05	0.00	0.19
UE-R2-S	6.61	2.55	0.06	0.67	0.57	6.66	2.47	0.06	0.48	0.50	6.74	2.33	0.05	0.30	0.42	7.78	1.54	0.05	0.00	0.28
UE-R2-LS	6.67	2.48	0.05	0.61	0.49	6.70	2.43	0.05	0.47	0.45	6.75	2.29	0.05	0.30	0.39	7.73	1.52	0.05	0.00	0.26
UE-R3-D	7.73	1.91	0.04	0.48	0.30	7.62	1.91	0.04	0.44	0.30	7.35	1.89	0.04	0.30	0.28	7.75	1.27	0.05	0.00	0.20
UE-R3-S	6.62	2.55	0.06	0.68	0.60	6.66	2.47	0.06	0.48	0.53	6.74	2.33	0.05	0.30	0.45	7.78	1.54	0.05	0.00	0.30
UE-R3-LS	6.66	2.49	0.06	0.62	0.53	6.69	2.43	0.05	0.47	0.48	6.74	2.30	0.05	0.30	0.42	7.74	1.52	0.05	0.00	0.28
CFSS 3F																				
UE-R0-D	7.31	2.24	0.04	0.48	0.31	7.26	2.18	0.04	0.44	0.30	7.10	2.13	0.04	0.30	0.26	7.71	1.30	0.05	0.00	0.17
UE-R0-S	6.61	2.53	0.06	0.68	0.57	6.66	2.49	0.06	0.48	0.49	6.74	2.36	0.05	0.30	0.40	7.76	1.56	0.05	0.00	0.26
UE-R0-LS	6.64	2.49	0.06	0.63	0.51	6.68	2.45	0.05	0.48	0.45	6.74	2.34	0.05	0.30	0.38	7.72	1.54	0.05	0.00	0.25
UE-R1-D	7.30	2.24	0.04	0.48	0.30	7.25	2.18	0.04	0.43	0.29	7.09	2.13	0.04	0.30	0.25	7.72	1.30	0.05	0.00	0.16
UE-R1-S	6.61	2.53	0.06	0.68	0.54	6.66	2.49	0.06	0.48	0.46	6.74	2.37	0.05	0.30	0.38	7.76	1.56	0.05	0.00	0.25
UE-R1-LS	6.63	2.50	0.06	0.64	0.49	6.68	2.47	0.05	0.48	0.44	6.74	2.35	0.05	0.30	0.37	7.73	1.54	0.05	0.00	0.24
UE-R2-D	7.32	2.24	0.04	0.48	0.30	7.26	2.18	0.04	0.44	0.29	7.10	2.13	0.05	0.30	0.25	7.72	1.30	0.05	0.00	0.16
UE-R2-S	6.61	2.53	0.06	0.68	0.53	6.66	2.49	0.06	0.48	0.46	6.74	2.37	0.05	0.30	0.38	7.76	1.56	0.05	0.00	0.24
UE-R2-LS	6.65	2.48	0.06	0.62	0.46	6.69	2.45	0.05	0.47	0.42	6.74	2.33	0.05	0.30	0.35	7.71	1.53	0.05	0.00	0.23
UE-R3-D	7.31	2.24	0.04	0.48	0.31	7.26	2.18	0.04	0.44	0.30	7.10	2.13	0.04	0.30	0.26	7.71	1.30	0.05	0.00	0.17
UE-R3-S	6.61	2.53	0.06	0.68	0.57	6.66	2.49	0.06	0.48	0.49	6.74	2.36	0.05	0.30	0.40	7.76	1.56	0.05	0.00	0.26
UE-R3-LS	6.64	2.49	0.06	0.63	0.50	6.68	2.45	0.05	0.48	0.45	6.74	2.34	0.05	0.30	0.38	7.72	1.54	0.05	0.00	0.25
CFSS 12F																				
UE-R0-D	6.85	2.47	0.08	0.76	0.40	6.76	2.40	0.07	0.50	0.33	6.74	2.26	0.07	0.30	0.26	7.66	1.42	0.07	0.00	0.15
UE-R0-S	6.66	2.47	0.08	0.82	0.52	6.65	2.36	0.07	0.50	0.41	6.63	2.24	0.07	0.30	0.33	7.66	1.50	0.07	0.00	0.19
UE-R0-LS	6.68	2.47	0.08	0.79	0.49	6.66	2.37	0.07	0.50	0.39	6.64	2.24	0.07	0.30	0.31	7.65	1.48	0.07	0.00	0.18
UE-R1-D	6.85	2.47	0.07	0.76	0.38	6.76	2.41	0.07	0.50	0.32	6.74	2.26	0.07	0.30	0.25	7.67	1.41	0.07	0.00	0.14
UE-R1-S	6.67	2.47	0.08	0.82	0.50	6.65	2.36	0.07	0.50	0.39	6.63	2.24	0.07	0.30	0.31	7.66	1.50	0.07	0.00	0.18
UE-R1-LS	6.68	2.47	0.08	0.80	0.47	6.66	2.37	0.07	0.50	0.37	6.64	2.24	0.07	0.30	0.30	7.65	1.49	0.07	0.00	0.18
UE-R2-D	6.85	2.47	0.08	0.76	0.38	6.77	2.40	0.07	0.50	0.31	6.74	2.25	0.07	0.30	0.25	7.67	1.41	0.07	0.00	0.14
UE-R2-S	6.66	2.47	0.08	0.82	0.49	6.65	2.36	0.07	0.50	0.39	6.63	2.24	0.07	0.30	0.31	7.66	1.50	0.06	0.00	0.18
UE-R2-LS	6.68	2.47	0.08	0.79	0.45	6.66	2.37	0.07	0.50	0.36	6.64	2.23	0.07	0.30	0.29	7.65	1.48	0.06	0.00	0.17
UE-R3-D	6.85	2.47	0.08	0.76	0.40	6.76	2.40	0.07	0.50	0.33	6.74	2.26	0.07	0.30	0.26	7.67	1.42	0.07	0.00	0.15
UE-R3-S	6.66	2.47	0.08	0.82	0.53	6.65	2.36	0.07	0.50	0.41	6.63	2.24	0.07	0.30	0.33	7.66	1.50	0.07	0.00	0.19
UE-R3-LS	6.68	2.47	0.08	0.79	0.49	6.66	2.37	0.07	0.50	0.39	6.64	2.24	0.07	0.30	0.31	7.65	1.48	0.07	0.00	0.18
Factor Lasso 1F	7.12	2.58	0.06	0.69	0.70	7.02	2.53	0.05	0.50	0.64	6.88	2.39	0.05	0.30	0.57	7.61	1.40	0.05	0.00	0.41
Factor Lasso 3F	6.93	2.58	0.06	0.66	0.69	6.88	2.51	0.05	0.50	0.62	6.82	2.38	0.05	0.30	0.54	7.65	1.49	0.05	0.00	0.38
Factor Lasso 12F	7.17	2.42	0.08	0.84	0.75	6.97	2.38	0.07	0.50	0.59	6.84	2.32	0.07	0.30	0.48	7.68	1.46	0.06	0.00	0.29
Factor HEAVY 1F	7.89	1.92	0.04	0.51	0.30	7.74	1.91	0.04	0.46	0.30	7.37	1.86	0.05	0.30	0.28	7.71	1.26	0.05	0.00	0.21
Factor HEAVY 3F	7.43	2.17	0.04	0.49	0.32	7.34	2.13	0.04	0.44	0.31	7.11	2.10	0.04	0.30	0.28	7.70	1.29	0.05	0.00	0.20
Factor HEAVY 12F	6.94	2.45	0.07	0.73	0.48	6.83	2.44	0.07	0.50	0.41	6.81	2.29	0.06	0.30	0.33	7.64	1.43	0.06	0.00	0.19
sRe-cDECO	7.20	2.46	0.11	1.02	0.47	7.14	2.18	0.09	0.50	0.31	7.20	2.04	0.08	0.30	0.24	8.12	1.41	0.07	0.00	0.13
EWMA	6.52	2.27	0.09	0.92	0.23	6.54	2.05	0.07	0.50	0.17	6.65	1.95	0.07	0.30	0.13	7.74	1.28	0.06	0.00	0.06
RW	7.15	2.39	0.08	0.68	2.59	7.18	2.23	0.07	0.48	2.33	7.40	1.99	0.07	0.30	2.05	8.56	1.11	0.07	0.00	1.34

Note: The table reports GMVP summary statistics for the four portfolios under consideration based on 754 one-step-ahead conditional covariance matrix predictions; σ_p stands for the annualized standard deviation of the portfolio returns. The average portfolio concentration CO, short position SP and turnover TO are given by Eqs. (3.29) to (3.31). Grey-shaded cells indicate that the respective model belongs to the 90% model confidence set.

Table 3.4.: Weekly GMVP forecasting results.

	Unrestricted GMVP					Restricted GMVP: 150/50					Restricted GMVP: 130/30					Restricted GMVP: Long-only				
Model	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO
CFSS 1F																				
UE-R0-D	8.81	1.38	0.04	0.48	0.41	8.70	1.41	0.04	0.44	0.41	8.32	1.44	0.04	0.30	0.38	7.68	1.12	0.05	0.00	0.28
UE-R0-S	7.34	2.02	0.06	0.66	0.68	7.32	2.09	0.05	0.48	0.60	7.25	2.01	0.05	0.30	0.52	7.69	1.33	0.05	0.00	0.35
UE-R0-LS	7.40	1.97	0.05	0.62	0.61	7.39	2.04	0.05	0.48	0.56	7.30	1.98	0.05	0.30	0.49	7.67	1.33	0.05	0.00	0.33
UE-R1-D	8.80	1.39	0.04	0.48	0.40	8.69	1.42	0.04	0.44	0.40	8.31	1.45	0.04	0.30	0.38	7.68	1.12	0.05	0.00	0.27
UE-R1-S	7.34	2.02	0.06	0.66	0.66	7.32	2.09	0.05	0.48	0.59	7.25	2.01	0.05	0.30	0.51	7.70	1.34	0.05	0.00	0.34
UE-R1-LS	7.39	1.98	0.05	0.62	0.61	7.38	2.05	0.05	0.48	0.56	7.29	1.99	0.05	0.30	0.48	7.69	1.33	0.05	0.00	0.33
UE-R2-D	8.83	1.38	0.04	0.49	0.40	8.72	1.41	0.04	0.44	0.40	8.33	1.44	0.04	0.30	0.38	7.69	1.12	0.05	0.00	0.27
UE-R2-S	7.35	2.01	0.06	0.67	0.66	7.34	2.08	0.05	0.49	0.59	7.26	2.00	0.05	0.30	0.50	7.69	1.33	0.05	0.00	0.34
UE-R2-LS	7.45	1.95	0.05	0.61	0.58	7.43	2.01	0.05	0.47	0.53	7.33	1.96	0.05	0.30	0.47	7.66	1.32	0.05	0.00	0.32
UE-R3-D	8.81	1.38	0.04	0.48	0.41	8.70	1.42	0.04	0.44	0.41	8.32	1.45	0.04	0.30	0.38	7.69	1.12	0.05	0.00	0.28
UE-R3-S	7.34	2.02	0.06	0.66	0.69	7.33	2.09	0.05	0.48	0.61	7.25	2.01	0.05	0.30	0.52	7.69	1.34	0.05	0.00	0.35
UE-R3-LS	7.41	1.97	0.05	0.62	0.61	7.40	2.03	0.05	0.48	0.56	7.30	1.98	0.05	0.30	0.49	7.67	1.33	0.05	0.00	0.34
CFSS 3F																				
UE-R0-D	8.37	1.59	0.04	0.48	0.42	8.29	1.60	0.04	0.44	0.41	8.04	1.60	0.04	0.30	0.36	7.64	1.09	0.05	0.00	0.23
UE-R0-S	7.32	2.00	0.06	0.66	0.63	7.29	2.07	0.05	0.49	0.55	7.23	2.01	0.05	0.30	0.46	7.69	1.32	0.05	0.00	0.29
UE-R0-LS	7.37	1.96	0.05	0.62	0.57	7.35	2.03	0.05	0.48	0.51	7.27	1.98	0.05	0.30	0.43	7.66	1.31	0.05	0.00	0.28
UE-R1-D	8.37	1.59	0.04	0.47	0.41	8.29	1.60	0.04	0.44	0.40	8.04	1.60	0.04	0.30	0.35	7.63	1.09	0.05	0.00	0.22
UE-R1-S	7.32	2.00	0.06	0.66	0.61	7.29	2.07	0.05	0.49	0.53	7.23	2.01	0.05	0.30	0.44	7.70	1.32	0.05	0.00	0.29
UE-R1-LS	7.36	1.97	0.05	0.62	0.57	7.34	2.04	0.05	0.48	0.51	7.26	1.99	0.05	0.30	0.43	7.69	1.31	0.05	0.00	0.28
UE-R2-D	8.38	1.59	0.04	0.48	0.41	8.30	1.60	0.04	0.44	0.40	8.05	1.60	0.04	0.30	0.35	7.65	1.09	0.05	0.00	0.22
UE-R2-S	7.33	1.99	0.06	0.66	0.61	7.30	2.06	0.05	0.49	0.53	7.23	2.00	0.05	0.30	0.44	7.69	1.31	0.05	0.00	0.28
UE-R2-LS	7.40	1.95	0.05	0.61	0.54	7.39	2.01	0.05	0.48	0.49	7.29	1.97	0.05	0.30	0.41	7.66	1.30	0.05	0.00	0.27
UE-R3-D	8.37	1.59	0.04	0.48	0.42	8.29	1.60	0.04	0.44	0.41	8.04	1.60	0.04	0.30	0.36	7.64	1.09	0.05	0.00	0.23
UE-R3-S	7.32	2.00	0.06	0.66	0.64	7.29	2.07	0.05	0.49	0.55	7.23	2.01	0.05	0.30	0.46	7.69	1.32	0.05	0.00	0.29
UE-R3-LS	7.37	1.96	0.05	0.62	0.57	7.36	2.03	0.05	0.48	0.51	7.27	1.98	0.05	0.30	0.43	7.67	1.31	0.05	0.00	0.28
CFSS 12F																				
UE-R0-D	7.80	1.83	0.07	0.75	0.49	7.72	1.87	0.07	0.50	0.41	7.51	1.85	0.07	0.30	0.34	7.66	1.26	0.07	0.00	0.19
UE-R0-S	7.58	1.97	0.07	0.81	0.56	7.50	1.97	0.07	0.50	0.45	7.36	1.96	0.07	0.30	0.36	7.64	1.32	0.06	0.00	0.21
UE-R0-LS	7.61	1.96	0.07	0.79	0.53	7.53	1.97	0.07	0.50	0.43	7.38	1.95	0.07	0.30	0.35	7.63	1.31	0.06	0.00	0.21
UE-R1-D	7.81	1.83	0.07	0.75	0.48	7.73	1.88	0.07	0.50	0.40	7.51	1.86	0.07	0.30	0.33	7.66	1.27	0.07	0.00	0.19
UE-R1-S	7.59	1.98	0.07	0.80	0.54	7.51	1.98	0.07	0.50	0.43	7.36	1.96	0.07	0.30	0.35	7.64	1.32	0.06	0.00	0.21
UE-R1-LS	7.62	1.97	0.07	0.79	0.52	7.54	1.98	0.07	0.50	0.42	7.38	1.96	0.07	0.30	0.34	7.64	1.32	0.06	0.00	0.20
UE-R2-D	7.80	1.84	0.07	0.76	0.48	7.72	1.87	0.07	0.50	0.40	7.51	1.85	0.07	0.30	0.32	7.67	1.26	0.07	0.00	0.19
UE-R2-S	7.58	1.97	0.07	0.81	0.54	7.50	1.97	0.07	0.50	0.43	7.36	1.95	0.07	0.30	0.34	7.64	1.32	0.06	0.00	0.20
UE-R2-LS	7.62	1.96	0.07	0.79	0.50	7.54	1.96	0.07	0.50	0.41	7.39	1.94	0.07	0.30	0.33	7.64	1.31	0.06	0.00	0.20
UE-R3-D	7.80	1.83	0.07	0.75	0.49	7.72	1.87	0.07	0.50	0.42	7.51	1.85	0.07	0.30	0.34	7.66	1.26	0.07	0.00	0.19
UE-R3-S	7.58	1.97	0.07	0.81	0.56	7.51	1.97	0.07	0.50	0.45	7.36	1.96	0.07	0.30	0.36	7.64	1.32	0.06	0.00	0.21
UE-R3-LS	7.62	1.96	0.07	0.79	0.53	7.54	1.97	0.07	0.50	0.43	7.38	1.95	0.07	0.30	0.35	7.64	1.31	0.06	0.00	0.21
Factor Lasso 1F	7.51	2.11	0.06	0.68	0.68	7.33	2.12	0.05	0.50	0.63	7.17	2.06	0.05	0.30	0.56	7.50	1.43	0.05	0.00	0.39
Factor Lasso 3F	7.41	2.10	0.05	0.66	0.69	7.25	2.12	0.05	0.50	0.62	7.08	2.08	0.05	0.30	0.53	7.47	1.44	0.05	0.00	0.35
Factor Lasso 12F	7.61	2.11	0.08	0.84	0.73	7.40	2.05	0.07	0.50	0.57	7.11	2.06	0.07	0.30	0.46	7.41	1.36	0.06	0.00	0.29
Factor HEAVY 1F	8.60	1.47	0.04	0.49	0.26	8.49	1.48	0.04	0.47	0.27	8.00	1.54	0.04	0.30	0.26	7.55	1.25	0.04	0.00	0.16
Factor HEAVY 3F	8.20	1.51	0.04	0.44	0.27	8.11	1.52	0.04	0.42	0.27	7.81	1.55	0.04	0.30	0.25	7.57	1.25	0.04	0.00	0.16
Factor HEAVY 12F	7.72	1.85	0.06	0.64	0.74	7.64	1.88	0.06	0.49	0.67	7.47	1.87	0.06	0.30	0.55	7.55	1.35	0.06	0.00	0.32
sRe-cDECO	7.10	2.48	0.11	1.00	0.69	7.04	2.28	0.09	0.50	0.46	7.11	2.08	0.08	0.30	0.35	7.59	1.42	0.07	0.00	0.19
EWMA	7.21	2.09	0.09	0.93	0.14	7.14	2.02	0.08	0.50	0.10	7.12	1.97	0.07	0.30	0.07	7.52	1.39	0.07	0.00	0.03
RW	7.43	1.65	0.09	0.93	2.67	7.37	1.73	0.07	0.50	2.04	7.39	1.73	0.07	0.30	1.67	7.77	1.34	0.07	0.00	0.97

Note: The table reports GMVP summary statistics for the four portfolios under consideration based on 150 five-step-ahead conditional covariance matrix predictions; σ_p stands for the annualized standard deviation of the portfolio returns. The average portfolio concentration CO, short position SP and turnover TO are given by Eqs. (3.29) to (3.31). Grey-shaded cells indicate that the respective model belongs to the 90% model confidence set.

Table 3.5.: Daily MVP with momentum signal forecasting results.

	Unrestricted MVP					Restricted MVP: 150/50					Restricted MVP: 130/30					Restricted MVP: Long-only				
Model	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO
CFSS 1F																				
UE-R0-D	8.34	2.92	0.05	0.67	0.36	8.14	2.80	0.05	0.49	0.34	8.10	2.64	0.06	0.30	0.31	12.48	1.53	0.07	0.00	0.19
UE-R0-S	7.29	3.43	0.07	0.84	0.66	7.53	3.28	0.07	0.50	0.51	7.99	3.05	0.07	0.30	0.42	12.67	1.48	0.08	0.00	0.21
UE-R0-LS	7.35	3.39	0.07	0.79	0.59	7.55	3.26	0.06	0.50	0.48	7.97	3.04	0.06	0.30	0.40	12.63	1.49	0.08	0.00	0.20
UE-R1-D	8.34	2.93	0.05	0.66	0.35	8.14	2.81	0.05	0.49	0.33	8.09	2.65	0.06	0.30	0.30	12.48	1.52	0.07	0.00	0.19
UE-R1-S	7.29	3.43	0.07	0.84	0.63	7.53	3.29	0.07	0.50	0.49	7.98	3.06	0.07	0.30	0.40	12.67	1.48	0.08	0.00	0.20
UE-R1-LS	7.33	3.40	0.07	0.80	0.58	7.54	3.27	0.06	0.50	0.47	7.97	3.06	0.06	0.30	0.38	12.64	1.49	0.08	0.00	0.20
UE-R2-D	8.34	2.92	0.05	0.67	0.35	8.14	2.80	0.05	0.49	0.33	8.10	2.64	0.06	0.30	0.30	12.48	1.52	0.07	0.00	0.19
UE-R2-S	7.28	3.43	0.07	0.84	0.62	7.53	3.28	0.07	0.50	0.48	7.98	3.06	0.07	0.30	0.40	12.67	1.48	0.08	0.00	0.20
UE-R2-LS	7.35	3.38	0.07	0.78	0.54	7.55	3.26	0.06	0.50	0.45	7.95	3.05	0.06	0.30	0.37	12.62	1.49	0.08	0.00	0.20
UE-R3-D	8.34	2.92	0.05	0.67	0.36	8.14	2.80	0.05	0.49	0.34	8.10	2.64	0.06	0.30	0.31	12.48	1.53	0.07	0.00	0.19
UE-R3-S	7.29	3.43	0.07	0.84	0.66	7.54	3.28	0.07	0.50	0.52	7.99	3.05	0.07	0.30	0.42	12.67	1.48	0.08	0.00	0.21
UE-R3-LS	7.35	3.39	0.07	0.79	0.59	7.55	3.26	0.06	0.50	0.48	7.97	3.04	0.06	0.30	0.40	12.63	1.49	0.08	0.00	0.20
CFSS 3F																				
UE-R0-D	8.01	3.12	0.05	0.67	0.38	7.94	2.95	0.05	0.49	0.34	8.04	2.70	0.06	0.30	0.29	12.51	1.55	0.08	0.00	0.18
UE-R0-S	7.30	3.39	0.07	0.85	0.62	7.53	3.24	0.07	0.50	0.48	7.96	3.02	0.07	0.30	0.38	12.65	1.52	0.08	0.00	0.20
UE-R0-LS	7.34	3.36	0.07	0.80	0.56	7.53	3.24	0.06	0.50	0.45	7.94	3.01	0.06	0.30	0.37	12.62	1.53	0.08	0.00	0.19
UE-R1-D	8.01	3.12	0.05	0.66	0.36	7.94	2.95	0.05	0.49	0.32	8.04	2.70	0.06	0.30	0.28	12.52	1.54	0.08	0.00	0.18
UE-R1-S	7.29	3.39	0.07	0.84	0.59	7.52	3.25	0.07	0.50	0.46	7.95	3.03	0.07	0.30	0.37	12.65	1.51	0.08	0.00	0.19
UE-R1-LS	7.32	3.37	0.07	0.81	0.55	7.53	3.24	0.06	0.50	0.44	7.94	3.02	0.06	0.30	0.35	12.63	1.52	0.08	0.00	0.19
UE-R2-D	8.01	3.11	0.05	0.67	0.36	7.93	2.94	0.05	0.49	0.32	8.04	2.71	0.06	0.30	0.28	12.50	1.54	0.08	0.00	0.18
UE-R2-S	7.29	3.40	0.07	0.85	0.59	7.52	3.25	0.07	0.50	0.45	7.95	3.03	0.07	0.30	0.36	12.65	1.51	0.08	0.00	0.19
UE-R2-LS	7.34	3.35	0.07	0.79	0.52	7.54	3.24	0.06	0.50	0.42	7.93	3.02	0.06	0.30	0.34	12.61	1.53	0.08	0.00	0.19
UE-R3-D	8.01	3.12	0.05	0.67	0.38	7.94	2.95	0.05	0.49	0.34	8.04	2.70	0.06	0.30	0.29	12.51	1.55	0.08	0.00	0.18
UE-R3-S	7.30	3.39	0.07	0.85	0.62	7.53	3.25	0.07	0.50	0.48	7.96	3.02	0.07	0.30	0.39	12.65	1.52	0.08	0.00	0.20
UE-R3-LS	7.34	3.36	0.07	0.80	0.56	7.53	3.24	0.06	0.50	0.45	7.94	3.01	0.06	0.30	0.37	12.62	1.53	0.08	0.00	0.19
CFSS 12F																				
UE-R0-D	7.48	3.41	0.09	0.92	0.46	7.58	3.21	0.08	0.50	0.35	7.87	2.86	0.08	0.30	0.28	12.48	1.54	0.08	0.00	0.17
UE-R0-S	7.33	3.39	0.09	0.98	0.59	7.48	3.21	0.08	0.50	0.41	7.83	2.93	0.08	0.30	0.33	12.51	1.53	0.08	0.00	0.18
UE-R0-LS	7.34	3.40	0.09	0.95	0.55	7.49	3.23	0.08	0.50	0.40	7.82	2.92	0.08	0.30	0.32	12.50	1.53	0.08	0.00	0.18
UE-R1-D	7.49	3.42	0.09	0.92	0.44	7.58	3.22	0.08	0.50	0.33	7.87	2.86	0.08	0.30	0.27	12.49	1.54	0.08	0.00	0.17
UE-R1-S	7.33	3.39	0.09	0.98	0.56	7.48	3.22	0.08	0.50	0.40	7.83	2.93	0.08	0.30	0.32	12.51	1.53	0.08	0.00	0.17
UE-R1-LS	7.34	3.39	0.09	0.96	0.53	7.49	3.23	0.08	0.50	0.38	7.82	2.93	0.08	0.30	0.31	12.50	1.53	0.08	0.00	0.17
UE-R2-D	7.48	3.41	0.09	0.92	0.44	7.57	3.21	0.08	0.50	0.33	7.86	2.86	0.08	0.30	0.27	12.48	1.54	0.08	0.00	0.17
UE-R2-S	7.32	3.39	0.09	0.98	0.55	7.47	3.22	0.08	0.50	0.39	7.82	2.93	0.08	0.30	0.31	12.51	1.53	0.08	0.00	0.17
UE-R2-LS	7.34	3.40	0.09	0.95	0.51	7.48	3.23	0.08	0.50	0.37	7.82	2.92	0.08	0.30	0.30	12.50	1.53	0.08	0.00	0.17
UE-R3-D	7.48	3.41	0.09	0.92	0.46	7.58	3.21	0.08	0.50	0.35	7.87	2.86	0.08	0.30	0.28	12.48	1.55	0.08	0.00	0.17
UE-R3-S	7.33	3.39	0.09	0.98	0.59	7.48	3.21	0.08	0.50	0.41	7.83	2.93	0.08	0.30	0.33	12.51	1.53	0.08	0.00	0.18
UE-R3-LS	7.34	3.40	0.09	0.95	0.55	7.49	3.23	0.08	0.50	0.40	7.82	2.92	0.08	0.30	0.32	12.50	1.53	0.08	0.00	0.18
Factor Lasso 1F	7.73	3.38	0.07	0.82	0.74	7.77	3.27	0.06	0.50	0.62	7.97	3.01	0.06	0.30	0.53	12.57	1.48	0.08	0.00	0.25
Factor Lasso 3F	7.58	3.38	0.07	0.81	0.73	7.68	3.27	0.06	0.50	0.60	7.95	2.98	0.06	0.30	0.51	12.62	1.49	0.08	0.00	0.24
Factor Lasso 12F	7.77	3.21	0.09	0.97	0.79	7.74	3.08	0.08	0.50	0.57	8.00	2.79	0.08	0.30	0.46	12.64	1.49	0.08	0.00	0.21
Factor HEAVY 1F	8.48	2.81	0.05	0.68	0.35	8.22	2.73	0.05	0.49	0.33	8.10	2.58	0.06	0.30	0.30	12.43	1.49	0.07	0.00	0.19
Factor HEAVY 3F	8.13	2.96	0.05	0.66	0.36	8.02	2.86	0.05	0.49	0.33	8.09	2.62	0.06	0.30	0.29	12.48	1.47	0.07	0.00	0.18
Factor HEAVY 12F	7.56	3.32	0.08	0.89	0.52	7.59	3.16	0.07	0.50	0.40	7.82	2.90	0.07	0.30	0.32	12.42	1.54	0.08	0.00	0.18
sRe-cDECO	7.86	3.02	0.12	1.16	0.51	8.00	2.83	0.09	0.50	0.33	8.29	2.72	0.08	0.30	0.27	12.78	1.47	0.08	0.00	0.14
EWMA	7.12	3.37	0.10	1.10	0.29	7.43	3.04	0.08	0.50	0.20	7.83	2.88	0.08	0.30	0.17	12.41	1.50	0.08	0.00	0.14
RW	7.85	3.43	0.09	0.87	2.73	7.94	3.37	0.08	0.50	2.27	8.34	3.08	0.08	0.30	1.89	12.90	1.54	0.08	0.00	0.77

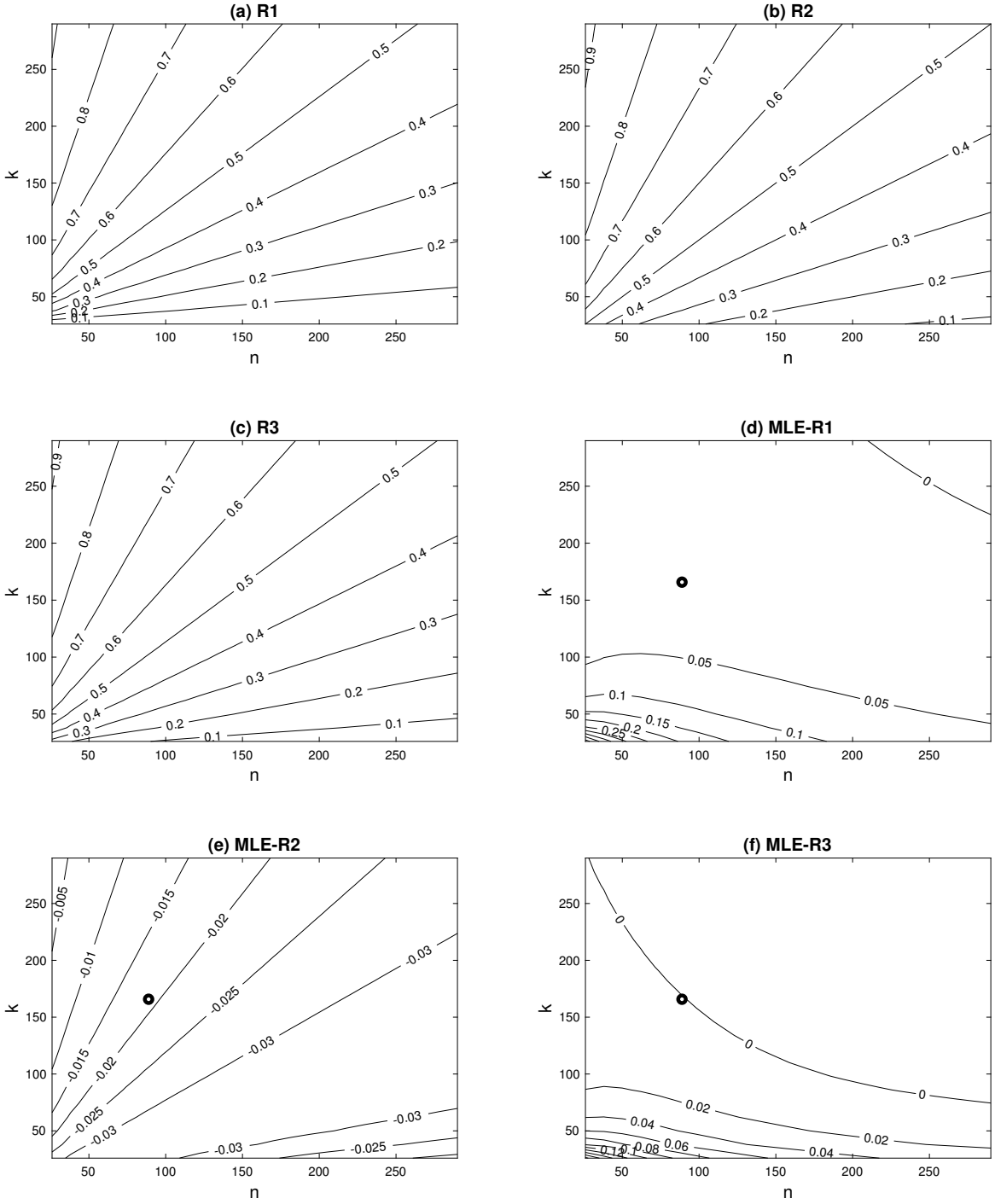
Note: The table reports MVP with momentum signal summary statistics for the four portfolios under consideration based on 754 one-step-ahead conditional covariance matrix predictions; σ_p stands for the annualized standard deviation of the portfolio returns. The average portfolio concentration CO, short position SP and turnover TO are given by Eqs. (3.29) to (3.31). Grey-shaded cells indicate that the respective model belongs to the 90% model confidence set.

Table 3.6.: Weekly MVP with momentum signal forecasting results.

	Unrestricted MVP					Restricted MVP: 150/50					Restricted MVP: 130/30					Restricted MVP: Long-only				
Model	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO	σ_p	SR	CO	SP	TO
CFSS 1F																				
UE-R0-D	9.45	2.31	0.05	0.67	0.52	8.98	2.31	0.05	0.49	0.50	8.67	2.27	0.06	0.30	0.47	12.03	1.44	0.07	0.00	0.34
UE-R0-S	7.91	2.95	0.07	0.83	0.78	8.03	2.85	0.06	0.50	0.63	8.10	2.76	0.06	0.30	0.53	12.06	1.40	0.08	0.00	0.35
UE-R0-LS	8.00	2.89	0.06	0.78	0.71	8.09	2.80	0.06	0.50	0.60	8.14	2.72	0.06	0.30	0.51	12.05	1.41	0.08	0.00	0.35
UE-R1-D	9.44	2.31	0.05	0.66	0.51	8.97	2.32	0.05	0.49	0.49	8.65	2.28	0.06	0.30	0.46	12.03	1.44	0.07	0.00	0.34
UE-R1-S	7.92	2.95	0.07	0.82	0.76	8.03	2.85	0.06	0.50	0.62	8.10	2.76	0.06	0.30	0.52	12.08	1.40	0.08	0.00	0.35
UE-R1-LS	7.98	2.91	0.06	0.79	0.70	8.08	2.82	0.06	0.50	0.59	8.14	2.73	0.06	0.30	0.51	12.07	1.41	0.08	0.00	0.34
UE-R2-D	9.46	2.29	0.05	0.67	0.51	8.99	2.30	0.05	0.49	0.50	8.68	2.26	0.06	0.30	0.46	12.03	1.44	0.07	0.00	0.34
UE-R2-S	7.93	2.94	0.07	0.83	0.76	8.04	2.84	0.06	0.50	0.61	8.12	2.75	0.06	0.30	0.52	12.05	1.40	0.08	0.00	0.35
UE-R2-LS	8.05	2.86	0.06	0.77	0.67	8.13	2.78	0.06	0.50	0.57	8.17	2.70	0.06	0.30	0.50	12.04	1.41	0.08	0.00	0.34
UE-R3-D	9.45	2.31	0.05	0.67	0.53	8.98	2.31	0.05	0.49	0.51	8.66	2.27	0.06	0.30	0.47	12.03	1.44	0.07	0.00	0.34
UE-R3-S	7.92	2.95	0.07	0.83	0.78	8.03	2.85	0.06	0.50	0.63	8.10	2.76	0.06	0.30	0.53	12.06	1.40	0.08	0.00	0.35
UE-R3-LS	8.01	2.89	0.06	0.78	0.71	8.10	2.80	0.06	0.50	0.60	8.15	2.72	0.06	0.30	0.51	12.05	1.41	0.08	0.00	0.35
CFSS 3F																				
UE-R0-D	9.15	2.39	0.05	0.66	0.54	8.80	2.38	0.05	0.49	0.49	8.64	2.29	0.06	0.30	0.44	12.05	1.44	0.07	0.00	0.33
UE-R0-S	7.94	2.90	0.07	0.83	0.73	8.06	2.81	0.06	0.50	0.59	8.14	2.70	0.06	0.30	0.49	12.04	1.41	0.08	0.00	0.34
UE-R0-LS	8.01	2.85	0.06	0.79	0.67	8.10	2.78	0.06	0.50	0.56	8.18	2.67	0.06	0.30	0.47	12.03	1.42	0.08	0.00	0.34
UE-R1-D	9.14	2.39	0.05	0.66	0.52	8.79	2.38	0.05	0.49	0.48	8.62	2.29	0.06	0.30	0.43	12.06	1.45	0.07	0.00	0.33
UE-R1-S	7.94	2.90	0.07	0.82	0.71	8.05	2.81	0.06	0.50	0.57	8.14	2.70	0.06	0.30	0.48	12.06	1.41	0.08	0.00	0.34
UE-R1-LS	7.99	2.87	0.07	0.79	0.67	8.09	2.79	0.06	0.50	0.55	8.18	2.68	0.06	0.30	0.46	12.05	1.42	0.08	0.00	0.34
UE-R2-D	9.16	2.38	0.05	0.66	0.53	8.81	2.37	0.05	0.49	0.49	8.66	2.28	0.06	0.30	0.43	12.05	1.44	0.08	0.00	0.33
UE-R2-S	7.95	2.90	0.07	0.83	0.71	8.07	2.80	0.06	0.50	0.57	8.15	2.69	0.06	0.30	0.48	12.04	1.41	0.08	0.00	0.34
UE-R2-LS	8.04	2.83	0.06	0.78	0.64	8.13	2.77	0.06	0.50	0.54	8.21	2.65	0.06	0.30	0.46	12.02	1.42	0.08	0.00	0.33
UE-R3-D	9.15	2.39	0.05	0.66	0.54	8.80	2.38	0.05	0.49	0.50	8.63	2.29	0.06	0.30	0.44	12.05	1.45	0.07	0.00	0.33
UE-R3-S	7.94	2.90	0.07	0.83	0.74	8.06	2.81	0.06	0.50	0.59	8.14	2.70	0.06	0.30	0.49	12.05	1.41	0.08	0.00	0.34
UE-R3-LS	8.01	2.85	0.06	0.78	0.67	8.11	2.78	0.06	0.50	0.56	8.19	2.67	0.06	0.30	0.47	12.03	1.42	0.08	0.00	0.34
CFSS 12F																				
UE-R0-D	8.43	2.70	0.08	0.91	0.61	8.47	2.69	0.08	0.50	0.49	8.58	2.53	0.08	0.30	0.42	12.03	1.40	0.08	0.00	0.32
UE-R0-S	8.14	2.89	0.09	0.96	0.68	8.29	2.79	0.08	0.50	0.52	8.36	2.68	0.08	0.30	0.43	12.01	1.40	0.08	0.00	0.32
UE-R0-LS	8.18	2.87	0.08	0.94	0.65	8.31	2.78	0.07	0.50	0.50	8.39	2.65	0.08	0.30	0.43	12.01	1.40	0.08	0.00	0.32
UE-R1-D	8.42	2.70	0.08	0.90	0.60	8.47	2.69	0.08	0.50	0.48	8.57	2.54	0.08	0.30	0.41	12.03	1.41	0.08	0.00	0.31
UE-R1-S	8.15	2.89	0.09	0.96	0.66	8.30	2.79	0.07	0.50	0.51	8.36	2.68	0.08	0.30	0.42	12.01	1.40	0.08	0.00	0.32
UE-R1-LS	8.19	2.88	0.08	0.94	0.64	8.31	2.79	0.07	0.50	0.50	8.39	2.66	0.08	0.30	0.42	12.01	1.40	0.08	0.00	0.31
UE-R2-D	8.42	2.70	0.09	0.92	0.60	8.47	2.69	0.08	0.50	0.48	8.59	2.53	0.08	0.30	0.41	12.03	1.40	0.08	0.00	0.31
UE-R2-S	8.14	2.89	0.09	0.96	0.66	8.29	2.79	0.08	0.50	0.50	8.36	2.67	0.08	0.30	0.42	12.01	1.40	0.08	0.00	0.31
UE-R2-LS	8.18	2.86	0.08	0.94	0.62	8.31	2.78	0.07	0.50	0.49	8.40	2.64	0.08	0.30	0.41	12.01	1.40	0.08	0.00	0.31
UE-R3-D	8.42	2.70	0.08	0.91	0.61	8.47	2.69	0.08	0.50	0.49	8.58	2.54	0.08	0.30	0.42	12.03	1.40	0.08	0.00	0.32
UE-R3-S	8.15	2.89	0.09	0.96	0.68	8.29	2.79	0.08	0.50	0.52	8.36	2.68	0.08	0.30	0.43	12.01	1.40	0.08	0.00	0.32
UE-R3-LS	8.18	2.87	0.08	0.94	0.65	8.31	2.78	0.07	0.50	0.51	8.39	2.65	0.08	0.30	0.43	12.01	1.40	0.08	0.00	0.32
Factor Lasso 1F	8.09	2.96	0.07	0.82	0.74	8.03	2.88	0.06	0.50	0.64	8.03	2.77	0.06	0.30	0.55	12.08	1.35	0.08	0.00	0.35
Factor Lasso 3F	8.01	2.95	0.07	0.81	0.75	7.97	2.90	0.06	0.50	0.63	8.01	2.79	0.06	0.30	0.53	12.06	1.39	0.08	0.00	0.34
Factor Lasso 12F	8.16	3.00	0.09	0.97	0.80	8.03	2.93	0.08	0.50	0.59	8.06	2.71	0.08	0.30	0.49	11.97	1.41	0.08	0.00	0.31
Factor HEAVY 1F	9.02	2.28	0.05	0.66	0.38	8.81	2.22	0.05	0.50	0.38	8.62	2.01	0.05	0.30	0.36	12.34	1.32	0.07	0.00	0.29
Factor HEAVY 3F	8.78	2.25	0.05	0.62	0.37	8.65	2.20	0.05	0.49	0.37	8.61	1.99	0.05	0.30	0.35	12.31	1.28	0.07	0.00	0.28
Factor HEAVY 12F	8.17	2.66	0.07	0.80	0.78	8.26	2.57	0.07	0.50	0.65	8.43	2.31	0.07	0.30	0.54	12.02	1.43	0.08	0.00	0.33
sRe-cDECO	7.66	3.14	0.12	1.15	0.79	7.84	2.87	0.09	0.50	0.52	8.20	2.58	0.08	0.30	0.43	12.14	1.37	0.08	0.00	0.29
EWMA	7.65	3.09	0.10	1.09	0.32	7.90	2.89	0.08	0.50	0.28	8.13	2.68	0.08	0.30	0.26	12.09	1.43	0.08	0.00	0.28
RW	7.98	2.80	0.11	1.11	2.83	8.15	2.78	0.08	0.50	1.93	8.13	2.75	0.08	0.30	1.51	12.10	1.46	0.08	0.00	0.59

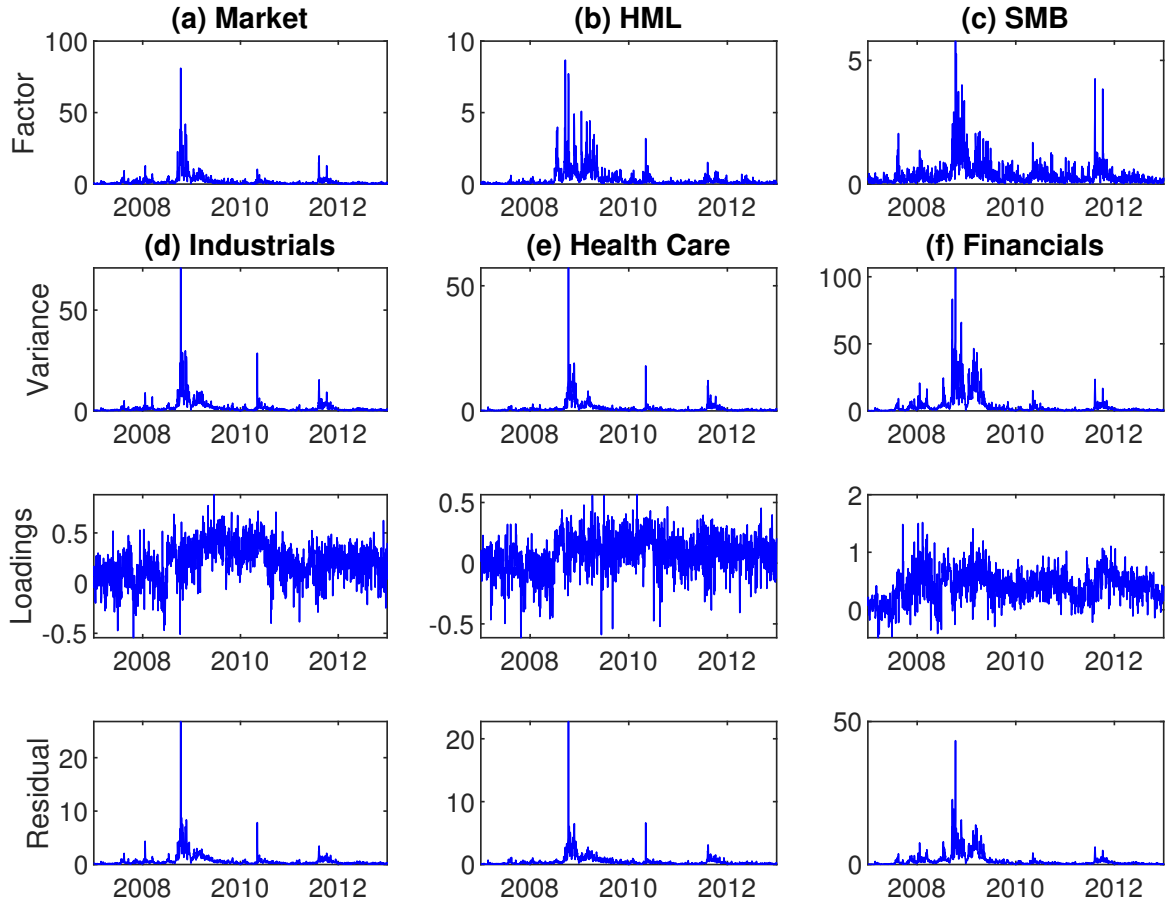
Note: The table reports MVP with momentum signal summary statistics for the four portfolios under consideration based on 150 five-step-ahead conditional covariance matrix predictions; σ_p stands for the annualized standard deviation of the portfolio returns. The average portfolio concentration CO, short position SP and turnover TO are given by Eqs. (3.29) to (3.31). Grey-shaded cells indicate that the respective model belongs to the 90% model confidence set.

Figure 3.1.: Smoothing parameter restrictions.



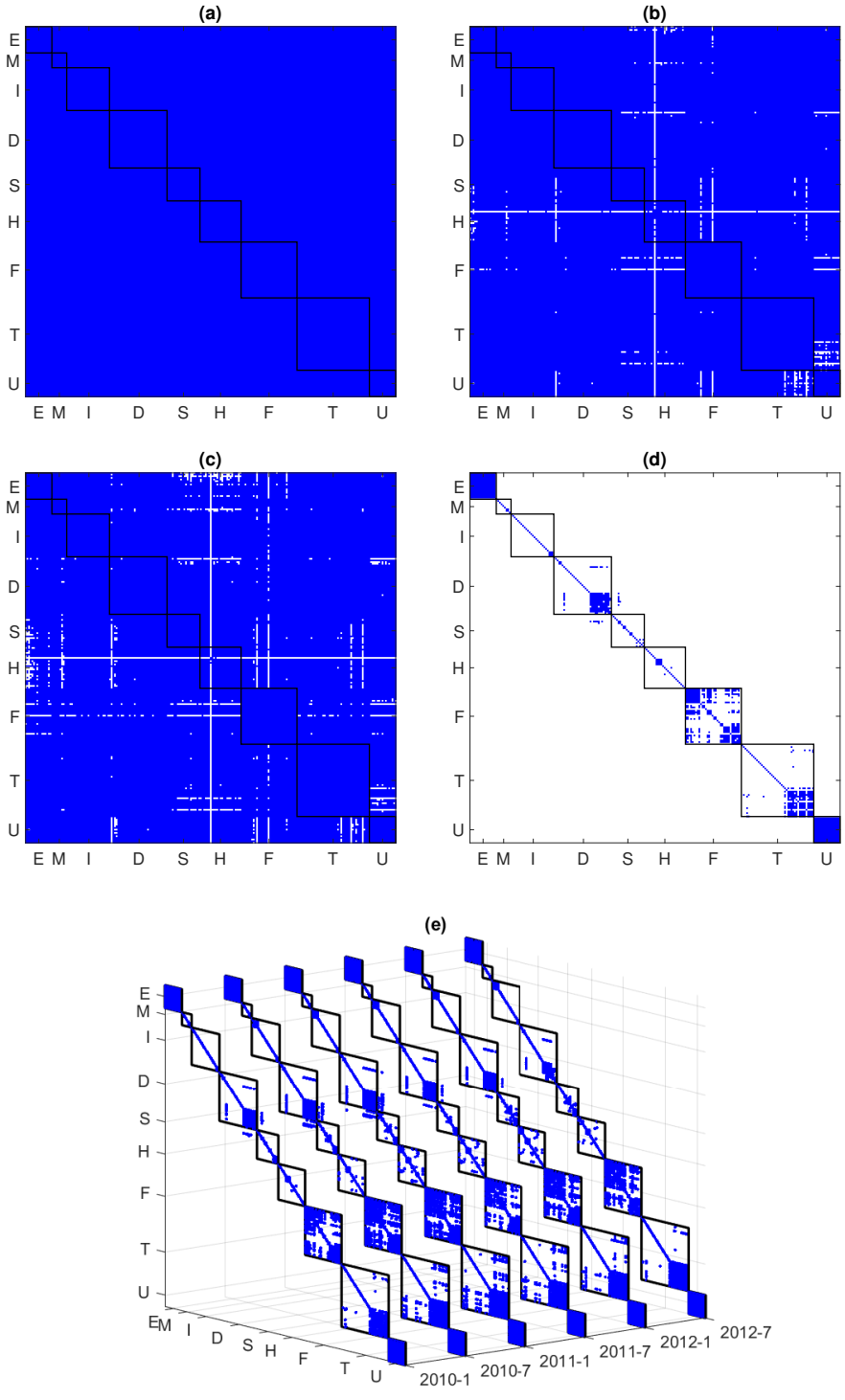
Panels (a) – (c): Smoothing parameter of the UE model under restrictions (R1), (R2) and (R3). Panels (d) – (f): Differences of the ML estimate for λ (when fixing n and k) and the values obtained under the respective restrictions. The black circles 'o' mark the respective unrestricted ML estimates of the d.o.f. parameters.

Figure 3.2.: Time-series plots of selected realized factor and aggregated (residual) asset variances.



Panels (a) – (c): Time series plots of the realized Fama-French factor variances; Columns (d) – (f): Time series plots of average realized asset variances (top), factor loadings on the market (middle) and residual variances (bottom) for the Industrials, Health Care and Financials sector.

Figure 3.3.: Sparsity pattern of residual correlations.



The figure displays the significant entries of the residual covariance matrices, relative to (a) zero, (b) one (Market), (c) three (Market + FF), and (d) 12 (Market + FF + ETFs) observed factors. Panel (e) displays the the residual sparsity for 12 factors at different points in time. The sector labels are listed in Table 3.1.

Chapter 4.

Modeling Realized Covariance Measures with Heterogeneous Liquidity: A Generalized Matrix-Variate Wishart State-Space Model

4.1. Introduction

The modeling and forecasting of covariance matrices of asset returns is an important field in financial econometrics with potential applications in portfolio allocation, hedging, and risk management. Recent approaches increasingly focus on the direct modeling of time-series of daily realized covariance matrices. These measures are computed from intraday asset return information and represent non-parametric consistent ex-post estimates of the daily integrated covariance matrices of asset returns as obtained from a continuous-time diffusion for the underlying intraday price process (see, e.g., Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2004; Park and Linton, 2012; Lunde et al., 2016). It is now widely documented that models for realized covariance matrices provide more precise forecasts than multivariate GARCH (MGARCH) and multivariate stochastic volatility (MSV) models, which only exploit daily asset return information.

Pioneering contributions on the modeling and forecasting of realized covariance matrices are provided by Gourieroux et al. (2009), Chiriac and Voev (2011), Bauer and Vorkink (2011), Noureldin et al. (2012), Golosnoy et al. (2012) and Jin and Maheu

This chapter is based on:
Modeling Realized Covariance Measures with Heterogeneous Liquidity: A Generalized Matrix-Variate Wishart State-Space Model; Authors: Bastian Gribisch and Jan P. Hartkopf.
Currently returned for revision and resubmission to The Journal of Econometrics for possible publication.

(2013). More recent approaches and model extensions are, e.g., found in Opschoor et al. (2017), Opschoor and Lucas (2019), Vassallo et al. (2019) and Sheppard and Xu (2019). The vast majority of models essentially treat the realized covariance matrix as the true integrated covariance matrix and apply observation-driven reduced form autoregressive models in order to forecast the realized measures. This approach is convenient from the perspective of parameter estimation and forecasting since the likelihood of observation-driven models is directly available, but comes with the disadvantage of neglected measurement errors in the lagged realized covariances driving the covariance dynamics. The effect of neglected measurement errors is, e.g., discussed in Bollerslev et al. (2016), Bollerslev et al. (2018) and Bekierman and Manner (2018), who suggest model extensions in order to mitigate the resulting bias in the persistence estimates.

An alternative to the observation-driven approach is to model the realized covariance matrix as a noisy measurement of the time-varying latent integrated covariance matrix. The resulting parameter-driven state-space framework directly accounts for measurement errors via the measurement density specification. The state-space approach has, e.g., been advocated by Asai et al. (2012a,b), who find considerable bias reductions in estimated integrated volatilities if measurement errors in realized volatilities are accounted for. Further applications are found in Barndorff-Nielsen and Shephard (2002), Takahashi et al. (2009) and Bekierman and Gribisch (2016).

Parameter-driven models come with the additional advantage of increased flexibility in modeling the (co)variance dynamics via the additional idiosyncratic innovation process driving the latent states. This helps to generate overdispersion, heavier tails and further features, which can put such models at an advantage over observation-driven models. The state-space approach also appears as a rather natural modeling framework for realized covariance measures since the latter represent consistent but noisy estimates of the latent integrated covariance state of an underlying continuous-time SV diffusion.

Contributions on parameter-driven state-space models for matrix-valued realized

(co)variance measures are sparse, which is explained by the relative complexity of model estimation and forecasting in high-dimensional settings combined with the nonlinear restriction of positive-definiteness (p.d.) for the covariance measure. The associated likelihood obtains as a high-dimensional integral, typically without closed form expression. Estimation then has to be based on simulation based techniques like Bayesian Monte Carlo Markov Chain (MCMC) or Simulated Maximum Likelihood (SML). Gribisch (2018) employs a linear Gaussian latent dynamic factor model for the distinct elements of the matrix-logarithm of a realized covariance matrix and Gribisch et al. (2020, see also Chapter 2 herein) propose a factor state-space model based on observed risk factors and a Wishart measurement density. Both approaches rely on Bayesian MCMC techniques for model estimation and model selection. Windle and Carvalho (2014) propose a simple yet flexible state-space framework via combining a Wishart measurement density with a sparsely parameterized generalized matrix-variate Beta type-I transition for the latent precision matrix, resulting in a closed form expression for the likelihood function.

The Wishart has become a popular approach for modeling realized covariance matrices (see, e.g., Golosnoy et al., 2012; Noureldin et al., 2012; Asai and McAleer, 2015; Gorgi et al., 2019). However, we find that the Wishart is rather ill-suited in the empirically realistic setting of heterogeneous liquidity across assets: As the Wishart's scale matrix is tied to the conditional mean of the covariance matrix, the measurement error's covariance structure is – ceteris paribus – entirely determined by a single scalar-valued d.o.f. parameter. Using a simple stylized iid model for $k \geq 1$ intraday asset returns, this d.o.f. parameter corresponds to the number of intraday return observations common to all assets. Since the (co)variances of Wishart random matrix elements are inverse proportional to the d.o.f. this induces an attenuation of the measurement error variance for less liquid assets, resulting in underestimation of the state persistence and overestimation of the state variation. This may induce severe effects on the forecasting performance of Wishart-based state-space models.

Accounting for heterogeneous liquidity appears crucial in high-dimensional state-

space modeling of realized covariance matrices. In order to alleviate the attenuation of measurement errors we consider the Riesz distribution introduced by Hassairi and Lajmi (2001) as a flexible generalization of the Wishart. Although the Riesz has been analyzed in a couple of papers on multivariate distributions and graphical models (see, e.g., Andersson and Klein, 2010; Veleva, 2009) we have not found any applications to financial data. The Riesz nests the Wishart and generalizes the Wishart-implied data generating process (DGP) for intraday returns to heterogeneous liquidity across assets. As a result, the Riesz is parameterized by a scale matrix and one d.o.f. parameter per asset, implying a substantial increase in flexibility due to $k-1$ additional d.o.f. parameters for modeling the measurement error covariance matrix. We analyze the stochastic properties of the Riesz, derive the implied covariance structure for the measurement error and develop a Riesz State-Space (RSS) model, which can be easily analyzed by Bayesian MCMC procedures. Our empirical results indicate that the additional flexibility of the Riesz appears important for covariance forecasting in state-space frameworks.

A particular advantage of the RSS approach is that it allows for a numerically efficient, yet simple Bayesian analysis using fairly standard MCMC techniques: We exploit the Cholesky-structure of the Riesz and factorize the measurement density in order to obtain a conditionally independent sequence of state-space models for conditional regression coefficients and error variances of asset returns. By modeling the error variances and regression betas as independent autoregressive state processes we can employ basic MCMC sampling techniques, completely parallelized over the number of assets, which greatly simplifies applications to high-dimensional asset return vectors while preserving positive definiteness of the realized covariance measures without parametric restrictions. The model structure moreover allows for straightforward implementation of efficient shrinkage prior restrictions on the conditional state variances in order to alleviate the risk of overfitting in high-dimensional state-space settings.

The rest of the paper is organized as follows: Section 4.2 details the Wishart and

the Riesz distribution and their stochastic properties, and Section 4.3 introduces the RSS model. The proposed MCMC procedure for the Bayesian posterior analysis and model comparisons as well as the construction of forecasts are discussed in Section 4.4. Section 4.5 presents the empirical application to New York Stock Exchange (NYSE) data and Section 4.6 concludes. Additional material and proofs of the Propositions are given in Appendix C.

4.2. Wishart and Riesz Measurement Densities for Realized Covariance Matrices

Consider a $(k \times k)$ realized covariance matrix $C_t = (c_{ij,t})$ observed as a noisy but consistent measurement for the 'true' but latent time-varying integrated covariance matrix $\Sigma_t = (\sigma_{ij,t})$ for trading days $t = 1, \ldots, T$. Further denote the k-dimensional vector of observed asset log-returns for intraday trading period ℓ at trading day t by $r_{t\ell} = (r_{t\ell,1}, \ldots, r_{t\ell,k})'$, $\ell = 1, \ldots, n$.

The day t realized covariance matrix as defined by Barndorff-Nielsen and Shephard (2004) obtains as

$$C_t = \sum_{\ell=1}^{n} r_{t\ell} r_{t\ell}', \tag{4.1}$$

where n denotes the total number of intraday return vectors for trading day t. For simplicity and w.l.o.g. we assume n to be constant across trading days. Under rather general regularity conditions on the continuous log-price process, Barndorff-Nielsen and Shephard (2004) show that C_t is a consistent estimator of Σ_t as $n \to \infty$. The literature offers numerous refinements of the classical realized covariance estimator in Eq. (4.1), accounting for market microstructure noise, jumps and non-synchronicity of intraday asset prices (see e.g. Zhang et al., 2005, Christensen and Kinnebrock, 2010, and Lunde et al., 2016).

Taking the integrated covariance matrix Σ_t as a latent state variable observed

through the noisy realized covariance matrix C_t it is reasonable to model C_t by a state-space approach with a measurement density $f(C_t|\Sigma_t)$ relating the measurements C_t to the states Σ_t and a transition density $f(\Sigma_t|\Sigma_{1:t-1})$ designed to approximate the observed dynamics of C_t.[1]

4.2.1. The Wishart

The realized covariance measure in (4.1) allows for a direct derivation of the Wishart as a measurement density for realized covariance matrices. Let

$$r_{t\ell} = \frac{L_t}{\sqrt{n}} u_{t\ell}, \tag{4.2}$$

where $u_{t\ell} = (u_{t\ell,1}, \ldots, u_{t\ell,k})'$ denotes a k-dimensional vector of iid standard normal intraday return innovations, $L_t = T_t^{\Sigma}\big[D_t^{\Sigma}\big]^{\frac{1}{2}}$ is obtained from the Cholesky decomposition $\Sigma_t = T_t^{\Sigma} D_t^{\Sigma} T_t^{\Sigma'}$ and the scaling factor $1/\sqrt{n}$ ensures $\mathrm{Var}(r_t) = \mathrm{Var}\big(\sum_{\ell=1}^{n} r_{t\ell}\big) = \Sigma_t$. Summarizing the intraday return innovations of trading day t in the $(k \times n)$ innovation matrix $U_t = [u_{t1}, \ldots, u_{tn}]$ we rewrite Eq. (4.1) as

$$C_t = \frac{1}{n} L_t U_t U_t' L_t', \qquad \text{with} \quad U_t U_t' \mid n \sim \mathcal{W}_k(n, I_k), \tag{4.3}$$

where $\mathcal{W}_k(n, I_k)$ denotes the Wishart distribution with scalar-valued d.o.f. parameter n and identity scale matrix. From the scaling property of the Wishart we obtain[2]

$$C_t \mid \Sigma_t, n \sim \mathcal{W}_k\left(n, \frac{1}{n}\Sigma_t\right). \tag{4.4}$$

The Wishart pdf, its characteristic function and the associated first and second order moments are summarized in the following proposition.

Proposition 1. *Let C_t denote a positive-definite $(k \times k)$ realized covariance matrix with $C_t|\Sigma_t, n \sim \mathcal{W}_k\big(n, \frac{1}{n}\Sigma_t\big)$, where n denotes the associated scalar-valued d.o.f. pa-*

[1] The notation $A_{s:\tau}$ is used to denote the collection $\{A_s, \ldots, A_\tau\}$.

[2] See, e.g., Muirhead (2005) and Gupta and Nagar (2000), for details on the properties of the Wishart distribution.

rameter and Σ_t *the* $(k \times k)$ *p.d. scale matrix. For* $n > k-1$ *the Wishart pdf is given by*

$$f(C_t|\Sigma_t, n) = \frac{|C_t|^{(n-k-1)/2}|\Sigma_t/n|^{-n/2}}{2^{nk/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left((n-i+1)/2\right)} \exp\left\{-\frac{n}{2}\text{tr}\left(\Sigma_t^{-1}C_t\right)\right\}, \tag{4.5}$$

with characteristic function $\varphi_{\mathcal{W}}(\Theta) = |I_k - 2\iota\Theta\Sigma_t/n|^{-\frac{n}{2}}$*, where* Θ *is a symmetric k-dimensional matrix and* ι *denotes the imaginary number. We have* $E\left[C_t|\,\Sigma_t, n\right] = \Sigma_t$ *and*

$$\text{Cov}\left[\text{vec}(C_t)|\Sigma_t, n\right] = \frac{1}{n}\left(I_{k^2} + K_{kk}\right)\left(\Sigma_t \otimes \Sigma_t\right), \tag{4.6}$$

where $\text{vec}(\cdot)$ *denotes the operator that stacks all columns of a matrix into a vector and* K_{kk} *denotes the commutation matrix (see, e.g., Lütkepohl, 1996, p. 115). Proof: See Muirhead (2005).*

From Eq. (4.6) we obtain $\text{Cov}[c_{ij,t}, c_{lm,t}|\Sigma_t, n] = \frac{1}{n}(\sigma_{il,t} \cdot \sigma_{jm,t} + \sigma_{im,t} \cdot \sigma_{jl,t})$, which implies that the covariance structure of the Wishart measurement error for Σ_t is proportional to the sum of the crossproducts of the integrated (co)variances.

The Wishart distribution is frequently applied for the stochastic modeling of realized covariance matrices (see, e.g., Philipov and Glickman, 2006; Golosnoy et al., 2012; Noureldin et al., 2012; Jin and Maheu, 2013; Bauwens et al., 2014; Asai and McAleer, 2015; Sheppard and Xu, 2019; Gorgi et al., 2019; Gribisch et al., 2020, for recent contributions). However, the Wishart DGP for intraday asset returns appears empirically unrealistic and the Wishart can merely serve as an approximation to the true (but unknown) measurement density for C_t. Taking Eq. (4.2) as a parametric model for asset returns ignores intraday time-variation of the spot covariance, potential non-normality of the return innovation process and heterogeneous liquidity across assets. Moreover, in practical applications the basic realized covariance measure of Barndorff-Nielsen and Shephard (2004) is typically replaced by noise-robust alternatives as, e.g., the composite realized kernel estimator of Lunde et al. (2016), for which

small-sample distributional results are not available. The Wishart is therefore interpreted as a Quasi Likelihood (QL) and the integer d.o.f. n is generalized to a real number on $(k-1, \infty)$ (cf. Muirhead, 2005) and treated as an additional free model-parameter driving the covariance structure of the measurement error conditional on $\mathrm{E}\left[C_t \middle| \Sigma_t, n\right] = \Sigma_t$.

Taking a closer look at the Wishart DGP for intraday asset returns, the assumption of time-constant spot covariances appears hard to relax but may serve as a valid approximation for highly persistent covariance dynamics over the trading day. Moreover, assuming normality for intraday return innovations appears rather unproblematic in continuous volatility mixtures, where mixing over the Σ_t-dynamics induces leptokurtic daily return observations (cf. Gribisch et al., 2020). The assumption of homogeneous liquidity however is critical and – if violated – implies an attenuation of measurement errors for the less liquid assets in the data. The Wishart covariance in (4.6) reflects a homogeneous liquidity structure via the scalar d.o.f. parameter n, which induces an even covariance scaling across assets. Since the Wishart requires $n > k-1$, high-dimensional applications involve increasingly low measurement errors, which may not be justified for less liquid assets. The resulting attenuation of measurement errors for the less liquid assets induces overfitting and potentially poor out-of-sample performance since the Σ_t-dynamics are bound to mimic the noise.

4.2.2. The Riesz

In order to alleviate the Wishart-implied attenuation of measurement errors we now generalize the Wishart DGP to the empirically realistic setting of heterogeneous liquidity across assets. Let n_i denote the number of intraday return observations for the ith asset at $t = 1, \ldots, T$, with $n_1 \leq n_2 \leq \cdots \leq n_k$. In analogy to the Wishart DGP in (4.2) we write

$$r_{t\ell} = L_t Q \tilde{u}_{t\ell}, \qquad \text{with} \quad Q = \mathrm{diag}\left(1/\sqrt{n_1}, \ldots, 1/\sqrt{n_k}\right), \tag{4.7}$$

where $\tilde{u}_{t\ell} = (\tilde{u}_{t\ell,1}, \ldots, \tilde{u}_{t\ell,k})'$ denotes a k-dimensional vector of intraday return innovations. The heterogeneous liquidity pattern across assets may now be generated by a monotone missing data structure for the intraday innovation process (cf. Veleva, 2009): We denote the $(k \times n)$ intraday innovation matrix by $\tilde{U}_t$ with the ℓth column given by $\tilde{u}_{t\ell}$ and the ith row containing n_i iid standard normal random variables, $n_i > i-1$, followed by $n - n_i$ missing data points replaced by zeros:

$$\tilde{U}_t = \begin{bmatrix} \tilde{u}_{t1,1} & \ldots & \tilde{u}_{tn_1,1} & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ \tilde{u}_{t1,k-1} & \ldots & \tilde{u}_{tn_1,k-1} & \ldots & \tilde{u}_{tn_{k-1},k-1} & 0 & \ldots & 0 \\ \tilde{u}_{t1,k} & \ldots & \tilde{u}_{tn_1,k} & \ldots & \tilde{u}_{tn_{k-1},k} & \tilde{u}_{tn_{k-1}+1,k} & \ldots & \tilde{u}_{tn_k,k} \end{bmatrix}. \tag{4.8}$$

The specific assumption of a monotone missing data pattern on $\tilde{U}_t$ comes with the particular advantage of providing a closed form distribution for C_t. Combining the intraday return DGP in (4.7) with the innovation process in (4.8) yields

$$C_t = L_t Q \tilde{U}_t \tilde{U}_t' Q L_t', \qquad \text{with} \quad \tilde{U}_t \tilde{U}_t' \mid \bar{n} \sim \mathcal{BG}_k^I\big(\bar{n}/2, I/2\big), \tag{4.9}$$

where $\mathcal{BG}_k^I$ denotes a k-dimensional Bellman-Gamma type-I distribution with identity scale and d.o.f. vector $\bar{n} = (n_1, \ldots, n_k)'$ (see Gupta and Nagar, 2000 and Veleva, 2009, for details). By the Cholesky-scaling properties of the Bellman-Gamma we obtain

$$C_t \mid \Sigma_t, \bar{n} \sim \mathcal{R}_k\big(\bar{n}, \Sigma_t\big), \tag{4.10}$$

where $\mathcal{R}_k$ denotes the k-dimensional Riesz distribution of Hassairi and Lajmi (2001) with d.o.f. vector $\bar{n}$ and scale matrix Σ_t (see Veleva, 2009 and Andersson and Klein, 2010). The Riesz pdf, its characteristic function and the associated first and second order moments are given in the following proposition.

Proposition 2. *Let C_t denote a positive-definite $(k \times k)$ realized covariance matrix with $C_t|\Sigma_t, \bar{n} \sim \mathcal{R}_k(\bar{n}, \Sigma_t)$, where $\bar{n} = (n_1, \ldots, n_k)'$ denotes the associated d.o.f. vector*

and Σ_t *the* $(k \times k)$ *p.d. scale matrix. For* $n_i > i-1$ *the pdf of the Riesz distribution is given by*

$$f(C_t|\Sigma_t,\bar{n}) = \frac{1}{\pi^{k(k-1)/4}} \left[\prod_{i=1}^{k} \frac{\left(\frac{n_i}{2}\right)^{\frac{n_i}{2}} \left[c_{i\cdot i-1,t}\right]^{\frac{n_i-k-1}{2}}}{\Gamma\left(\frac{n_i-i+1}{2}\right) \left[\sigma_{i\cdot i-1,t}\right]^{\frac{n_i}{2}}}\right] \exp\left\{-\mathrm{tr}\left(\Sigma_t^{-\frac{\bar{n}}{2}} C_t\right)\right\}, \tag{4.11}$$

with characteristic function $\varphi_{\mathcal{R}}(\Theta) = \prod_{i=1}^{k} \left(\frac{n_i}{2\sigma_{i\cdot i-1,t}}\right)^{\frac{n_i}{2}} \left(\xi_{i\cdot i-1,t}(\Theta)\right)^{\frac{n_i}{2}}$*, where* $\xi_{i\cdot i-1,t}(\Theta)$ *is the i'th diagonal element of* D_t^{Ξ} *from the Cholesky decomposition* $\Xi_t = \left(\Sigma_t^{-\frac{\bar{n}}{2}} - \iota\Theta\right) = T_t^{\Xi} D_t^{\Xi} T_t^{\Xi\prime}$*,* ι *denotes the imaginary number,* Θ *is a symmetric* k*-dimensional matrix,* $\sigma_{i\cdot i-1,t}$*,* $i = 1,\ldots,k$*, are the diagonal elements of* D_t^{Σ} *in the Cholesky decomposition* $\Sigma_t = T_t^{\Sigma} D_t^{\Sigma} T_t^{\Sigma\prime}$*,* $c_{i\cdot i-1,t}$ *are the diagonal elements of* D_t^{C} *in the Cholesky decomposition* $C_t = T_t^{C} D_t^{C} T_t^{C\prime}$*, and*

$$\Sigma_t^{-\frac{\bar{n}}{2}} = \left[T_t^{\Sigma\prime}\right]^{-1} \mathrm{diag}\left(\frac{n_1}{2\sigma_{1\cdot 0,t}},\ldots,\frac{n_k}{2\sigma_{k\cdot k-1,t}}\right) \left[T_t^{\Sigma}\right]^{-1}. \tag{4.12}$$

We obtain $E\left[C_t|\,\Sigma_t,\bar{n}\right] = \Sigma_t$ *and*

$$\begin{aligned}\mathrm{Cov}\left[\mathrm{vec}(C_t)|\Sigma_t,\bar{n}\right] = {} & \frac{n_k}{4}(I_{k^2}+K_{kk})\left(\Sigma_t^{\frac{\bar{n}}{2}} \otimes \Sigma_t^{\frac{\bar{n}}{2}}\right) \\ & + \frac{1}{4}\sum_{i=1}^{k-1}(n_i - n_{i+1})(I_{k^2}+K_{kk})\left\{\left(\Xi_{it}^{*} \otimes (\Sigma_t^{\frac{\bar{n}}{2}} - \Xi_{it}^{*}/2)\right)\right\}(I_{k^2}+K_{kk}),\end{aligned} \tag{4.13}$$

where $\Xi_{it}^{*} = \Sigma_t^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma_t^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma_t^{\frac{\bar{n}}{2}}$*, and* $E_i = [I_i, 0_{(i\times m-i)}]'$*. Proof: See Appendix C.1.*

Note that the Riesz nests the Wishart $\mathcal{W}_k(n, \Sigma_t/n)$ for $n_1 = \cdots = n_k = n$, and – similar to the Wishart – the integer Riesz d.o.f. are readily generalized to real numbers with $n_i > i-1$.

The assumption of non-decreasing liquidity across assets, i.e., $n_1 \leq n_2 \leq \cdots \leq n_k$, identifies a liquidity-based sorting of the assets in the underlying return vector. This sorting appears natural regarding the way the Riesz DGP is constructed, but constitutes an overidentifying restriction since the existence of the Riesz pdf in Eq. (4.11) only requires $n_i > i-1$, which may be satisfied although $n_i > n_j$ for $j > i$ (see, e.g.,

Andersson and Klein, 2010). For practical applications we suggest to sort the assets according to a global liquidity measure[3], e.g, the median number of trades over a specific time horizon, and (for the sake of flexibility) relax the overidentifying restrictions on the d.o.f. to $n_i > i - 1$.

The Riesz constitutes a generalization of the Wishart to missing data structures and allows for $k - 1$ additional d.o.f. in fitting the distributional characteristics of realized covariance matrices. In particular, the measurement error covariance in (4.13) appears far more flexible in adapting the stochastic properties of potentially complex measurement errors as compared to its Wishart analogue in (4.6). Moreover, the Riesz alleviates the attenuation bias of the Wishart, since the lower boundary for the single Wishart d.o.f. $n > k - 1$ is replaced by $n_i > i - 1$ for $i = 1, \ldots, k$, which is clearly less binding in high-dimensional scenarios. For empirical applications to realized covariance data for stocks with heterogeneous liquidity structure we would expect the estimated Wishart d.o.f. $\hat{n}$ to lie somewhere in between the corresponding Riesz estimates $\hat{n}_1$ to $\hat{n}_k$. Relative to the Riesz, the Wishart would then induce a downward bias of the measurement error variance for the less liquid stocks in the liquidity sorted return vector.

It is important to note that, although the Riesz and Wishart DGPs are motivated from the standard realized covariance measure of Barndorff-Nielsen and Shephard (2004), see Eq. (4.1), which provides the basic intuition of realized covariance estimation and – under certain assumptions (see above) – allows to obtain closed form small-sample results, this does not restrict their applicability to more refined and robust realized measures such as, e.g., the composite realized kernel of Lunde et al. (2016) or the CholCov estimator of Boudt et al. (2017), which by their data-efficient nature automatically account for the heterogeneous liquidity across assets and for which small-sample distributional results are not available. Overall, the Riesz-implied measurement error covariance offers a QL framework with additional d.o.f. to adapt to the specific properties of the realized measure at hand. Our empirical application

[3]Compare Boudt et al. (2017) for a similar approach in a Cholesky realized covariance setting.

in Section 4.5 employs the data-efficient and microstructure-noise robust composite realized kernel measure of Lunde et al. (2016) and our findings show that the Riesz significantly improves the model-fit and alleviates the attenuation of measurement errors relative to the Wishart.

Alternative distributions, which have been considered for the modeling of p.d. realized covariance measures, are the inverse Wishart and the Matrix-F distribution. The inverse Wishart implies a Wishart distribution for the precision matrix, therefore imposing similar restrictions on the measurement errors as the Wishart itself (see, e.g., Grothe et al., 2019, for an application of the inverse Wishart distribution). The Matrix-F has been analyzed by Opschoor et al. (2017), Opschoor and Lucas (2019) and Vassallo et al. (2019) for the class of observation driven generalized autoregressive score (GAS) models. The authors find that replacing the Wishart by a Matrix-F (a Wishart-inverted-Wishart mixture, which nests the inverted Wishart distribution; see Konno, 1991, and Gupta and Nagar, 2000, for details) results in significant improvements in model fit, since the Matrix-F introduces a second d.o.f. parameter, which accounts for leptokurtic tail-behaviour. Recent results of Gribisch et al. (2020) however indicate that there is no evidence against the Wishart and in favor of the Matrix-F distribution for parameter-driven state-space models, since the marginalization of the conditional Wishart w.r.t. the latent state variables in Σ_t suffices to capture the tail behavior of realized covariance data.

4.3. The Riesz State-Space Model

Based on the preceding discussion we take the integrated covariance matrix Σ_t as a latent state variable observed through the realized covariance matrix C_t and combine the Riesz measurement density for C_t with a suitably chosen transition density $f(\Sigma_t|\Sigma_{1:t-1})$ in order to obtain a tractable state-space model designed to mimic the persistent dynamics of time series of p.d. realized covariance matrices.

Here we exploit that the Riesz allows for a convenient likelihood factorization which

greatly simplifies statistical inference in high-dimensional applications. The Riesz likelihood factorization is conceptually based on the Cholesky decomposition $\Sigma_t = T_t^{\Sigma} D_t^{\Sigma} T_t^{\Sigma'}$ as implied by the recursive sequence of linear regressions

$$r_{i,t} = \beta'_{i\cdot i-1,t} r_{1:i-1,t} + \epsilon_{i,t}, \quad i = 1, \ldots, k, \tag{4.14}$$

where $r_{i,t}$ denotes the day t log-return of the ith asset, $r_{1:i-1,t} = (r_{1,t}, \ldots, r_{i-1,t})'$, $\beta_{i\cdot i-1,t} = \Sigma^{-1}_{[i-1]t}\sigma_{i-,t}$, with $\sigma_{i-,t} = (\sigma_{i1,t}, \ldots, \sigma_{ii-1,t})'$, and $\text{Var}[\epsilon_{it}] = \sigma_{i\cdot i-1,t} = \sigma_{ii,t} - \sigma'_{i-,t}\Sigma_{[i-1],t}\sigma_{i-,t}$, where $\Sigma_{[i-1],t}$ indicates the upper left $(i-1 \times i-1)$ submatrix of Σ_t, $i = 2, \ldots, k$. For $i = 1$ we obtain $\beta_{i\cdot i-1,t} = 0$ and $\sigma_{1\cdot 0,t} = \sigma_{11,t}$. With $r_t = (r_{1,t}, \ldots, r_{k,t})'$ and $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{k,t})'$ we have $\epsilon_t = V_t r_t$, where V_t denotes a lower triangular matrix of negative β-coefficients and ones on the diagonal. The Cholesky decomposition of Σ_t then results from $D_t^{\Sigma} = \text{Var}(\epsilon_t) = V_t \Sigma_t V_t'$ and $T_t^{\Sigma} = V_t^{-1}$ (compare Darolles et al., 2018).

The Riesz likelihood factorization is detailed in the following proposition.

Proposition 3. *The conditional Riesz pdf $f(C_t|\Sigma_t, \bar{n})$ provided in Proposition 2 can be equivalently expressed as*

$$f(C_t|\Sigma_t, \bar{n}) = \prod_{i=1}^{k} f(c_{i\cdot i-1,t}|\sigma_{i\cdot i-1,t}, n_i) f(b_{i\cdot i-1,t}|\beta_{i\cdot i-1,t}, n_i, \sigma_{i\cdot i-1,t}, C_{[i-1],t}) \tag{4.15}$$

with

$$c_{i\cdot i-1,t} \mid \sigma_{i\cdot i-1,t}, n_i \sim \Gamma\left(\frac{n_i - i + 1}{2},\ \frac{2}{n_i}\sigma_{i\cdot i-1,t}\right), \tag{4.16}$$

$$b_{i\cdot i-1,t} \mid \beta_{i\cdot i-1,t}, n_i, \sigma_{i\cdot i-1,t}, C_{[i-1],t} \sim \mathcal{N}\left(\beta_{i\cdot i-1,t},\ \frac{1}{n_i}\sigma_{i\cdot i-1,t} C^{-1}_{[i-1],t}\right), \tag{4.17}$$

where $b_{i\cdot i-1,t} = C^{-1}_{[i-1]t} c_{i-,t}$, $\beta_{i\cdot i-1,t} = \Sigma^{-1}_{[i-1]t}\sigma_{i-,t}$, $b_{1\cdot 0,t} = \beta_{1\cdot 0,t} = 0\ \forall\ t$, $\sigma_{i-,t} = (\sigma_{i1,t}, \ldots, \sigma_{ii-1,t})'$, $\Sigma_{[i-1],t}$ denotes the upper left $(i-1 \times i-1)$ submatrix of Σ_t, $\Gamma(\cdot,\cdot)$ denotes the gamma distribution in shape-scale parameterization, and $C_{[i-1],t}$ and $c_{i-,t}$ are defined by analogy. Proof: See Appendix C.2.

The likelihood representation in Proposition 3 factorizes the conditional density of C_t in k multiplicative components which treat the realized regression coefficients $b_{i\cdot i-1,t}$ and realized error variances $c_{i\cdot i-1,t}$ as noisy measurements for their population counterparts $\beta_{i\cdot i-1,t}$ and $\sigma_{i\cdot i-1,t}$, which in turn resemble the Cholesky decomposition of Σ_t. The likelihood factorization suggests to model the dynamics of the matrix-variate p.d. state variable Σ_t by treating the regression betas $\beta_{i\cdot i-1,t}$ and error variances $\sigma_{i\cdot i-1,t}$ as unobserved dynamic state processes which can be modeled conveniently in real space, preserving positivity of Σ_t without parametric restrictions.[4] Since the k likelihood components represent mutually independent functions in the asset-specific states, combining them with independent prior assumptions allows for the implementation of parallelized computing routines for joint Bayesian MCMC sampling of the k state sequences $\left\{\sigma_{i\cdot i-1,t}, \beta_{i\cdot i-1,t}\right\}_{t=1}^{T}$, $i = 1, \ldots, k$. The likelihood factorization in Proposition 3 therefore greatly simplifies statistical inference in high-dimensional state-space scenarios. Details on the MCMC implementation of the RSS model are given in Section 4.4 and Appendix C.3.

We now complete the Riesz state-space model by specifying the transition densities for the state processes. Here we combine independent random walk (RW) processes for the $k(k-1)/2$ elements of $\{\beta_{i\cdot i-1,t}\}_{i=1}^{k}$ with heterogeneous autoregressive (HAR) processes (Corsi, 2009) for the logarithmic error variances $\{\log \sigma_{i\cdot i-1,t}\}_{i=1}^{k}$. While the RW assumption is rather typical for time-varying coefficient models and preserves sparsity in model parametrization, the observed logarithmic error variances $\log c_{i\cdot i-1,t}$ feature a particularly strong long-memory like autocorrelation structure which appears to be well approximated by HAR dynamics (cf. Gribisch et al., 2020).

For $x_{it} = \log \sigma_{i\cdot i-1,t}$, $i = 1, \ldots, k$, we impose mutually independent Gaussian HAR processes of the form

$$x_{it} - \gamma_i = \phi_{i1}\bar{x}_{i[t-1:t-1]} + \phi_{i2}\bar{x}_{i[t-1:t-5]} + \phi_{i3}\bar{x}_{i[t-1:t-22]} + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, \nu_i^2), \quad (4.18)$$

[4] The dynamic modeling of Cholesky factors is frequently applied for multivariate volatility modeling, see, e.g., Chiriac and Voev (2011), Lopes et al. (2016), Shirota et al. (2017), and Darolles et al. (2018).

where $\bar{x}_{i[t-1:t-h]} = \sum_{\tau=1}^{h}(x_{it-\tau} - \gamma_i)/h$ for $h = 1, 5, 22$ represents daily, weekly and monthly lags, respectively. The process is initiated by the stationary distribution of x_{it} as implied by the HAR parameters $\gamma_i, \phi_{i1}, \phi_{i2}, \phi_{i3}$ and ν_i.

For the $i-1$ distinct elements in $\beta_{i\cdot i-1,t} = (\beta_{1i,t}, \ldots, \beta_{i-1i,t})'$ we impose mutually independent Gaussian RW processes

$$\beta_{ji,t} = \beta_{ji,t-1} + \zeta_{ji,t}, \qquad \zeta_{ji,t} \sim \mathcal{N}(0, \tau_{ji}^2), \tag{4.19}$$

for $j = 1, \ldots, i-1$, with initial value $\beta_{ji,0} \mid \bar{\beta}_{ji}, p_{ji0} \sim \mathcal{N}(\bar{\beta}_{ji}, p_{ji0})$.

The vector of parameters, denoted by θ, for the Riesz State-Space (RSS) model as defined by Eqs. (4.15), (4.18) and (4.19) consists of $k + 5k + 3k(k-1)/2$ parameters, which are the k Riesz d.o.f. parameters, the HAR-parameters $(\gamma, \phi_1, \phi_2, \phi_3, \nu^2)$ for the k state processes $\{x_{it}\}$ and the RW parameters $(\bar{\beta}, p_0, \tau^2)$ for the $k(k-1)/2$ time-varying regression coefficients $\{\beta_{ji,t}\}$. Hence, for a large number of assets the actual amount of model parameters is fairly huge. For instance with $k = 30$ assets as in our empirical application below we have for the RSS model 465 state processes and 1485 parameters to be estimated. However, as we use a data set covering $T = 3272$ trading days with $Tk(k+1)/2 = 1,521,480$ covariance observations we have 1024 observations per parameter, which can be expected to provide enough information for reliable statistical inference.

Nevertheless, the unrestricted state dynamics in (4.18) and (4.19) clearly involve a risk of overfitting the covariance dynamics since the number of model parameters grows quadratically in the number of assets k. In the Bayesian RSS setting this risk is readily alleviated by imposing shrinkage priors on the RW parameters of the $k(k-1)/2$ beta coefficients, automatically reducing time-varying coefficients to constant ones. This greatly reduces the number of significant model parameters and helps to reduce the impact of estimation noise on the model's forecasting performance. An alternative way of imposing sparsity on the model structure is to rely on common factor structures (see, e.g., Gribisch et al., 2020). As opposed to a factor approach, the shrinkage RSS

specification has the advantage of being completely data-driven and can therefore be expected to be more flexible in fitting the observed covariance dynamics. For the RSS model it turns out that the implementation of shrinkage prior restrictions is essentially straightforward. Details on the implementation of the shrinkage approach and the Bayesian MCMC scheme are provided in the upcoming section and in Appendix C.3.

4.4. Bayesian Posterior Analysis and Forecasting

We apply MCMC methods for a Bayesian posterior analysis of the RSS model and use the Gibbs approach to simulate from the joint posterior of the parameters and states

$$\pi(\theta, \{x_{i,1:T}\}, \{\beta_{i\cdot i-1,1:T}\}|C_{1:T}). \tag{4.20}$$

In the following we outline the MCMC algorithm for the RSS model without shrinkage priors, labeled *basic RSS* model, and provide details on the minor modifications needed in order to introduce shrinkage prior restrictions on the time-varying beta process in (4.19), resulting in the *shrinkage RSS* model.

4.4.1. MCMC sampling

The properties of the Riesz distribution and the dynamic model structure of the RSS model allow for fast and numerically efficient Gibbs sampling. In particular, the factorization of the measurement density in (4.15) combined with independent priors for the structural model parameters and state processes implies that the RSS model can be devoted in k functionally independent state-space models, one for each of the k state sequences $\{(\beta'_{i\cdot i-1,t}, \sigma_{i\cdot i-1,t})'\}_{t=1}^{T}$, $i = 1, \ldots, k$.

The Gibbs sampling algorithm for Bayesian MCMC estimation of the complete RSS model can then be decomposed into k independent Gibbs samplers, each updating the structural model parameters and states associated with the respective state-space model. The k Gibbs samplers can be run completely parallel on multi-core computers,

resulting in a significant reduction of the computational burden, which is typically involved in high-dimensional sate-space modeling. It is important to note that this parallelization is not possible if the Riesz is replaced by a Wishart measurement density, since the single Wishart d.o.f. parameter n induces functional dependence between the k state-space models, such that parallelization of the Gibbs sampler is only possible for the single Gibbs sweep, conditional on the previously simulated d.o.f..

Our proposed MCMC implementation for the basic RSS model consists of the following Gibbs sampling steps 1.) – 3.), parallelized over $i = 1, \ldots, k$:

1.) Sampling of the beta states $\beta_{i\cdot i-1,0:T}$ and $\bar{\beta}_{ji}$, τ_{ji}, p_{ji0} for $j = 1, \ldots, i-1$,

2.) sampling of the log variance states $x_{i,1:T}$ and γ_i, ϕ_{1i}, ϕ_{2i}, ϕ_{3i}, ν_i^2,

3.) sampling of the Riesz degree of freedom parameter n_i.

The ith MCMC algorithm, $i = 1, \ldots, k$, repeatedly draws from the full conditional posterior distributions associated with steps 1.) – 3.), which are detailed below. After dropping the draws from the first cycles as burn-in we use the draws from the next S cycles for the purpose of approximating the joint posterior in Eq. (4.20). Bayesian point estimates (posterior means) of the model parameters and latent state variables are then obtained as sample averages over the corresponding Gibbs draws.

In step 1.) for full conditional sampling of the beta states it proves advantageous in terms of mixing and convergence of the Markov chains to re-write the state equation in (4.19) into its so-called *non-centered parameterization* (see, e.g., Bitto and Frühwirth-Schnatter, 2019, and Huber et al., 2020). Eq. (4.19) is then reformulated as $\beta_{ji,t} = \bar{\beta}_{ji} + \tau_{ji}\tilde{\beta}_{ji,t}$, $t = 0, \ldots, T$, where $\tilde{\beta}_{ji,t}$ is an independent RW process with iid standard normal increments,

$$\tilde{\beta}_{ji,t} = \tilde{\beta}_{ji,t-1} + \tilde{\zeta}_{ji,t}, \qquad \tilde{\zeta}_{ji,t} \overset{iid}{\sim} \mathcal{N}(0,1), \tag{4.21}$$

and initial value $\tilde{\beta}_{ji,0} \mid \tilde{p}_{ji0} \sim \mathcal{N}(0, \tilde{p}_{ji0})$, such that $p_{ji0} = \tau_{ji}^2 \tilde{p}_{ji0}$. Using the above transformation and conditioning on $\bar{\beta}_{ji}$, τ_{ji}, $\tilde{p}_{ji0}$, for $j = 1, \ldots, i-1$, n_i and $x_{i,1:T}$, we

obtain from (4.17) the linear Gaussian measurement equation

$$b_{i\cdot i-1,t} = \bar{\beta}_{i\cdot i-1} + T_i \tilde{\beta}_{i\cdot i-1,t} + \varepsilon_{i\cdot i-1,t}, \tag{4.22}$$

where $\bar{\beta}_{i\cdot i-1} = (\bar{\beta}_{1i}, \ldots, \bar{\beta}_{i-1i})'$, $T_i = \text{diag}(\tau_{1i}, \ldots, \tau_{i-1i})$ and $\varepsilon_{i\cdot i-1,t}$ is a zero mean Gaussian innovation vector with conditional covariance given in (4.17). The measurement equation (4.22) together with the state transitions for the $\tilde{\beta}_{ji,t}$'s in Eq. (4.21) defines a linear Gaussian state-space model for the $(i-1)$-dimensional process $\tilde{\beta}_{i\cdot i-1,0:T}$. Hence efficient simulation of the trajectory of normalized regression coefficients $\tilde{\beta}_{i\cdot i-1,0:T}$ can be carried out via the precision sampling algorithm of Chan and Jeliazkov (2009). In order to increase the computational efficiency and speed of simulation – in particular for high-dimensional applications – we follow the ideas of Carriero et al. (2019), Koop et al. (2019) and Kastner and Huber (2020), and employ the Cholesky decomposition $C^{-1}_{[i-1],t} = H_t \Delta_t H_t'$, which allows to augment the jth equation in the system (4.22) with the contemporaneous values of the first $j-1$ elements in $\varepsilon_{i\cdot i-1,t}$, $j = 1, \ldots, i-1$, i.e.,

$$b_{ji,t} = \bar{\beta}_{ji} + \tau_{ji}\tilde{\beta}_{ji,t} + \sum_{l=1}^{j-1} h_{lj,t}\tilde{\varepsilon}_{li,t} + \tilde{\varepsilon}_{ji,t}, \qquad \tilde{\varepsilon}_{ji,t} \sim \mathcal{N}(0, \sigma_{i\cdot i-1,t}\delta_{j\cdot j-1,t}/n_i), \tag{4.23}$$

where $\tilde{\varepsilon}_{li,t}$ and $\tilde{\varepsilon}_{ji,t}$ are independent for $l \neq j$, $\delta_{j\cdot j-1,t}$ denotes the jth diagonal element of Δ_t and $h_{lj,t}$ is the lth entry in the jth row of H_t. This allows to sample the elements of $\beta_{i\cdot i-1,t}$ via a sequence of one-dimensional precision samplers.[5]

For the basic RSS model sampling of $\bar{\beta}_{ji}$ and τ^2_{ji} can be conducted conditionally on $\beta_{i\cdot i-1,0:T}$ and $x_{i,1:T}$, by applying standard conjugate Normal and inverse Gamma priors, respectively.

For the shrinkage RSS specification we implement a hierarchical Normal-Gamma shrinkage prior for τ_{ji}, as recently proposed by Bitto and Frühwirth-Schnatter (2019). The Normal-Gamma prior allows for a data-driven shrinkage of the beta process variances τ^2_{ji} to zero, resulting in time-constant beta-coefficients over the entire observation

[5] After sampling $\tilde{\beta}_{i\cdot i-1,0:T}$ we obtain a draw for $\beta_{i\cdot i-1,0:T}$ by the transformation $\beta_{ji,t} = \bar{\beta}_{ji} + \tau_{ji}\tilde{\beta}_{ji,t}$.

period.[6] We further replace the Gaussian prior for $\bar{\beta}_{ji}$ by a second Normal-Gamma prior, inducing shrinkage of the initial values of the beta process to zero. This allows to shrink static coefficients to coefficients which are not significant over the entire observation period.

The hierarchical Normal-Gamma shrinkage priors for $\bar{\beta}_{ij}$ and τ_{ij} are given as follows:

$$\begin{aligned}
&\bar{\beta}_{ji} \mid s_{ji}^{\beta} \sim \mathcal{N}(0, s_{ji}^{\beta}), \quad s_{ji}^{\beta} \mid a_i^{\beta}, b_i^{\beta} \sim \Gamma(a_i^{\beta}, 2/(a_i^{\beta} b_i^{\beta})), \quad a_i^{\beta} \sim \mathcal{E}(d_1), \quad b_i^{\beta} \sim \Gamma(d_2, 1/d_3), \\
&\tau_{ji} \mid s_{ji}^{\tau} \sim \mathcal{N}(0, s_{ji}^{\tau}), \quad s_{ji}^{\tau} \mid a_i^{\tau}, b_i^{\tau} \sim \Gamma(a_i^{\tau}, 2/(a_i^{\tau} b_i^{\tau})), \quad a_i^{\tau} \sim \mathcal{E}(d_1), \quad b_i^{\tau} \sim \Gamma(d_2, 1/d_3),
\end{aligned} \tag{4.24}$$

where $i = 2, \ldots, k$, $j = 1, \ldots, i-1$, and $\mathcal{E}$ denotes the exponential distribution. $a_i^{\bullet}$ ($\bullet \in \{\beta, \tau\}$) controls the tail behavior of the prior, with smaller values inducing heavier tails, and $b_i^{\bullet}$ controls the global level of shrinkage, with larger values inducing more prior mass for $s_{ji}^{\bullet}$ at zero. Full conditional updating of the hierarchical prior parameters $a_i^{\bullet}$, $b_i^{\bullet}$, and $\{s_{ji}^{\bullet}\}$ is detailed in the Appendix C.3.[7] Conditional on $\{s_{ji}^{\bullet}\}$, updating of $\{\bar{\beta}_{ji}\}$ and $\{\tau_{ji}\}$ exploits that Eq. (4.22) obeys a Gaussian seemingly unrelated regression structure, hence the resulting posteriors are Gaussian. To further increase the sampling efficiency we redraw the values of $\bar{\beta}_{ji}$ and τ_{ji} through interweaving into the state equation of the centered parameterization in Eq. (4.19) (see, e.g., Yu and Meng, 2011, and Bitto and Frühwirth-Schnatter, 2019, for details on interweaving). For $\{\tilde{p}_{ji0}\}$ we impose a conjugate inverse Gamma prior.

In step 2.) the full conditional posterior for $x_{i,1:T}$ obtains as

$$\begin{aligned}
&\pi(x_{i,1:T} | \beta_{i\cdot i-1,1:T}, \gamma_i, \phi_{i1}, \phi_{i2}, \phi_{i3}, \nu_i^2, n_i, C_{1:T}) \\
&\quad \propto \prod_{t=1}^{T} \exp\left\{ -\frac{n_i}{2} \left[x_{it} + e^{-x_{it}} (c_{i\cdot i-1,t} + (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})' C_{[i-1],t} (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})) \right] \right\} \\
&\qquad \times f_{\mathcal{N}}\Big(x_{it} \,\Big|\, \gamma_i + \phi_{i1} \bar{x}_{i[t-1:t-1]} + \phi_{i2} \bar{x}_{i[t-1:t-5]} + \phi_{i3} \bar{x}_{i[t-1:t-22]} \,,\, \nu_i^2 \Big),
\end{aligned} \tag{4.25}$$

[6] See e.g. Park and Casella (2008), Carvalho et al. (2010), and Frühwirth-Schnatter and Wagner (2011) for details on shrinkage priors.

[7] See also Appendix C.6 for a note on random number generation for the Generalized Inverse Gaussian distribution, which is needed for the sampling scheme of the shrinkage prior parameters.

which resembles a conditional, nonlinear and non-Gaussian state-space model for $x_{i,1:T}$. We sample the trajectory $x_{i,1:T}$ from its posterior in one block by employing the Particle Gibbs with Ancestor sampling (PG-AS) approach of Lindsten et al. (2014).

For sampling of the HAR parameters we select independent conjugate Normal-inverted-Gamma priors and simulate directly from their full conditional Gaussian distributions given the $x_{i,1:T}$ state trajectory (see Gribisch et al., 2020).

In step 3.) we select a uniform prior $p(n_i)$ on $(i-1,\infty)$ and sample the Riesz degree of freedom parameter n_i using the Metropolis-Hastings (MH) algorithm. The target density is given by

$$\begin{aligned}&\pi(n_i \,|\{x_{i,1:T}\},\{\beta_{i\cdot i-1,1:T}\},C_{1:T})\\ &\qquad\propto p(n_i)\prod_{t=1}^{T} f(c_{i\cdot i-1,t}|\sigma_{i\cdot i-1,t},n_i)f(b_{i\cdot i-1,t}|\beta_{i\cdot i-1,t},\sigma_{i\cdot i-1,t},n_i,C_{[i-1]t}).\end{aligned} \tag{4.26}$$

The proposal density for the MH step is chosen as a tailored t distribution. We select for the mean and variance the mode and inverse Hessian obtained by numerical optimization of the logarithmic target density kernel in (4.26) over n_i, and set the d.o.f. parameter to 18.

The chosen hyperparameters for the basic RSS model and its shrinkage version are given in Appendix C.3.

4.4.2. Model comparison

For the purpose of comparing alternative model specifications we rely upon the Deviance Information Criterion (DIC) based on the likelihood function (Spiegelhalter et al., 2002).

Let θ denote the list of all static model parameters. Then the conditional DIC is given by

$$\text{DIC} = -2\log p(C_{1:T}|\hat{\theta}) + 2p_D, \tag{4.27}$$

with small values of the criterion preferred. The term $p(C_{1:T}|\hat{\theta})$ represents the likelihood function evaluated at the posterior estimates for the parameters rewarding good fits, and p_D is the effective sample size, penalizing good fits achieved by means of excessively rich parameterizations. The effective sample size is defined as

$$p_D = -2\left[\mathrm{E}_{post}[\log p(C_{1:T}|\theta)] - \log p(C_{1:T}|\hat{\theta})\right], \tag{4.28}$$

where $\mathrm{E}_{post}[\log p(C_{1:T}|\theta)]$ is the mean of the conditional log-likelihood function taken w.r.t. the posterior distribution of θ.

The conditional likelihood function in Eqs. (4.27) and (4.28) obtains as

$$p(C_{1:T}|\theta) = \int \prod_{t=1}^{T} f(C_t \mid \Sigma_t; \theta) f(\Sigma_t \mid \Sigma_{1:t-1}; \theta)\ \mathrm{d}\Sigma_{1:T}. \tag{4.29}$$

Under the factorization of the Riesz density given in Proposition 3 and the independent priors for the state processes $\{x_{i,1:T}, \beta_{i\cdot i-1,1:T}\}$ the conditional likelihood in Eq. (4.29) as a function in $\{x_{i,1:T}, \beta_{i\cdot i-1,1:T}\}$ factorizes into k functionally independent components, such that $p(C_{1:T}|\theta)$ is proportional to

$$\prod_{i=1}^{k} \iint \Bigg[\prod_{t=1}^{T} \exp\left\{ -\frac{n_i}{2}\left[x_{it} + e^{-x_{it}}(c_{i\cdot i-1,t} + (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t}) C_{[i-1],t} (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})' \right] \right\} \\ \times f(x_{it} \mid x_{i,t-22:t-1}) f(\beta_{i\cdot i-1,t} \mid \beta_{i\cdot i-1,t-1}) \Bigg] \mathrm{d}\beta_{i\cdot i-1,1:T}\ \mathrm{d}x_{i,1:T}, \tag{4.30}$$

where $f(x_{it} \mid x_{i,t-22:t-1})$ and $f(\beta_{i\cdot i-1,t} \mid \beta_{i\cdot i-1,t-1})$ are the Gaussian transition densities of the latent states. For a given value of θ the k integrals w.r.t. the $\beta_{i\cdot i-1,1:T}$'s and $x_{i,1:T}$'s can be taken as likelihood functions of independent mixed linear/nonlinear (non-)Gaussian state-space models, which can be easily evaluated in parallel using a Rao-Blackwellized particle filter (RBPF) approach (see, e.g., Schön et al., 2005).

Using this RBPF for the likelihood evaluation we can estimate the posterior mean of the conditional log-likelihood function in Eq. (4.28) by the arithmetic mean over the Gibbs draws of $\{\theta^{(i)}\}_{i=1}^{S}$, that is $\widehat{\mathrm{E}}_{post}\left[\log p(C_{1:T}|\theta)\right] = \frac{1}{S}\sum_{i=1}^{S} \log p(C_{1:T}|\theta^{(i)})$.

4.4.3. Forecasting

Using the Gibbs sampler outlined in Section 4.4.1, we can perform out-of-sample point- and density forecasting for the realized covariance matrix of asset returns C_{t+1}. A density forecast for C_{t+1} obtains as

$$p(C_{t+1}|C_{1:t}) = \int f(C_{t+1}|\Sigma_{t+1};\theta)\ f(\Sigma_{t+1}|\Sigma_{1:t};\theta)\ \pi(\Sigma_{1:t},\theta|C_{1:t})\ \mathrm{d}\Sigma_{1:t+1}\mathrm{d}\theta, \qquad (4.31)$$

where $\pi(\Sigma_{1:t},\theta|C_{1:t})$ denotes the joint posterior density of the state variables in $\Sigma_{1:t}$ and the parameter vector θ for the observed data up to period t, and $f(C_{t+1}|\Sigma_{t+1};\theta)$ is the measurement density for the realized covariance of the assets. The forecasting density (4.31) evaluated at the ex-post observed value for C_{t+1} defines the period $t+1$ Bayesian predictive density (Geweke, 2005).

The predictive density according to Eq. (4.31) can be approximated via MC integration, i.e.,

$$p(C_{t+1}|C_{1:t}) \simeq \frac{1}{S}\sum_{i=1}^{S} f(C_{t+1}|\Sigma_{t+1}^{(i)};\theta^{(i)}), \qquad (4.32)$$

where $\{\Sigma_{t+1}^{(i)}\}$ and $\theta^{(i)}$ are simulated draws from the convolution of the posterior and transitions, $f(\Sigma_{t+1}|\Sigma_{1:t};\theta)\ \pi(\Sigma_{1:t},\theta|C_{1:t})$, based on Gibbs simulations from the joint posterior $\pi(\Sigma_{1:t},\theta|C_{1:t})$. Using the simulated draws $\{\Sigma_{t+1}^{(i)}\}$ the point forecast of C_{t+1} given by $\mathrm{E}(C_{t+1}|C_{1:t}) = \mathrm{E}(\Sigma_{t+1}|C_{1:t})$ can be approximated by

$$\mathrm{E}(C_{t+1}|C_{1:t}) \simeq \frac{1}{S}\sum_{i=1}^{S} \Sigma_{t+1}^{(i)}. \qquad (4.33)$$

Note that the point forecast in Eq. (4.33) does account for parameter estimation uncertainty. However, since we shall compare the predictive performance of the RSS model with alternative forecasting approaches, which are not estimated by Bayesian techniques and for which it is not clear how to obtain the posterior distribution, we rely on the alternative conditional point forecast $\mathrm{E}(C_{t+1}|C_{1:t},\theta)$. For its computation

we set the parameters θ equal to their posterior mean estimates $\hat{\theta}_{1:t}$ based on the data observed from 1 to t and simulate $\{\Sigma_{t+1}^{(i)}\}$ from $f(\Sigma_{t+1}|\Sigma_{1:t};\hat{\theta}_{1:t})\ \pi(\Sigma_{1:t}|C_{1:t},\hat{\theta}_{1:t})$.

4.5. Empirical Application

4.5.1. Data

We use the RSS model to analyze the dynamics of daily realized covariance matrices for 30 stocks traded at the NYSE. The stocks are sorted by increasing liquidity and cover nine relevant industry sectors according to the Global Industry Classification Standard (GICS). The list of stocks is provided in Table 4.1.

The daily realized covariance matrices C_t are computed by the composite realized kernel method of Lunde et al. (2016) based on 1-minute intraday returns.[8] The composite realized kernel performs a variance/correlation decomposition of the covariance matrix and estimates the individual asset variances by the noise-robust univariate realized kernel method of Barndorff-Nielsen et al. (2008), which exploits all available data information for the respective asset. The estimation of the correlations for all asset pairs is based on bivariate realized kernels with refresh time sampling (see Barndorff-Nielsen et al., 2011, and Lunde et al., 2016, for details). Hence by the individual computation of the realized variances and the refresh time sampling for all bivariate relationships the composite realized kernel directly accounts for the heterogeneous liquidity of the assets.

The data comprises $30 \times 31/2 = 465$ time series of realized variances and covariances for the sample period from January 2, 2002 to December 31, 2014, covering 3272 trading days. See Figure 4.1 for time-series plots of the realized variances for three randomly selected stocks (Dover, Nike, and Microsoft) together with the according sample autocorrelation functions (ACFs). The time-series feature a very strong serial correlation structure, which is characteristical for all assets in the data set.

[8] The intraday return data for the 30 stocks has been obtained from QuantQuote.com.

4.5.2. Estimation results

We estimate the RSS model with and without shrinkage priors for the complete sample covering 3272 trading days. The prior assumptions are overall fairly uninformative and detailed in the Technical Appendix C.3.

For parameter estimation we run the MCMC algorithm proposed in Section 4.4.1 for 25,000 iterations, where the first 15,000 are discarded. The PG-AS procedure in the MCMC updating step for the state-trajectories $\{x_{i,1:T}\}$ is implemented using 50 particles. For the computation of the DIC criterion we run the RBPF using 25,000 particles. The MCMC algorithm is implemented in MATLAB. In order to evaluate the sampling efficiency of the proposed MCMC procedure for estimating the parameters of the RSS model we compute the inefficiency factors for the posterior samples of the parameters.[9] Its benchmark value expected under perfect mixing of the MCMC draws for the parameters is equal to one. For the RSS Shrinkage model the values of the inefficiency factors range from 1.01 to 74.02 with an average value of 20.74 indicating a high sampling efficiency with a very fast mixing rate.

Our in-sample analysis is threefold: We start with analyzing the impact of the shrinkage prior specification on the estimated β-dynamics under the RSS model. We then turn to the empirical effect of the attenuation bias of the Wishart relative to the Riesz distribution and finally assess the in-sample fit of various model specifications using the DIC criterion.

Figure 4.2 illustrates the impact of the Normal-Gamma shrinkage prior restrictions on the β-estimates. Panels a) to e) report five selected time series of realized regression coefficients b_{ji} together with corresponding smoothed β-estimates (posterior means). The estimates are obtained under the RSS model with and without shrinkage priors on the conditional RW variances. Panel f) summarizes the estimated RW standard deviations for all 435 β-processes via histograms. The results show a strong

[9] The inefficiency factor for the posterior sample of a parameter is defined as IF $= 1 + 2[B/(B-1)] \sum_{j=1}^{B} K(j/B)\, \hat{\rho}(j)$, where $\hat{\rho}(j)$ denotes the lag j sample autocorrelation of the MCMC draws of the parameter, $K(\cdot)$ is the Parzen kernel function and B is the bandwidth which we set equal to $B = 100$ (for details, see Kim et al., 1998).

data-driven shrinkage effect for those sequences, where the realized betas imply only moderate time-variation accompanied with relatively high measurement error noise (panels c) to e)). The realized betas in panels a) and b) in contrast indicate a powerful signal for time-variation in the β processes, resulting in virtually identical beta estimates across the two RSS specifications.

We now analyze the Wishart-implied restriction of equal d.o.f. across assets. For this purpose we estimate a Wishart State-Space (WSS) model, which is obtained from the RSS by setting $n_1 = n_2 = \cdots = n_k = n$. Figure 4.3 reports the d.o.f. estimates together with corresponding 95% posterior high-density regions. The Riesz d.o.f. show highly significant variation and an increasing trend, ranging from 9.63 to 223.18. The assumption of constant d.o.f. is clearly rejected by the data. The estimated Wishart d.o.f. amounts to $\hat{n} = 173.06$ with a very tight 95% posterior high-density region and results in a significant attenuation of the implied measurement errors for the less liquid assets relative to the Riesz (compare the discussion in Section 4.2).

Figure 4.4 provides an illustration of the Wishart-implied attenuation bias. The Figure reports time-series of measurement error standard deviations for the 1st, 2nd, 10th, 20th and 30th asset for the time period from January 2nd, 2008, to December 30th, 2009, computed at estimated d.o.f. and smoothed estimates of the integrated variances according to the RSS Shrinkage model (blue) and the WSS Shrinkage model (red). As expected, the Wishart attenuates the dispersion of the measurement error, resulting in far lower standard deviations as compared to the Riesz – in particular for the less liquid assets, which are sorted first in the underlying return vector.

Figure 4.5 illustrates the effect of the Wishart-implied attenuation on the state dynamics. Panel a) indicates a significant increase in the conditional standard deviation of the beta-coefficients. The induced state variation compensates for the reduced measurement error and may negatively affect the forecasting performance. Similar results are found for the 30 $\sigma_{i.i-1,t}$-processes (compare panel e)). Moreover, the Wishart induces a significant downward bias of the lag-1 HAR parameter ϕ_1 which is of considerable importance for the short-term forecasting performance.

Table 4.2 contains the DIC values for RSS and WSS specifications with and without shrinkage. Since the computation of the DIC is computationally prohibitive for medium- and high-dimensional settings, we focus on six randomly chosen five-dimensional sub-portfolios from the 30-dimensional data set of Section 4.5.1. We also report DICs for the Uhlig Extension (UE) state-space model of Windle and Carvalho (2014) as a natural Bayesian state-space competitor. The UE model combines the Wishart measurement for C_t in Eq. (4) with a generalized matrix-variate Beta type-I transition density for $\Omega_t = \Sigma_t^{-1}$, resulting in a Matrix-F predictive distribution.[10] In particular the model assumes

$$\Omega_t = \Omega_{t-1}^{1/2} \Psi_t \Omega_{t-1}^{1/2'} / \lambda, \qquad \text{with} \quad \Psi_t \sim \mathcal{B}_k^I(m/2, n/2), \tag{4.34}$$

where $A^{1/2}$ denotes the lower triangular Cholesky factor of A and λ is a scalar-valued smoothing parameter, which is set to $\lambda = [1 + n/(m - k - 1)]^{-1}$ as suggested by the authors. Windle and Carvalho (2014) derive a closed form expression for the predictive distribution of C_t, which turns out to be of Matrix-F type. Bayesian estimation of the UE model is conducted based on the MCMC scheme proposed by the authors.

The DIC results show a clear preference for the RSS Shrinkage specification. The Wishart measurement and the Matrix-F UE model are outperformed in any case. This is particularly remarkable for the UE model, which is very sparsely parameterized. Hence the flexible RSS state dynamics together with the Riesz measurement density appear to outweigh the penalty for the comparably huge number of model parameters. For the RSS and the WSS model the shrinkage specifications provide the best in-sample fit in five and four out of six cases, respectively.

Summarizing the in-sample findings, the Wishart-implied restriction of equal d.o.f. for all assets is clearly rejected by the data and implies a considerable reduction of the measurement error variance for the less liquid assets, inducing spurious state variation accompanied by a significant reduction of short-term persistence. The WSS state

[10] See Gupta and Nagar (2000), for details on the matrix-variate Beta type-I distribution.

dynamics are bound to capture the suppressed measurement error variation, inducing overfitting and presumably poor out-of-sample performance. The Riesz distribution alleviates the attenuation of measurement errors and the RSS specification provides a significant improvement in in-sample fit relative to the according WSS specification and the sparsely parameterized Matrix-F UE model of Windle and Carvalho (2014).

4.5.3. Out-of-sample forecasting results

We now analyze the out-of-sample forecasting performance of the RSS model for the realized covariance matrix of asset returns C_t. As competing forecasting approaches we consider a simple RW forecast for the realized covariance matrix, the Exponentially Weighted Moving Average (EWMA) model (Morgan, 1996), the UE Matrix-F state-space approach of Windle and Carvalho (2014) and the Lasso VAR approach of Callot et al. (2017). The EWMA is popular in industry practice while the Lasso VAR approach is found to be highly effective in predicting large covariance matrices of asset returns. While the RW, EWMA and Lasso VAR approaches are observation-driven forecasting models, the Matrix-F UE approach represents a parameter-driven state-space competitor to the RSS model (see Section 4.5.2 for details on the UE model).

The EWMA is given by

$$\mathrm{E}(C_{t+1} \mid C_{1:t}) = (1-\lambda)C_t + \lambda \mathrm{E}(C_t \mid C_{1:t-1}), \tag{4.35}$$

where we set the smoothing parameter λ to its typically selected value of 0.96 (cf. Callot et al., 2017).

The Lasso VAR model of Callot et al. (2017) employs a vector autoregression (VAR) of order ℓ for the unique lower-triangular elements of the matrix logarithm of C_t.[11] We consider VAR(1), VAR(20) and VAR-HAR structures, where the VAR-HAR involves a regression of the matrix-log on its first lag, the sum of the first 5 lags, and the

[11] For details on the matrix logarithm see, e.g., Bauer and Vorkink (2011) and Gribisch (2018).

sum of the first 22 lags (cf. Corsi, 2009). Estimation of the VAR parameters is conducted equation-wise using the Lasso of Tibshirani (1996), where the penalty parameter is chosen by the Bayesian Information Criterion (cf. Callot et al., 2017). Positive definiteness of the covariance forecast is ensured via the inverse of the matrix logarithm.

We focus on 1-day-ahead predictions $\widehat{C}_t = \mathrm{E}(C_{t+1} \mid C_{t-2519+1:t}, \hat{\theta})$, which are obtained by re-estimating the model parameters every 10 trading days on a rolling window with 2519 daily observations and then producing a sequence of new 1-day-ahead forecasts based on the updated parameter estimates. We consider three out-of-sample forecasting periods each covering one year with about 251 trading days, starting at January 3rd, 2012, and ending at December 31th, 2014.

Statistical forecast evaluation

In order to assess the point forecasting accuracy we follow Ledoit et al. (2003) and use the root-mean-squared-error (RMSE) based on the Frobenius norm comparing the covariance matrix forecast $\widehat{C}_t$ and the ex-post observed value for C_t. This RMSE is given by

$$\mathrm{RMSE} = \frac{1}{T^\star} \sum_t ||C_t - \widehat{C}_t|| = \frac{1}{T^\star} \sum_t \Big[\sum_{i,j} (c_{ijt} - \hat{c}_{ijt})^2 \Big]^{1/2}, \tag{4.36}$$

where $T^\star$ denotes the number of forecasting periods. Since the RMSE is susceptible to outliers in the realized covariance data we also consider the QLIKE loss

$$\mathrm{QLIKE} = \frac{1}{T^\star} \sum_t \log |\widehat{C}_t| + \mathrm{tr}\left(\widehat{C}_t^{\,-1} C_t \right), \tag{4.37}$$

which is known to be robust to noisy (co)variance proxies (see Patton, 2011, and Laurent et al., 2013).

For assessing the significance of differences in the RMSE and QLIKE losses across models we rely on the model confidence set (MCS) approach of Hansen et al. (2011).

The MCS identifies the model or set of models having the best forecasting performance at a given confidence level. The MCS is computed for a confidence level of 90% using a block bootstrap with block length $\lfloor (T^\star)^{1/3} \rfloor$ and 10,000 bootstrap replications.

Table 4.3 reports the RMSE and QLIKE results for the out-of-sample periods 2012, 2013 and 2014 as well as all three periods aggregated together. The RSS approach significantly outperforms its competitors in most cases, in particular for the QLIKE measure. The lowest RMSE losses are obtained under the shrinkage specification. For the QLIKE loss the non-shrinkage version of the RSS model performs best. The Lasso-HAR and the EWMA approach are the strongest competitors under the RMSE and the QLIKE, respectively, but are significantly outperformed by the RSS models. The only exception is the Lasso-HAR in 2013 for the RMSE loss.

Figure 4.6 reports time-series plots of period-wise accumulated Bayesian predictive densities for C_{t+1} computed for the RSS Shrinkage model, the WSS Shrinkage model and the UE approach, which are all estimated by Bayesian inference techniques. Since the evaluation of the $(k \times k)$-dimensional measurement density $f(C_{t+1}|\Sigma_{t+1}^{(i)}; \theta)$ in Eq. (4.32) turns out to be numerically very unstable in high-dimensional applications[12], we focus on the six five-dimensional sub-portfolios discussed in Section 4.5.2. The results show evidence for the Riesz state-space model as opposed to its Wishart and Matrix-F competitors. Only for subsets two and five the Riesz is outperformed by the Matrix-F UE specification.

Global-minimum-variance-portfolio forecasts

In order to provide an economic evaluation of the forecasting performance we use the predicted covariance matrices of the assets to construct optimal investment portfolios (c.f. Bauwens et al., 2016, and Callot et al., 2017). We consider a risk-averse investor who relies on conditional mean-variance analysis to allocate resources across the 30 firms in Table 4.1.

[12] For high-dimensional Riesz measurement densities (as in our application for $k = 30$ assets) the density evaluation suffers from frequent floating-point underflows (see Kastner, 2019, for a discussion of similar computational problems).

For a given covariance matrix forecast $\widehat{C}_t$ computed in period $t-1$ the investor obtains portfolio weights $\hat{w}_t$ by solving the minimization problem

$$\widehat{w}_t = \arg\min_{w_t} w_t' \widehat{C}_t w_t, \tag{4.38}$$

subject to $\sum_{i=1}^{k} w_{it} = 1$ and additional side-conditions which involve constraints on the absolute portfolio weights, leverage, and the expected portfolio return $w_{t+1}' \hat{\mu}_{t+1} = \mu_{target}$, where $w_t = (w_{1t}, \ldots, w_{kt})'$ denotes the vector of period-t portfolio weights to be selected in period $t-1$ and μ_{target} is the target expected return from t to $t+1$. We set μ_{target} to 10% per year and $\hat{\mu}_{t+1}$ is computed by a moving average of the 30 asset returns over the previous 100 trading days. The side-conditions are given in Table 4.4 and discussed below.

We use the following measures for assessing the relative capabilities of the competing models for optimal portfolio allocation:

1. Average out-of-sample portfolio return: $\mu_p = \frac{1}{T-t_0} \sum_{t=t_0+1}^{T} r_t^p$, where $r_t^p = \widehat{w}_t' r_t$ denotes the portfolio return.

2. Accumlated out-of-sample portfolio return: $\mu_p^{acc} = \prod_{t=t_0+1}^{T} (1 + r_t^p) - 1$.

3. Out-of-sample portfolio standard deviation: $\sigma_p = \sqrt{\frac{1}{T-t_0} \sum_{t=t_0+1}^{T} (r_t^p - \mu_p)^2}$.

4. Sharpe ratio: $\text{SR}_p = \frac{\mu_p}{\sigma_p}$.

5. Minimum weight over all assets and time periods: $\min_{t_0+1 \le t \le T} \min_{1 \le i \le k} (\widehat{w}_{it})$.

6. Maximum weight over all assets and time periods: $\max_{t_0+1 \le t \le T} \max_{1 \le i \le k} (\widehat{w}_{it})$.

7. Proportion of leverage: $\frac{1}{k(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{k} 1(\widehat{w}_{it} < 0)$.

8. Average Turnover: $\text{TO} = \frac{1}{k(T-t_0)} \sum_{t=t_0+1}^{T} \sum_{i=1}^{k} |\widehat{w}_{it} - \widehat{w}_{it}^{hold}|$, where $\widehat{w}_{it}^{hold} = \widehat{w}_{it-1} \frac{1+r_{it-1}}{1+r_{t-1}^p}$.

The results for the complete out-of-sample phase from 2012 to 2014 are summarized in Table 4.4. We consider three sets of side-conditions, which are given as captions

of the respective panels in the table. Panel a) considers the global minimum variance portfolio (GMVP) without additional side-conditions. Panel b) involves a no-leverage condition and panels c) and d) impose conditions on the weights and the target return. Panel d) corresponds to the setting of Callot et al. (2017). All four settings show overall good and competitive results for the RSS specifications. The RSS performance also appears robust over the four portfolio settings and involves comparably low portfolio standard deviations. We however note that the RW and the UE approach are strong competitors w.r.t portfolio performance as measured by the average return and the Sharpe ratio. The lowest turnover is always generated by the EWMA model.

Figure 4.7 shows time-series plots of randomly selected portfolio weight forecasts obtained under the RSS, WSS and UE models. The RSS forecasts appear smoothest, inducing lower turnover, leverage and less extreme positive and negative weights compared to its state-space competitors (compare Table 4.4). This finding can be interpreted as a direct consequence of the reduced variability of the RSS state process compared to its Wishart state-space competitors, where the Wishart attenuates measurement errors, resulting in spurious state variation, which in turn induces extreme portfolio weight forecasts.

4.6. Conclusion

We propose a new matrix-variate Riesz state-space (RSS) model for time series of k-dimensional positive-definite realized covariance matrices of asset returns. The RSS model is based on the Riesz measurement density, which was introduced by Hassairi and Lajmi (2001) and can be interpreted as a generalization of the Wishart to missing data structures. We analyze the stochastic properties of the Riesz and provide a comparison to the nested Wishart distribution as a measurement density for realized covariance matrices. Importantly, we find that the Wishart attenuates the dispersion of measurement errors in high-dimensional applications. This attenuation is alleviated by the Riesz, which allows for more flexible covariance structures via $k-1$ additional

d.o.f. to capture the underlying liquidity pattern.

We combine the Riesz measurement density with a Cholesky-based state transition in order to obtain a tractable state-space model which enables straightforward Bayesian MCMC estimation completely parallelized over the assets. In order to cope with potential overfitting problems in high-dimensional applications we introduce sparsity on the state dynamics via implementing the hierarchical Normal-Gamma shrinkage prior approach of Bitto and Frühwirth-Schnatter (2019).

We provide an empirical application of Wishart- and Riesz state-space models to 30 assets selected from the S&P 500. Our results show that the Wishart implies a considerable reduction of the measurement error variance for the less liquid assets, inducing spurious state variation accompanied by a significant reduction of the short-term state persistence relative to the Riesz. Compared to relevant Wishart state-space competitors the Riesz specification provides a significant improvement in in-sample fit.

The in-sample findings are confirmed by an out-of-sample forecasting experiment including both statistical and economic evaluation criteria. The RSS model shows good forecasting performance and we can conclude that the RSS approach is a valuable tool for modeling and forecasting time-series of high-dimensional realized covariance matrices.

4.7. Tables and Figures

Table 4.1.: List of the stocks included in the data set.

#	Symbol	Company	Sector	Avg. trades	Med. trades	Subset
1	VAR	Varian Medical Systems	H	360.8	374	2
2	DOV	Dover Corp.	I	371.0	381	5
3	LXK	Lexmark Int., Inc.	T	374.5	382	3
4	CCE	Coca-Cola European Partners	S	372.4	385	4
5	PX	Pelangio Exploration Inc.	M	378.8	386	2
6	CB	Chubb Ltd.	F	377.0	386	3
7	APOL	Apollo Group, Inc.	D	381.7	387	1
8	DO	Diamond Offshore Drill., Inc.	E	375.5	388	4
9	CAG	Conagra Brands, Inc.	S	383.6	388	6
10	VRSN	VeriSign, Inc.	T	383.9	388	6
11	A	Agilent Technologies, Inc.	H	385.0	389	6
12	LMT	Lockheed Martin Corp.	I	385.4	389	5
13	ETFC	E*TRADE Fin. Corp.	F	381.1	389	6
14	CIEN	Ciena Corp.	T	384.8	389	4
15	EMR	Emerson Electric Company	I	385.1	390	5
16	NKE	Nike Inc.	D	382.3	390	2
17	TJX	TJX Companies Inc.	D	382.1	390	1
18	BAX	Baxter International	H	385.0	390	3
19	ALL	Allstate Corp.	F	385.5	390	5
20	SO	Southern Company	U	385.0	390	4
21	C	Citigroup Inc.	F	389.4	391	1
22	DD	DuPont de Nemours, Inc.	M	388.3	391	1
23	MSFT	Microsoft Corp.	T	389.5	391	3
24	SBUX	Starbucks Corp.	D	388.4	391	4
25	HD	The Home Depot, Inc.	D	389.1	391	3
26	KO	The Coca-Cola Company	S	388.7	391	2
27	JNJ	Johnson & Johnson	H	389.1	391	5
28	USB	U.S. Bancorp	F	387.4	391	6
29	MS	Morgan Stanley	F	388.7	391	2
30	EMC	EMC Corporation	T	389.1	391	1

Note: The stocks are sorted by increasing liquidity based on the median number of trades per day (Med. trades). Avg. trades denotes the average number of trades per day. Subset denotes the respective 5-dimensional subset the particular stock belongs to. The sector labels are: (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (M) Materials; (E) Energy; (U) Utilities; (F) Financials; (T) Information Technologies.

Table 4.2.: Deviance Information Criteria (DIC) results.

					DIC $\times 10^{-4}$				
Assets					RSS	RSS Shr.	WSS	WSS Shr.	UE
apol	tjx	c	dd	emc	4.6312	4.6148	4.6471	4.6461	30.8564
var	px	nke	ko	ms	0.9327	0.9327	1.0808	1.1317	15.6550
lxk	cb	bax	msft	hd	1.0049	0.9797	1.0250	1.0372	16.3226
cce	do	cien	so	sbux	5.0921	5.0848	5.0983	5.0913	32.2674
dov	lmt	emr	all	jnj	−1.6123	−1.6133	−1.6017	−1.6059	5.0853
cag	vrsn	a	etfc	usb	5.2281	5.2342	5.2755	5.2643	33.2150

Note: This table reports the DIC results for the RSS, WSS and UE models for the six subsets of the 30D data. The likelihood for the DICs in Eq. (4.29) is computed using a Rao-Blackwellized bootstrap particle filter with 25,000 particles. The assets belonging to the corresponding subsets are listed in Table 4.1.

Table 4.3.: Evaluation of point-forecast accuracy.

	Period							
Model	2012	2013	2014	Full	2012	2013	2014	Full
	RMSE				QLIKE			
RSS	12.40	9.04	9.97	10.46	9.89	9.48	10.45	9.94
RSS Shr.	12.40	9.02	9.96	10.45	9.91	9.48	10.47	9.96
WSS	13.28	9.65	10.03	10.98	10.37	9.85	10.85	10.36
WSS Shr.	13.15	9.56	9.98	10.89	10.29	9.79	10.76	10.28
UE	13.31	9.75	10.29	11.11	11.02	10.67	11.78	11.16
Lasso-VAR(1)	13.04	9.19	10.30	10.84	11.50	10.86	11.92	11.43
Lasso-VAR(20)	12.88	9.14	10.28	10.76	11.19	10.44	11.54	11.06
Lasso-VAR(HAR)	12.61	9.06	10.17	10.61	11.08	10.37	11.52	10.99
EWMA	14.18	9.63	10.86	11.55	11.31	9.80	11.61	10.90
RW	16.61	11.95	11.95	13.50	17.47	16.77	18.49	17.58

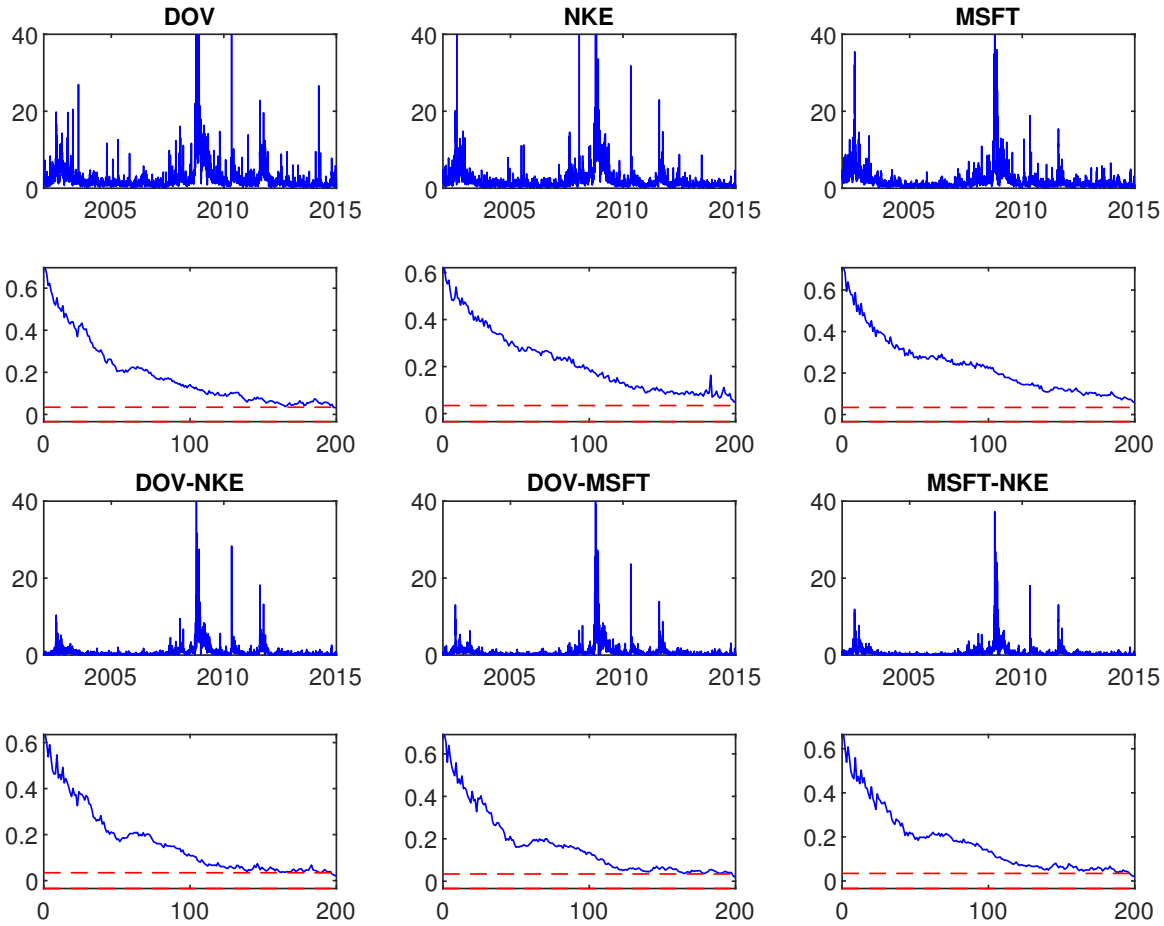
Note: This table reports the RMSE and QLIKE losses as given in Eqs. (4.36) and (4.37). Grey-shaded cells indicate that the model is in the 90% model confidence set.

Table 4.4.: GMVP forecasting results.

Model	μ_p	μ_p^{acc}	σ_p	SR_p	min$w$	max$w$	Leverage	TO
	Unrestricted							
RSS	0.0504	0.4507	0.4503	0.1120	-0.1151	0.4084	0.2809	0.0125
RSS Shr.	0.0514	0.4612	0.4497	0.1143	-0.1152	0.4254	0.2849	0.0138
WSS	0.0508	0.4538	0.4616	0.1100	-0.2027	0.5258	0.3282	0.0270
WSS Shr.	0.0517	0.4640	0.4616	0.1120	-0.1928	0.5412	0.3270	0.0262
UE	0.0566	0.5194	0.4641	0.1221	-0.1607	0.5187	0.3371	0.0165
Lasso-VAR(1)	0.0423	0.3635	0.4634	0.0912	-0.0603	0.4775	0.3631	0.0187
Lasso-VAR(20)	0.0433	0.3741	0.4598	0.0941	-0.0645	0.3982	0.3549	0.0126
Lasso-VAR(HAR)	0.0457	0.4003	0.4529	0.1010	-0.0752	0.3821	0.3441	0.0118
EWMA	0.0412	0.3532	0.4447	0.0926	-0.0980	0.3225	0.2885	0.0023
RW	0.0565	0.5152	0.5056	0.1117	-0.1908	0.8176	0.3940	0.0579
	$w_{it+1} \geq 0 \; \forall i$							
RSS	0.0455	0.3974	0.4608	0.0988	0.0000	0.4186	—	0.0098
RSS Shr.	0.0464	0.4073	0.4584	0.1013	0.0000	0.4333	—	0.0110
WSS	0.0465	0.4075	0.4702	0.0989	0.0000	0.5185	—	0.0197
WSS Shr.	0.0476	0.4187	0.4690	0.1014	0.0000	0.5350	—	0.0192
UE	0.0494	0.4385	0.4731	0.1045	0.0000	0.5008	—	0.0113
Lasso-VAR(1)	0.0369	0.3090	0.4739	0.0778	0.0000	0.4888	—	0.0137
Lasso-VAR(20)	0.0381	0.3209	0.4699	0.0810	0.0000	0.4134	—	0.0093
Lasso-VAR(HAR)	0.0389	0.3293	0.4646	0.0837	0.0000	0.3997	—	0.0089
EWMA	0.0402	0.3429	0.4585	0.0877	0.0000	0.3235	—	0.0016
RW	0.0460	0.4006	0.5052	0.0911	0.0000	0.8547	—	0.0380
	$0 \leq w_{it+1} \leq 0.2 \; \forall i; \; w_{t+1}'\hat{\mu}_{t+1} = \mu_{target}$							
RSS	0.0474	0.4160	0.4838	0.0979	0.0000	0.2000	—	0.0091
RSS Shr.	0.0471	0.4136	0.4823	0.0977	0.0000	0.2000	—	0.0100
WSS	0.0466	0.4077	0.4951	0.0942	0.0000	0.2000	—	0.0176
WSS Shr.	0.0470	0.4114	0.4947	0.0950	0.0000	0.2000	—	0.0171
UE	0.0485	0.4278	0.4975	0.0976	0.0000	0.2000	—	0.0102
Lasso-VAR(1)	0.0406	0.3445	0.4972	0.0816	0.0000	0.2000	—	0.0121
Lasso-VAR(20)	0.0407	0.3457	0.4949	0.0822	0.0000	0.2000	—	0.0085
Lasso-VAR(HAR)	0.0418	0.3575	0.4899	0.0853	0.0000	0.2000	—	0.0081
EWMA	0.0418	0.3574	0.4820	0.0866	0.0000	0.2000	—	0.0029
RW	0.0502	0.4444	0.5165	0.0971	0.0000	0.2000	—	0.0323
	$\lvert w_{it+1}\rvert \leq 0.2 \; \forall i; \; \sum_{i=1}^{k} \lvert w_{it+1}\rvert 1(w_{it} < 0) < 0.3; \; w_{t+1}'\hat{\mu}_{t+1} = \mu_{target}$							
RSS	0.0518	0.4649	0.4629	0.1119	-0.1068	0.2000	0.2874	0.0121
RSS Shr.	0.0522	0.4695	0.4635	0.1126	-0.0973	0.2000	0.2927	0.0132
WSS	0.0530	0.4779	0.4752	0.1116	-0.1500	0.2000	0.3258	0.0244
WSS Shr.	0.0534	0.4827	0.4751	0.1125	-0.1440	0.2000	0.3241	0.0237
UE	0.0583	0.5381	0.4734	0.1232	-0.1474	0.2000	0.3275	0.0154
Lasso-VAR(1)	0.0462	0.4039	0.4742	0.0974	-0.0848	0.2000	0.3522	0.0174
Lasso-VAR(20)	0.0473	0.4159	0.4725	0.1001	-0.0887	0.2000	0.3478	0.0121
Lasso-VAR(HAR)	0.0483	0.4263	0.4667	0.1034	-0.0838	0.2000	0.3386	0.0117
EWMA	0.0452	0.3943	0.4567	0.0989	-0.1036	0.2000	0.2935	0.0037
RW	0.0520	0.4657	0.5018	0.1037	-0.1744	0.2000	0.3619	0.0458

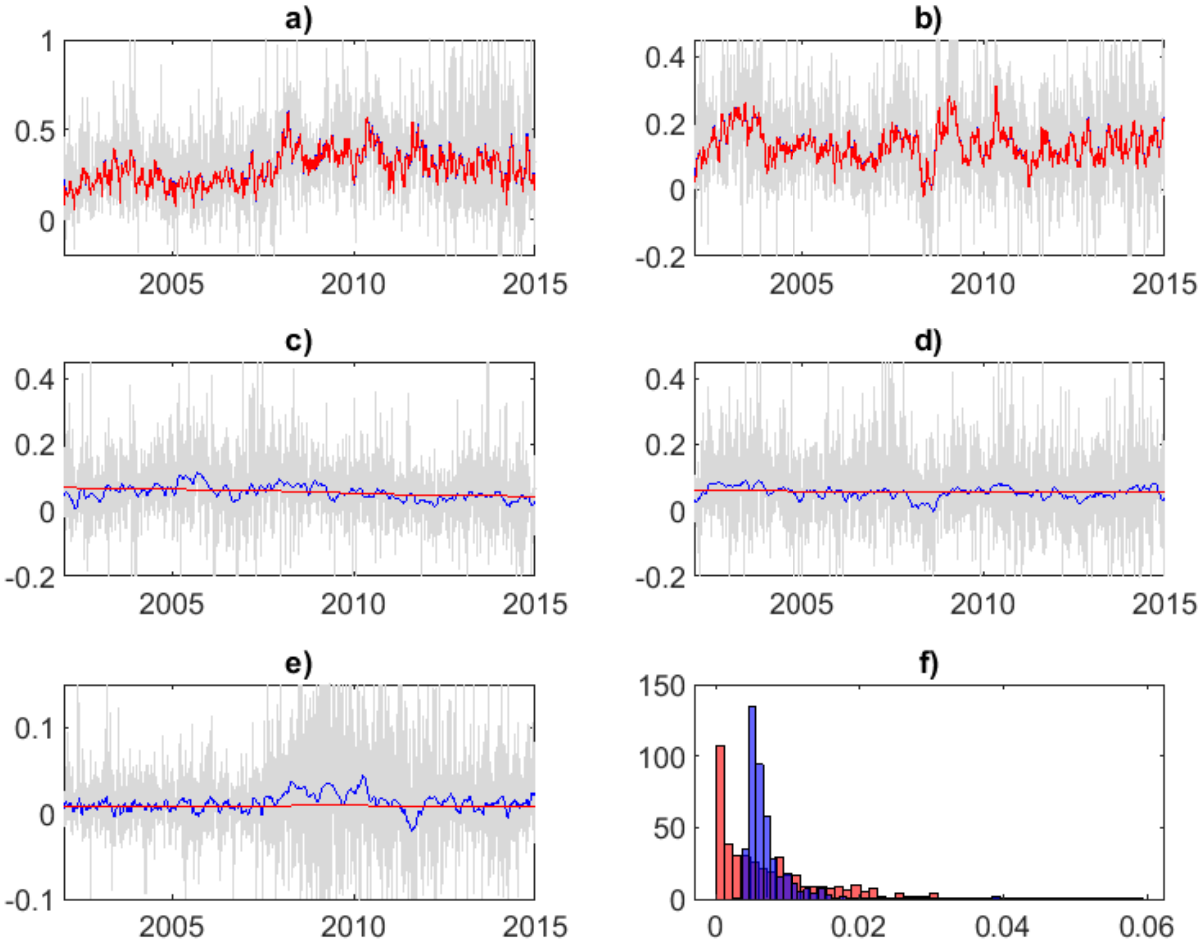
Note: This table reports GMVP summary statistics for different restrictions. $\hat{\mu}_{t+1}$ is computed as a moving average with window length $h = 100$. μ_{target} is set to 10% per year.

Figure 4.1.: Time series plots of selected realized (co)variances.



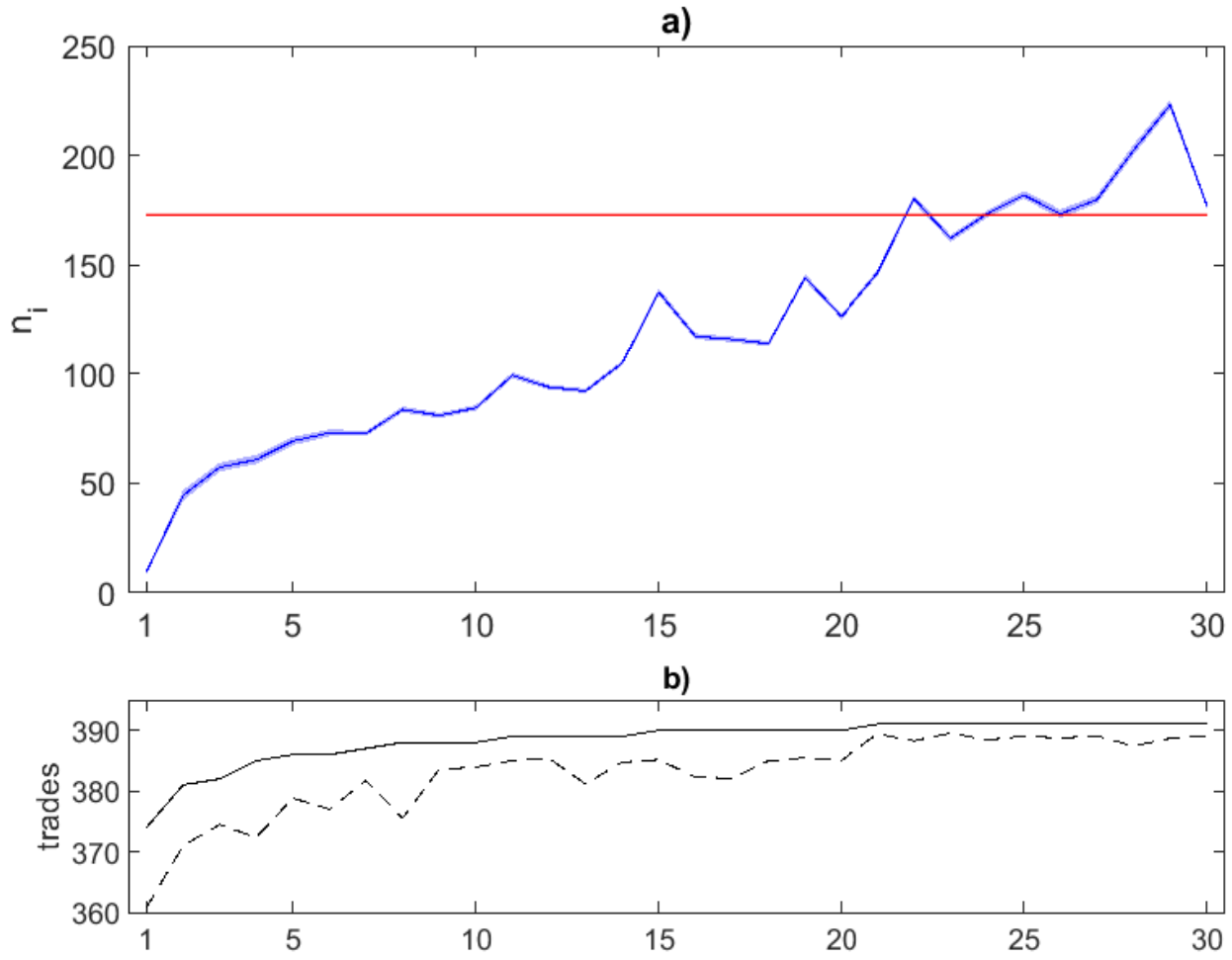
Time-series plots and according sample ACFs of the realized (co)variances of three stocks selected randomly from the 30 stocks in Table 4.1. The dashed lines indicate 95% Bartlett confidence bounds under the null of no autocorrelation.

Figure 4.2.: Time series plots of β estimates obtained under the Riesz distribution.



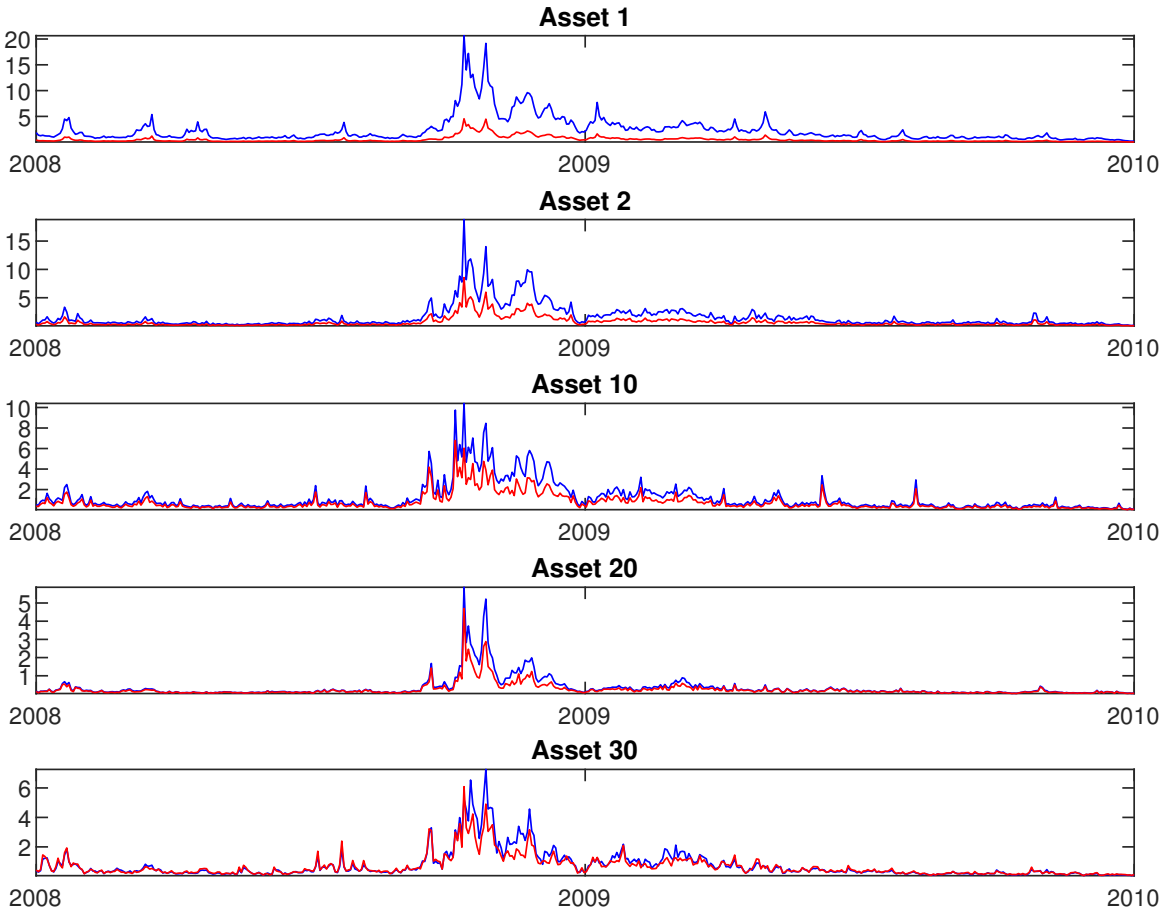
Panels a) – e): time-series plots of β estimates obtained under the Riesz distribution. Panel f): histograms of estimated conditional RW standard deviations. Red: smoothed estimates with shrinkage priors; blue: smoothed estimates without shrinkage priors. Gray: realized betas.

Figure 4.3.: Estimation results for the d.o.f parameters of the RSS and WSS model.



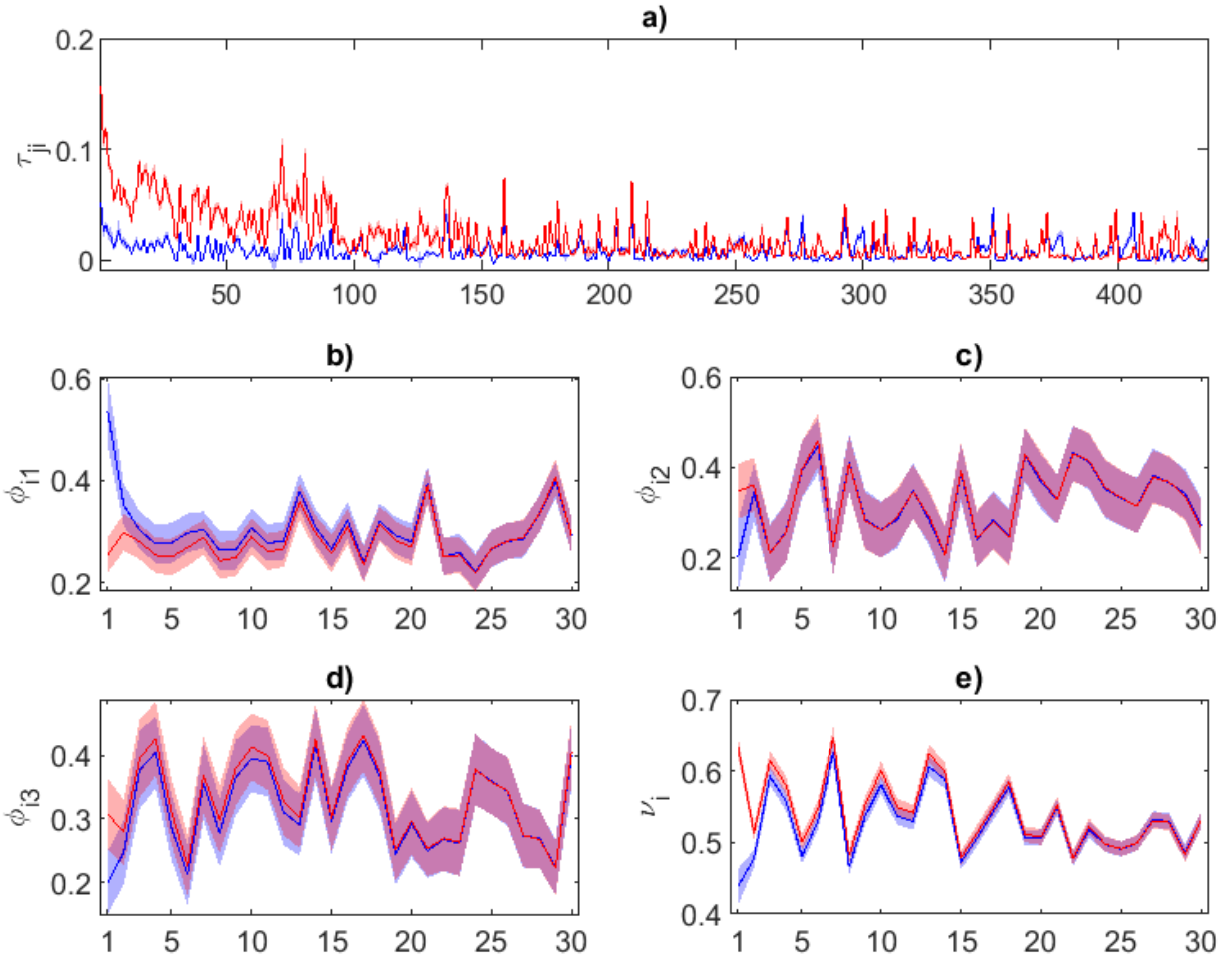
Panel a): Estimated d.o.f. for the RSS Shrinkage model (dark blue) and the corresponding WSS Shrinkage model (dark red). The light blue and light red shaded areas mark 95% posterior high density regions for the RSS Shrinkage model and the WSS Shrinkage model, respectively. Panel b): Liquidity measured according to the median (solid) and average number of trades per day (dashed), respectively.

Figure 4.4.: Time-series of measurement error standard deviations.



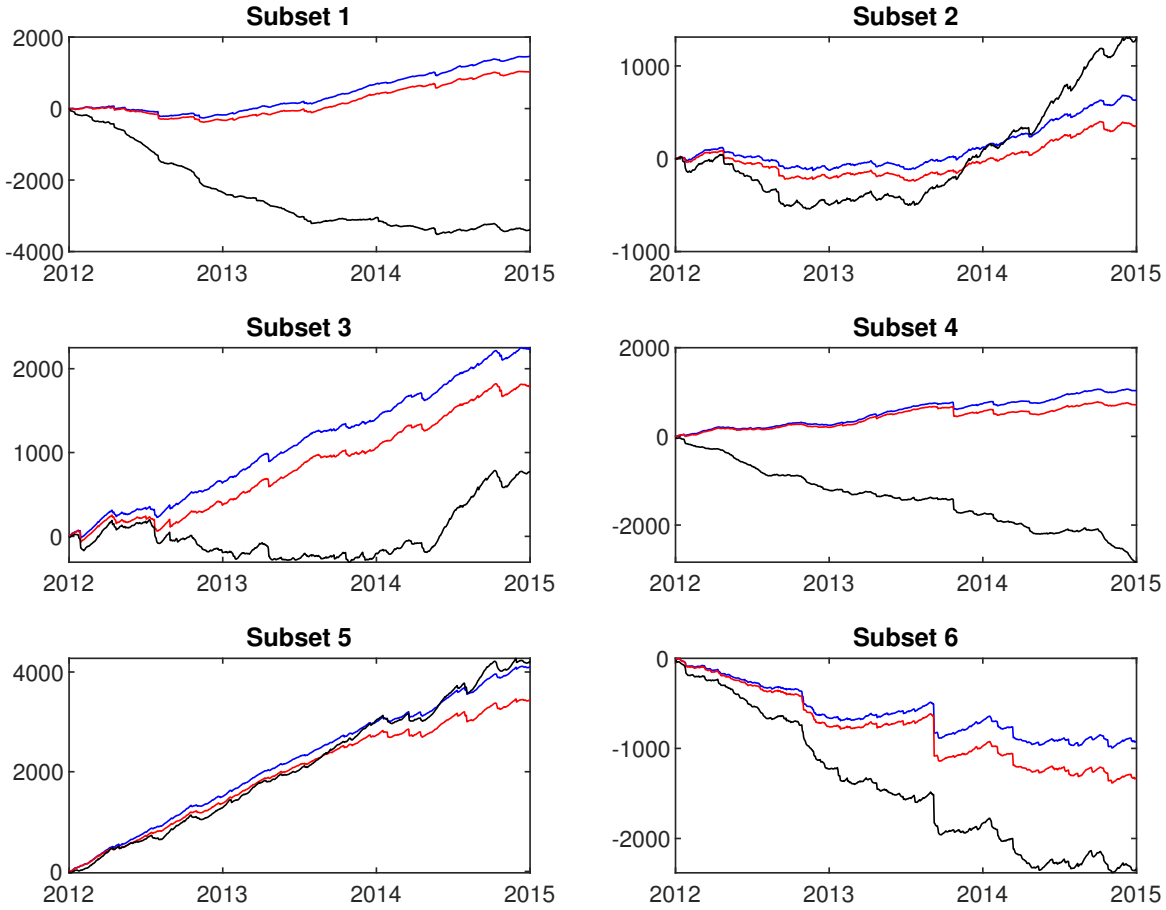
Time-series of measurement error standard deviations for the 1st, 2nd, 10th, 20th and 30th asset for the time period from January 2nd, 2008, to December 30th, 2009. The standard deviations are computed at estimated d.o.f. and smoothed estimates of the integrated variances according to the RSS Shrinkage model (blue) and the WSS shrinkage model (red) for the 30-dimensional data set detailed in Section 4.5.1.

Figure 4.5.: Estimation results for the RSS and WSS model.



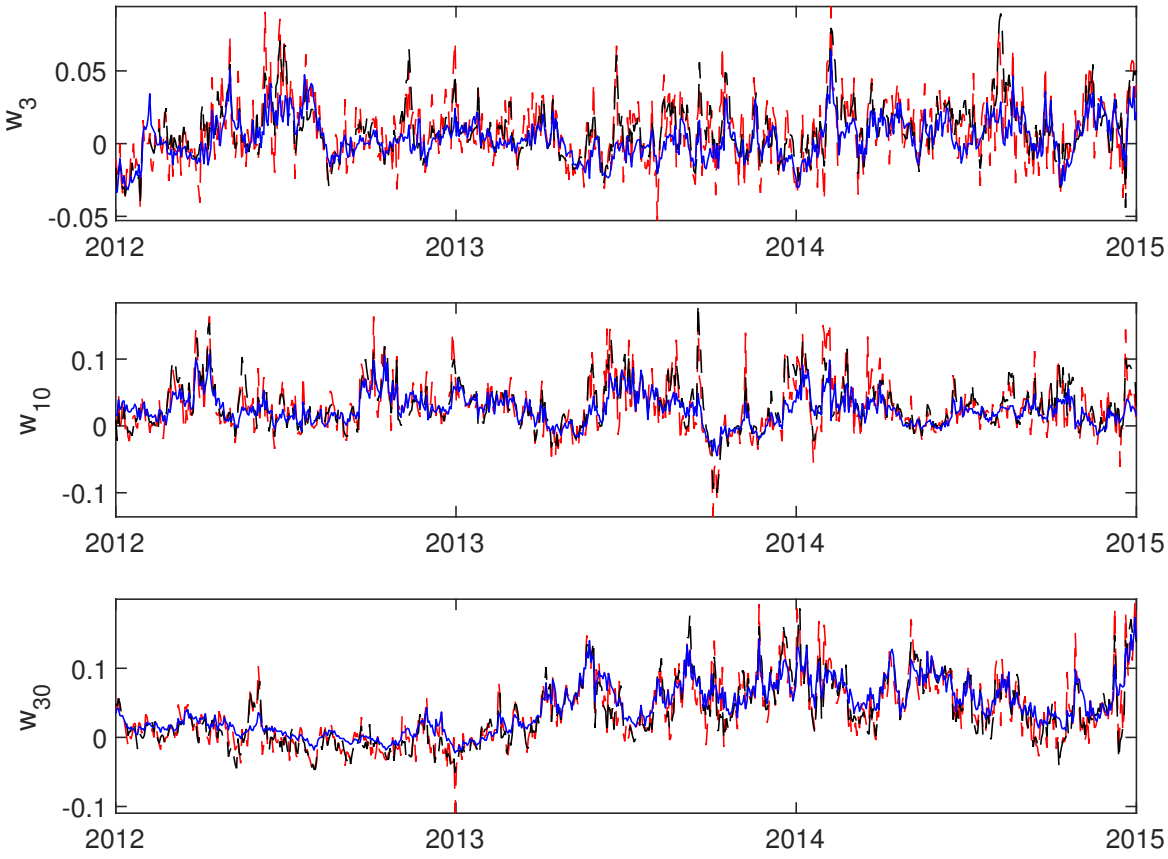
Estimation results for the state parameters of the RSS Shrinkage model (dark blue) and the corresponding WSS Shrinkage model (dark red). Panel a): estimated conditional standard deviations for the RW processes for the β_{ji} coefficients. Panel b) – e): estimated HAR parameters for the $\log \sigma_{i.i-1,t}$ processes. The light blue and light red shaded areas mark 95% posterior high density regions for the RSS Shrinkage model and the WSS Shrinkage model, respectively.

Figure 4.6.: Time-series plots of accumulated Bayesian predictive densities.



Time-series plots of period-wise accumulated Bayesian predictive densities for C_{t+1} in different five-dimensional subsets of the full data set. RSS Shrinkage: blue; WSS Shrinkage: red; Matrix-F UE: black.

Figure 4.7.: Time-series plots of GMVP weights.



Time-series plots of GMVP weights for DOV, NKE and MSFT stock under Callot et al. (2017) restrictions. RSS Shrinkage: solid blue; WSS Shrinkage: dashed red; Matrix-F UE: dashed black.

Appendix A.
Appendix for Chapter 2

A.1. Details on the MCMC Algorithm

A.1.1. PGAS algorithm for the sampling of $x^f_{k,1:T}$ and $x^e_{i,1:T}$

In steps 1.) and 3.) of the proposed MCMC sampling scheme in Section 2.3.1, we sample the latent state series $\{x^f_{k,1:T}\}$ and $\{x^e_{i,1:T}\}$ full conditional on the static model parameters and latent states $\{\tilde{\ell}_{k,1:T}\}$ and $\{\beta_{i,1:T}\}$ using particle Gibbs with ancestor sampling (PG-AS, Andrieu et al., 2010; Lindsten et al., 2014). Note, in addition to conditioning on the static parameters, the particle Gibbs algorithm also uses a draw of $x^f_{k,1:T}$ ($x^e_{i,1:T}$) from the previous MCMC iteration as reference trajectory.

In the following, the PG-AS algorithm is stated for sampling of the ith idiosyncratic component with AR(1) state dynamics. The algorithm is straightforwardly altered to sample the factor components or to incorporate latent HAR dynamics. The PG-AS algorithm proceeds in the following steps to sample $x^e_{i,1:T}$ from its posterior in Eq. (2.19):

1.) *For $t = 1$:* Initialize $M-1$ particles $x^{e(1)}_{i1}, \dots, x^{e(M-1)}_{i1}$ from $\mathcal{N}(m_{xi}, v^2_{xi})$, where m_{xi} and v^2_{xi} are the stationary mean and variance of the latent AR process, respectively, and deterministically set $x^{e(M)}_{i1} = x^e_{i1}$. Then calculate the weights

$$w^{(j)}_{i1} = \exp\left\{-\frac{n}{2}\left[x^{e(j)}_{i1} + \left(\beta'_{i1}C^f_1\beta_{i1} - 2\beta'_{i1}c^{rf}_{i1} + c^r_{i1}\right)\exp(-x^{e(j)}_{i1})\right]\right\}, \qquad \text{(A.1)}$$

and compute the normalized weights $\widetilde{w}^{(j)}_{i1} = w^{(j)}_{i1} / \sum_{l=1}^M w^{(l)}_{i1}$, for $j = 1, \dots, M$.

2.) *For $t = 2, \dots, T$:* Re-sample the state trajectories by drawing $\widetilde{x}^{e(1)}_{i,1:t-1}, \dots, \widetilde{x}^{e(M-1)}_{i,1:t-1}$ from the set $\{x^{e(j)}_{i,1:t-1}\}_{j=1}^M$ according to the probabilities $\{\widetilde{w}^{(j)}_{it-1}\}_{j=1}^M$, and draw an

index $J \in \{1, \ldots, M\}$ with probability

$$\Pr(J = j) = \frac{w_{it-1}^{(j)} f_{\mathcal{N}}(x_{it}^{e} \mid x_{it-1}^{e(j)})}{\sum_{l=1}^{M} w_{it-1}^{(l)} f_{\mathcal{N}}(x_{it}^{e} \mid x_{it-1}^{e(l)})}, \tag{A.2}$$

for $j = 1, \ldots, M$. Here $f_{\mathcal{N}}(x_{it}^{e} \mid x_{it-1}^{e(j)})$ denotes the Gaussian transition density of the latent states evaluated at the reference trajectory. Set $\widetilde{x}_{i1:t-1}^{e(M)} = x_{i1:t-1}^{e(J)}$. Then, sample $M-1$ draws $x_{it}^{e(1)}, \ldots, x_{it}^{e(M-1)}$ from the corresponding transitions $\mathcal{N}(\gamma_i^e + \phi_i^e(\widetilde{x}_{it-1}^{e(j)} - \gamma_i^e), [\nu_i^e]^2)$, $j = 1, \ldots, M-1$, and set $x_{it}^{e(M)} = x_{it}^{e}$ and $x_{i,1:t}^{e(j)} \equiv (\widetilde{x}_{i,1:t-1}^{e(j)}, x_{it}^{e(j)})$, $j = 1, \ldots, M$. Now, calculate the weights

$$w_{it}^{(j)} = \exp\left\{-\frac{n}{2}\left[x_{it}^{e(j)} + \left(\beta_{it}' C_t^f \beta_{it} - 2\beta_{it}' c_{it}^{rf} + c_{it}^{r}\right)\exp(-x_{it}^{e(j)})\right]\right\}, \tag{A.3}$$

and compute the normalized weights $\widetilde{w}_{it}^{(j)} = w_{it}^{(j)} / \sum_{l=1}^{M} w_{it}^{(l)}$, for $j = 1, \ldots, M$.

3.) *For $t = T$:* Sample an index $J^\star \in \{1, \ldots, M\}$ with probability $\Pr(J^\star = j) = \widetilde{w}_{iT}^{(j)}$. The trajectory $x_{i,1:T}^{e(J^\star)} \equiv (\widetilde{x}_{i,1:T-1}^{e(J^\star)}, x_{iT}^{e(J^\star)})$ is now used as approximate draw from the posterior distribution.

The PG-AS algorithm for the sampling of the factor components $x_{k,1:T}^{f}$, $k = 1, \ldots, q$, is obtained by setting the weights $w_{it}^{(j)}$ in Eqs. (A.1) and (A.3) to

$$w_{kt}^{(j)} = \exp\left\{-\frac{n}{2}\left[x_{kt}^{f(j)} + \left(\ell_{kt}' C_t^f \ell_{kt}\right)\exp(-x_{kt}^{f(j)})\right]\right\}, \tag{A.4}$$

for $t = 1, \ldots, T$.

A.1.2. FFBS algorithm for the sampling of $\tilde{\ell}_{k,1:T}$ and $\beta_{i,1:T}$

In steps 2.) and 4.) of the proposed MCMC sampling scheme, we sample the latent state series $\{\tilde{\ell}_{k,1:T}\}$ and $\{\beta_{i,1:T}\}$ full conditional on the static model parameters and latent states $\{x_{k,1:T}^{f}\}$ and $\{x_{i,1:T}^{e}\}$ via a Forward Filtering Backward Sampling algorithm (FFBS, de Jong and Shephard, 1995). In the following, the FFBS algorithm is stated for an arbitrary d-variate vector state-space model with latent state vector

$x_{1:T}$, which can straightforwardly be adjusted to match the corresponding models for $\{\tilde{\ell}_{k,1:T}\}$ and $\{\beta_{i,1:T}\}$.

The d-variate vector state-space model is given by

$$y_t \mid x_t \sim \mathcal{N}(x_t, F_t), \qquad x_t \mid x_{t-1} \sim \mathcal{N}(\Gamma + \Phi x_{t-1}, V), \tag{A.5}$$

with $x_0 \sim \mathcal{N}(\tilde{\Gamma}, \tilde{V})$, where $\tilde{\Gamma} = (I_d - \Phi)^{-1}\Gamma$ and $\text{vec}(\tilde{V}) = (I_{d^2} - \Phi \otimes \Phi)^{-1}\text{vec}(V)$.

Given the initial conditions $M_0 = \tilde{\Gamma}$ and $P_0 = \tilde{V}$, the forward filter cycles in expectation through the predictive and filtering distributions

$$x_t \mid y_{1:t-1} \sim \mathcal{N}(M_t^\star, P_t^\star), \qquad \text{and} \quad x_t \mid y_{1:t} \sim \mathcal{N}(M_t, P_t), \tag{A.6}$$

where

$$M_t^\star = \Gamma + \Phi M_{t-1}, \qquad P_t^\star = \Phi P_{t-1}\Phi' + V \tag{A.7}$$

$$M_t = M_t^\star + K_t(y_t - M_t^\star), \qquad P_t = P_t^\star - K_t(P_t^\star + F_t)K_t' \tag{A.8}$$

with $K_t = P_t^\star(P_t^\star + F_t)^{-1}$, for $t = 1, \ldots, T$.

After the forward filtering is finished, the sequence $x_{1:T}$ of latent states is drawn recursively from their posterior

$$\pi(x_{1:T} \mid y_{1:T}, \theta) = f(x_T \mid y_{1:T}, \theta) \prod_{t=1}^{T-1} f(x_t \mid y_{1:t}, x_{t+1}, \theta), \tag{A.9}$$

by first sampling from the full conditional distribution

$$x_T \mid y_{1:T}, \theta \sim \mathcal{N}(M_T, P_T), \tag{A.10}$$

and then recursively backward sampling from

$$x_t \mid y_{1:t}, x_{t+1}, \theta \sim \mathcal{N}(M_t + \tilde{K}_t(x_{t+1} - \Gamma - \Phi M_t), P_t - \tilde{K}_t(V + \Phi P_t \Phi')\tilde{K}_t'), \tag{A.11}$$

with $\tilde{K}_t = P_t\Phi(V + \Phi P_t\Phi')^{-1}$, for $t = T-1, \ldots, 1$, where M_t and P_t are found from the forward filtering algorithm in (A.7) and (A.8); see also Kim and Nelson (1999).

The FFBS algorithms for $\{\tilde{\ell}_{kt}\}_{k=2}^{q}$ and $\{\beta_{it}\}_{i=1}^{p}$ are readily set up by setting

$$y_t \equiv l_{kt} = (\tilde{C}_{kt}^f)^{-1}\tilde{c}_{kt}^f, \qquad x_t \equiv \tilde{\ell}_{kt}, \qquad F_t \equiv \frac{e^{x_{kt}^f}}{n}(\tilde{C}_{kt}^f)^{-1},$$
$$\Gamma \equiv \gamma_k^\ell, \qquad \Phi \equiv \Phi_k^\ell, \qquad V \equiv \Sigma_k^\ell,$$

and

$$y_t \equiv b_{it} = (C_t^f)^{-1}c_{it}^{rf}, \qquad x_t \equiv \beta_{it}, \qquad F_t \equiv \frac{e^{x_{it}^e}}{n}(C_t^f)^{-1},$$
$$\Gamma \equiv \gamma_i^\beta, \qquad \Phi \equiv \Phi_i^\beta, \qquad V \equiv \Sigma_i^\beta,$$

respectively. The vectors and matrices $\tilde{C}_{kt}^f, \tilde{c}_{kt}^f, \gamma_k^\ell, \Phi_k^\ell, \Sigma_k^\ell$ and $\gamma_i^\beta, \Phi_i^\beta, \Sigma_i^\beta$ are defined in Section 2.3.1.

A.1.3. Sampling of HAR parameters

The WFSS-HAR model assumes HAR dynamics for the individual time-series $\{x_{k,1:T}^f\}$ and $\{x_{i,1:T}^e\}$. Based on an arbitrary time-series $x_{i,1:T}$, this section describes how to sample the corresponding static parameters.

Conditional on $x_{i,1:T}$, joint updating for the HAR parameters γ_i, ϕ_{i1}, ϕ_{i2}, ϕ_{i3}, ν_i^2 is done by sampling from

$$\nu_i^2 \mid x_{i,1:T} \sim \mathcal{IG}\left(\frac{\nu_1}{2}, \frac{s_1}{2}\right), \tag{A.12}$$

$$(\gamma_i^*, \phi_{i1}, \phi_{i2}, \phi_{i3})' \mid x_{i,1:T}, \nu_i^2 \sim \mathcal{N}_4(M_1^{-1}b_1, \nu_i^2 M_1^{-1}), \tag{A.13}$$

with $\gamma_i^* = \gamma_i(1 - \phi_{i1} - \phi_{i2} - \phi_{i3})$, and

$$\nu_1 = \nu_0 + T, \tag{A.14}$$

$$s_1 = s_0 + y'(y - Xb) + (b_0 - b)'(M_0^{-1} + (X'X)^{-1})^{-1}(b_0 - b), \tag{A.15}$$

$$M_1 = M_0 + X'X, \tag{A.16}$$

$$b_1 = M_0 b_0 + X'Xb, \tag{A.17}$$

where ν_0, s_0, b_0 and M_0 are the prior hyperparameters, y is a $(T-22)\times 1$ vector with entries $x_{i,23:T}$, X is a $(T-22)\times 4$ matrix obtained by stacking the 1×4 vectors $X_t = (1, \bar{x}_{i[t-1:t-1]}, \bar{x}_{i[t-1:t-5]}, \bar{x}_{i[t-1:t-22]})$, for $t = 23, \ldots, T$, and $b = (X'X)^{-1}X'y$. A sample for γ_i is obtained by rescaling γ_i^* with $(1 - \phi_{i1} - \phi_{i2} - \phi_{i3})^{-1}$.

Sampling from the Normal-inverted-Gamma posterior in Eqs. (A.12) and (A.13) does not ensure stationarity of the underlying HAR process and invertibility of $(1 - \phi_{i1} - \phi_{i2} - \phi_{i3})$. However, this can be achieved via a Metropolis-step. Therefore, a proposal for $(\gamma_i^*, \phi_{i1}, \phi_{i2}, \phi_{i3})$ is drawn from (A.13). The proposal is accepted as new draw with probability one if the stationarity conditions for the HAR process and $(1 - \phi_{i1} - \phi_{i2} - \phi_{i3}) \neq 0$ are fulfilled, and is rejected otherwise.

A.1.4. Sampling of AR(1) parameters

The WFSS model assumes AR(1) dynamics for the individual time-series $\{\ell_{kj,1:T}\}$, and – in few model specifications – $\{\beta_{ik,1:T}\}$, $\{x^f_{k,1:T}\}$ and $\{x^e_{i,1:T}\}$. Based on an arbitrary time-series $x_{i,1:T}$, this section describes how to sample the corresponding static parameters.

Similar to the updating of the HAR parameters in Section A.1.3, joint updating for the AR(1) parameters γ_i, ϕ_i, ν_i^2 is done conditional on $x_{i,1:T}$, by sampling from

$$\nu_i^2 \mid x_{i,1:T} \sim \mathcal{IG}\left(\frac{\nu_1}{2}, \frac{s_1}{2}\right), \tag{A.18}$$

$$(\gamma_i^*, \phi_i)' \mid x_{i,1:T}, \nu_i^2 \sim \mathcal{N}_2(M_1^{-1}b_1, \nu_i^2 M_1^{-1}), \tag{A.19}$$

with $\gamma_i^* = \gamma_i(1-\phi_i)$. The posterior parameters ν_1, s_1, b_1 and M_1 are defined in Eqs. (A.14)–(A.17). Here, y is a $(T-1)\times 1$ vector with entries $x_{i,2:T}$, X is a $(T-1)\times 2$ matrix obtained by stacking the 1×2 vectors $X_t = (1, x_{i,t-1})$, for $t = 2,\ldots,T$, and $b = (X'X)^{-1}X'y$. A sample for γ_i is obtained by rescaling γ_i^* with $(1-\phi_i)^{-1}$.

Sampling from the Normal-inverted-Gamma posterior in Eqs. (A.18) and (A.19) does not ensure stationarity of the underlying AR(1) process and invertibility of $(1-\phi_i)$. However, this can be achieved via a Metropolis-step. Therefore, a proposal for (γ_i^*, ϕ_i) is drawn from (A.19). The proposal is accepted as new draw with probability one if the stationarity conditions for the AR(1) process and $(1-\phi_i) \neq 0$ are fulfilled, and is rejected otherwise.

A.1.5. Sampling of d.o.f. parameter n

In the last step we simulate the Wishart degrees of freedom parameter n, which is allowed to take values on a discrete grid, i.e., $n \in \{n_{(1)}, n_{(2)}, \ldots, n_{(N)}\}$, with $m < n_{(1)} < n_{(2)} < \cdots < n_{(N)}$. Given the discrete uniform prior, $\mathrm{pr}(n)$, which assigns the prior probability $\Pr(n = n_{(i)}) = \frac{1}{N}$, $i = 1,\ldots,N$, the posterior probability for each $n_{(i)}$ is given by

$$\begin{aligned}\Pr(n = n_{(i)} | \{x^e_{i,1:T}\}, \{\beta_{i,1:T}\}, \{x^f_{k,1:T}\}, \{\ell_{k,1:T}\}, C_{1:T}) \\ = \frac{\prod_{t=1}^{T} f_{\mathcal{W}}(C_t | n_{(i)}, \Sigma_t/n_{(i)})}{\sum_{j=1}^{N}\prod_{t=1}^{T} f_{\mathcal{W}}(C_t | n_{(j)}, \Sigma_t/n_{(j)})},\end{aligned} \tag{A.20}$$

where $f_{\mathcal{W}}(C_t|\cdot)$ is defined in Eq. (2.3). That is, we generate a sample from the multinomial distribution in Eq. (2.21), by sampling n from $\{n_{(1)}, n_{(2)}, \ldots, n_{(N)}\}$ with probabilities (A.20).

A.2. BPF Approximation of the Likelihood in Equation (2.24)

The Deviance Information Criteria (DIC) calculations in Eq. (2.22) are based on the conditional likelihood in Eq. (2.24). Since the integral in Eq. (2.24) is analytically

intractable we have to rely on approximations. For its evaluation we use the Bootstrap particle filter (BPF) algorithm of Gordon et al. (1993). Based on the factorization of the conditional likelihood in Eq. (2.25), we implement p independent BPFs to approximate the asset-specific likelihood contributions

$$\begin{aligned} &p_i(C^r_{1:T}|C^{fr}_{1:T}, C^f_{1:T}; \theta) \\ &\quad = \int \left[\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2}\left[(n-q)x^e_{it} + nc^e_{it}\exp(-x^e_{it}) \right] \right\} f_{\mathcal{N}}(x^e_{it} \mid x^e_{i1:t-1}) \right] \mathrm{d}x^e_{i,1:T}, \end{aligned} \tag{A.21}$$

for $i = 1, \ldots, p$, where $f_{\mathcal{N}}(x^e_{it} \mid x^e_{i1:t-1})$ denotes the Gaussian transition density of the latent x^e_{it}'s.

In the following we state the BPF algorithm for any asset-specific component i, whereby we drop the conditioning on $C^{fr}_{1:T}$, $C^f_{1:T}$ and θ and we focus on AR(1) dynamics for the latent idiosyncratic components for notational convenience. Likelihood estimates for the WFSS models with latent HAR dynamics are found in a similar fashion.

The BPF algorithm for estimating the conditional likelihood in Eq. (A.21) produces Monte Carlo estimates for the sequence of the period-t likelihood contributions $p_i(C^r_t|C^r_{1:t-1})$ by sequentially importance sampling (IS) and re-sampling using the transition densities $f_{\mathcal{N}}(x^e_{it} \mid x^e_{it-1})$ as IS densities for the states x^e_{it}. For a given value of the parameters θ the BPF algorithm proceeds in the following steps:

1.) *For $t = 1$:* Initialize M draws $x^{e(1)}_{i1}, \ldots, x^{e(M)}_{i1}$ from $\mathcal{N}(m_{xi}, v^2_{xi})$, where m_{xi} and v^2_{xi} are the stationary mean and variance of the latent AR process, respectively, and compute the period-1 likelihood estimate $\hat{p}_i(C^r_1)$ as

$$\hat{p}_i(C^r_1) = \frac{1}{M}\sum_{j=1}^{M} w^{(j)}_{i1}, \qquad w^{(j)}_{i1} = e^{\left\{-\frac{1}{2}\left[(n-q)x^{e(j)}_{i1} + nc^e_{i1}\exp(-x^{e(j)}_{i1})\right]\right\}}. \tag{A.22}$$

Then compute the normalized weights $\widetilde{w}^{(j)}_{i1} = w^{(j)}_{i1} / \sum_{l=1}^{M} w^{(l)}_{i1}$ and re-sample by drawing $\widetilde{x}^{e(1)}_{i1}, \ldots, \widetilde{x}^{e(M)}_{i1}$ from the set $\{x^{e(j)}_{i1}\}_{j=1}^{M}$ according to the probabilities $\{\widetilde{w}^{(j)}_{i1}\}_{j=1}^{M}$.

2.) *For $t = 2, \ldots, T$:* Sample M draws $x_{it}^{e(1)}, \ldots, x_{it}^{e(M)}$ from the corresponding transitions $\mathcal{N}(\gamma_i^e + \phi_i^e(\widetilde{x}_{it-1}^{e(j)} - \gamma_i^e), [\nu_i^e]^2)$, set $x_{i,1:t}^{e(j)} = (\widetilde{x}_{i,1:t-1}^{e(j)}, x_{it}^{e(j)})$, $j = 1, \ldots, M$, and compute the period-t likelihood estimate $\hat{p}_i(C_t^r | C_{1:t-1}^r)$ as

$$\hat{p}_i(C_t^r \mid C_{1:t-1}^r) = \frac{1}{M} \sum_{j=1}^{M} w_{it}^{(j)}, \quad w_{it}^{(j)} = e^{\left\{-\frac{1}{2}\left[(n-q)x_{it}^{e(j)} + nc_{it}^e \exp(-x_{it}^{e(j)})\right]\right\}}. \tag{A.23}$$

Then compute the normalized weights $\widetilde{w}_{it}^{(j)} = w_{it}^{(j)} / \sum_{l=1}^{M} w_{it}^{(l)}$, $j = 1, \ldots, M$, and re-sample by drawing $\widetilde{x}_{i,1:t}^{e(1)}, \ldots, \widetilde{x}_{i,1:t}^{e(M)}$ from the set $\{x_{i,1:t}^{e(j)}\}_{j=1}^{M}$ according to the probabilities $\{\widetilde{w}_{it}^{(j)}\}_{j=1}^{M}$.

3.) The BPF estimate for $p_i(C_{1:T}^r)$ is then obtained as $\hat{p}_i(C_1^r) \prod_{t=2}^{T} \hat{p}_i(C_t^r \mid C_{1:t-1}^r)$.

Finally, the estimate for the conditional 'full' likelihood is given by

$$\hat{p}(C_{1:T}^r | C_{1:T}^{fr}, C_{1:T}^f; \theta) = K_T(n, C_{1:T}^e) \prod_{i=1}^{p} \hat{p}_i(C_{1:T}^r | C_{1:T}^{fr}, C_{1:T}^f; \theta), \tag{A.24}$$

where $K_T(n, C_{1:T}^e) = (n/2)^{T(n-q)p/2} \left[\Gamma_p((n-q)/2)\right]^{-T} \prod_{t=1}^{T} |C_t^e|^{(n-q-p-1)/2}$.

A.3. A Matrix-F Mixture Factor State-Space Model

In the following we derive a Matrix-F mixture factor state-space model by exploiting the Wishart-mixture representation of the Matrix-F distribution. Therefore, we assume

$$C_t^{-1} \mid S_t \overset{ind}{\sim} \mathcal{W}_m(n_2, S_t^{-1}), \qquad S_t \mid \Sigma_t \overset{ind}{\sim} \mathcal{W}_m(n_1, \Sigma_t/c), \tag{A.25}$$

where S_t is the $m \times m$ mixture-scale matrix, $n_1, n_2 > m-1$ are the d.o.f. parameters, and c is a scalar parameter scaling Σ_t. The matrices C_t and Σ_t are partitioned as stated in Eq. (2.1).

The marginal density of C_t^{-1} given Σ_t is readily obtained by integrating the joint

density $f(C_t^{-1}, S_t \mid \Sigma_t)$ w.r.t. S_t, i.e.,

$$\begin{aligned} f(C_t^{-1} \mid \Sigma_t) &= \int f(C_t^{-1}, S_t \mid \Sigma_t)\, \mathrm{d}S_t \\ &= \left[2^{(n_1+n_2)m/2}\Gamma_m\left(\frac{n_1}{2}\right)\Gamma_m\left(\frac{n_2}{2}\right)\right] |c\Sigma_t^{-1}|^{n_1/2}|C_t^{-1}|^{(n_2-m-1)/2} \\ &\quad \times \int |S_t|^{(n_1+n_2-m-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[S_t(C_t^{-1}+c\Sigma_t^{-1})\right]\right\}\, \mathrm{d}S_t. \end{aligned} \tag{A.26}$$

The identity $|S_t|^{(n_1+n_2-m-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[S_t(C_t^{-1}+c\Sigma_t^{-1})\right]\right\}$ in the latter equation is the kernel of a $\mathcal{W}_m(n_1+n_2, (C_t^{-1}+c\Sigma_t^{-1})^{-1})$ distribution for S_t. Hence,

$$\begin{aligned} &\int |S_t|^{(n_1+n_2-m-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[S_t(C_t^{-1}+c\Sigma_t^{-1})\right]\right\}\, \mathrm{d}S_t \\ &\qquad = 2^{(n_1+n_2)m/2}\Gamma_m\left(\frac{n_1+n_2}{2}\right)|C_t^{-1}+c\Sigma_t^{-1}|^{-(n_1+n_2)/2}. \end{aligned} \tag{A.27}$$

Combining (A.26) and (A.27) yields

$$f(C_t^{-1} \mid \Sigma_t) = \frac{\Gamma_m\left(\frac{n_1+n_2}{2}\right)}{\Gamma_m\left(\frac{n_1}{2}\right)\Gamma_m\left(\frac{n_2}{2}\right)} |c\Sigma_t^{-1}|^{n_1/2}|C_t^{-1}|^{(n_2-m-1)/2}|C_t^{-1}+c\Sigma_t^{-1}|^{-(n_1+n_2)/2}, \tag{A.28}$$

which is the pdf of a matrix-variate F distribution with $n_1, n_2 > m-1$ d.o.f. and pos. def. parameter matrix $c\Sigma_t^{-1}$ (see, e.g., Konno, 1991; Opschoor et al., 2017). From the latter result, i.e., $C_t^{-1} \mid \Sigma_t \sim F_m(n_2, n_1, c\Sigma_t^{-1})$, the conditional distribution of C_t given Σ_t follows as (Konno, 1988)

$$C_t \mid \Sigma_t \sim F_m(n_1, n_2, \Sigma_t/c), \tag{A.29}$$

with conditional moments

$$\mathrm{E}[C_t \mid \Sigma_t] = \frac{c^{-1}n_1}{n_2-m-1}\Sigma_t,$$

and

$$\begin{aligned}\mathrm{Cov}[\mathrm{vec}(C_t) \mid \Sigma_t] = \frac{c^{-2} n_1 (n_1 + n_2 - m - 1)}{(n_2 - m)(n_2 - m - 1)(n_2 - m - 3)} \\ \times \left((I_{m^2} + K_{mm})(\Sigma_t \otimes \Sigma_t) + \frac{2}{n_2 - m - 1} \mathrm{vec}(\Sigma_t)\mathrm{vec}(\Sigma_t)' \right),\end{aligned}$$

where K_{mm} denotes the $m^2 \times m^2$ commutation matrix with $K_{mm}\mathrm{vec}(A) = \mathrm{vec}(A')$ (see Lütkepohl, 1996, Chapter 9).[1] The second order moment exists if $n_2 - m - 3 > 0$. By setting the scalar c in Eq. (A.25) to $n_1/(n_2 - m - 1)$ we can assure $\mathrm{E}[C_t \mid \Sigma_t] = \Sigma_t$.

Assuming a partition for the mixture-scale S_t similar to C_t and Σ_t, we can derive a factorization of $f(S_t \mid \Sigma_t)$ similar to that stated in Eq. (2.7). Futhermore, we can exploit that the posterior $\pi(\Sigma_{1:T} \mid C_{1:T}, S_{1:T})$ is proportional to $f(\Sigma_{1:T}) \prod_{t=1}^{T} f(S_t|\Sigma_t)$. Hence, we can conduct MCMC sampling of the individual components of $\Sigma_{1:T}$ based on a draw for $S_{1:T}$ from its posterior distribution. The posterior of $S_{1:T}$ is given by

$$\pi(S_{1:T} \mid C_{1:T}, \Sigma_{1:T}) \propto f(S_{1:T}, C_{1:T}^{-1}, \mid \Sigma_{1:T}) \propto \prod_{t=1}^{T} f(C_t^{-1} \mid S_t) f(S_t, \mid \Sigma_t). \tag{A.30}$$

From Eq. (A.27) we know, that (A.30) is proportional to the product of independent Wishart densities with $n_1 + n_2$ d.o.f. and scale matrices $(C_t^{-1} + c\Sigma_t^{-1})^{-1}$, for $t = 1, \ldots, T$. This result allows to sample $S_1, \ldots, S_T \mid C_{1:T}, \Sigma_{1:T}$ independently from their respective posteriors.

The results we obtained for an initial posterior analysis of this generalization show that there is no evidence against the Wishart in favor of the fat-tailed Matrix-F distribution. Actually, the posterior estimates for all the parameters of the Matrix-F generalization are virtually equal to their values obtained for the fitted WFSS model. However, draws from the posterior of n_2 showed that for the additional d.o.f. parameter very large values ($n_2 > 10^3$) were found (detailed results are not presented here). Since for $n_2 \to \infty$ the Matrix-F collapses to the Wishart, this result indicates that the original WFSS model suffices to capture the tail behavior of the realized

[1] A derivation for the second order moment can, e.g., be found in Appendix B.

covariance data.

A.4. Impact of Parameter Uncertainty on Forecast Accuracy for the WFSS Model

In this Appendix we compare the accuracy of point forecasts for the realized covariance matrix of the asset returns C_t^r obtained under the WFSS model using forecasts ignoring the uncertainty about the parameters with the accuracy of forecasts accounting for this uncertainty.

Recall that when ignoring parameter uncertainty, forecasts for C_t^r are taken to be the mean of the forecast density $\mathrm{E}(C_{t+1}^r|C_{1:t};\theta)$, where the parameters θ are set to their (time-sequentially updated) estimated values (see Eqs. 2.27 – 2.29). Forecasts which account for parameter uncertainty are given by the mean of the predictive density (Geweke, 2005). This predictive density, denoted by $p(C_{t+1}^r|C_{1:t})$, obtains by replacing in the forecast density given in Eq. (2.27) the period-t conditional posterior of the states $\pi(\Sigma_{1:t}|C_{1:t};\theta)$ given the parameters θ by the corresponding joint posterior $\pi(\Sigma_{1:t},\theta|C_{1:t})$, and then integrating w.r.t. $\Sigma_{1:t+1}$ as well as θ, so that

$$p(C_{t+1}^r|C_{1:t}) = \iint f(C_{t+1}^r|\Sigma_{t+1};\theta)\ f(\Sigma_{t+1}|\Sigma_{1:t};\theta)\ \pi(\Sigma_{1:t},\theta|C_{1:t})\ \mathrm{d}\theta\mathrm{d}\Sigma_{1:t+1}. \quad \text{(A.31)}$$

The mean of this predictive density given by $\mathrm{E}(C_{t+1}^r|C_{1:t}) = \mathrm{E}(\Sigma_{t+1}^r|C_{1:t})$ can straightforwardly be approximated by

$$\mathrm{E}(C_{t+1}^r|C_{1:t}) \simeq \frac{1}{S}\sum_{i=1}^{S}\Sigma_{t+1}^{r(i)}, \quad \text{(A.32)}$$

where $\{\Sigma_{t+1}^{r(i)}\}_{i=1}^S$ are simulated draws from the convolution $f(\Sigma_{t+1}|\Sigma_{1:t};\theta)\,\pi(\Sigma_{1:t},\theta|C_{1:t})$ based on Gibbs simulations from the period-t posterior $\pi(\Sigma_{1:t},\theta|C_{1:t})$.

For the sake of comparison, we use for those Bayesian forecasts of C_{t+1}^r the same parameter-updating scheme as it is used for the forecasts based on $\mathrm{E}(C_{t+1}^r|C_{1:t};\theta)$ which consists of re-estimating the parameters every 10 time periods (see Section

2.4.3). For the Bayesian forecasts according to Eq. (A.32) this is implemented by sampling every 10 periods $\{\Sigma_{1:t}^{(i)}, \theta^{(i)}\}_i$ from the corresponding period-t joint posterior $\pi(\Sigma_{1:t}, \theta | C_{1:t})$ to obtain $\{\Sigma_{t+1}^{r(i)}\}_i$ from sampling for each pair $(\Sigma_{1:t}^{(i)}, \theta^{(i)})$ from the state transition $f(\Sigma_{t+1} | \Sigma_{1:t}^{(i)}; \theta^{(i)})$. Then the Gibbs draws of the parameters $\{\theta^{(i)}\}_i$ from such an up-dating period are kept fixed for the next 9 periods $(t+s, s = 1, \ldots, 9)$ and are used to sample $\{\Sigma_{t+s}^{(i)}\}_i$ from the corresponding period-$(t+s)$ conditional posterior $\pi(\Sigma_{1:t+s} | C_{1:t+s}; \theta^{(i)})$ and to obtain $\{\Sigma_{t+s+1}^{r(i)}\}_i$ from $f(\Sigma_{t+s+1} | \Sigma_{1:t+s}^{(i)}; \theta^{(i)})$.

In Table A.1 we provide the average RMSE losses for the Bayesian point forecasts $\mathrm{E}(C_{t+1}^r | C_{1:t})$ under the WFSS specification with the largest set of parameters (the unrestricted full 9-factor WFSS with time-varying loadings and HAR variance dynamics, 9F-HAR-vβ) together with those of the corresponding forecasts $\mathrm{E}(C_{t+1}^r | C_{1:t}; \theta)$ which are reproduced from Table 2.3. From the results we see that in our WFSS application parameter uncertainty is of very limited importance for the accuracy of point forecasts. This is to be expected in view of the low posterior uncertainty about the parameters as indicated by their posterior standard deviations provided in Figure 2.5. In fact, those standard deviations are fairly small relative to their posterior means (see Figure 2.4).

A.5. Additional Posterior Summary Results

We carefully checked the convergence of the proposed MCMC algorithm as well as its mixing rate (in terms of the correlation of the Gibbs draws). For this we monitored the trace plots of the Gibbs draws for the parameters and their sample autocorrelation function (ACF). Following the standard practice in the literature, we report in Section 2.4.2 as diagnostic summary statistics for the mixing rate the inefficiency factor (IF). For the unrestricted full WFSS model (9F-HAR-vβ) the IF values across all of its parameters range from 1.5 to 17.36 with an average value of 3.44 indicating a high sampling efficiency with a fast mixing rate of the MCMC algorithm. In Figure A.1 we provide the box plot of the IF values for all the parameters for the full 9F-HAR-vβ

WFSS model and in Figures A.2 – A.4 the trace plots of the Gibbs draws (including those of the burn-in period) together with the respective ACFs for the parameter with the smallest IF value (Figure A.2), the median IF value (Figure A.3) and the largest IF value (Figure A.4). Those plots corroborate the fast mixing rate of the MCMC algorithm.

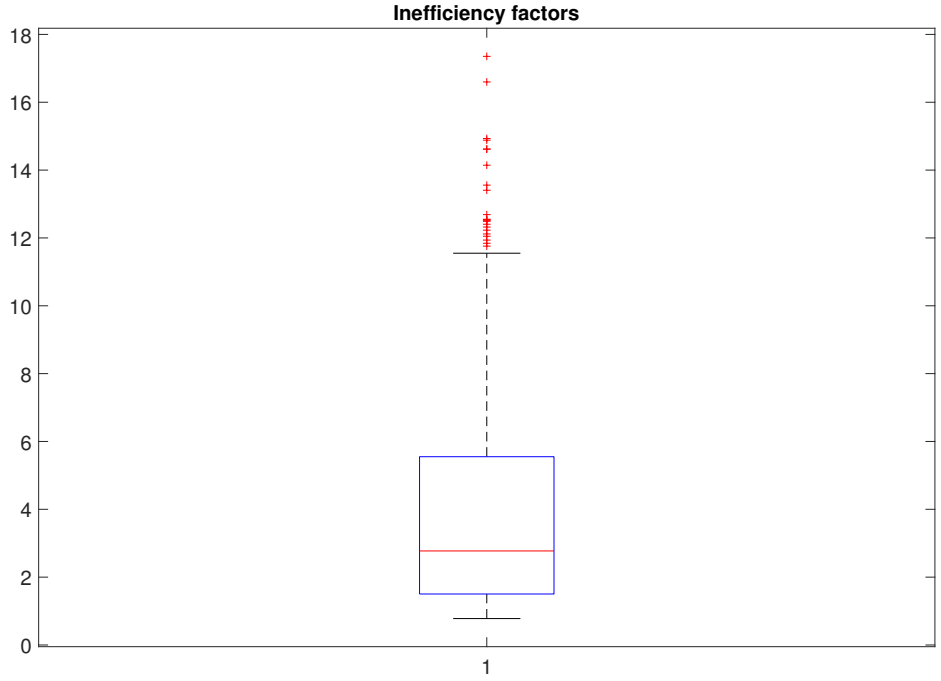
A.6. Additional Tables and Figures

Table A.1.: RMSE evaluation of point-forecast accuracy for the 9F-HAR-vβ WFSS model.

	Period 2011			Period 2012			Period 2011 - 2012		
Forecast rule	RMSE	RMSE^{v}	RMSE^{c}	RMSE	RMSE^{v}	RMSE^{c}	RMSE	RMSE^{v}	RMSE^{c}
$\mathrm{E}(C^r_{t+1}\|C_{1:t};\theta)$	51.24	16.87	33.54	23.02	10.63	14.03	36.51	13.61	23.36
$\mathrm{E}(C^r_{t+1}\|C_{1:t})$	51.58	16.94	33.78	23.06	10.65	14.06	36.69	13.65	23.48

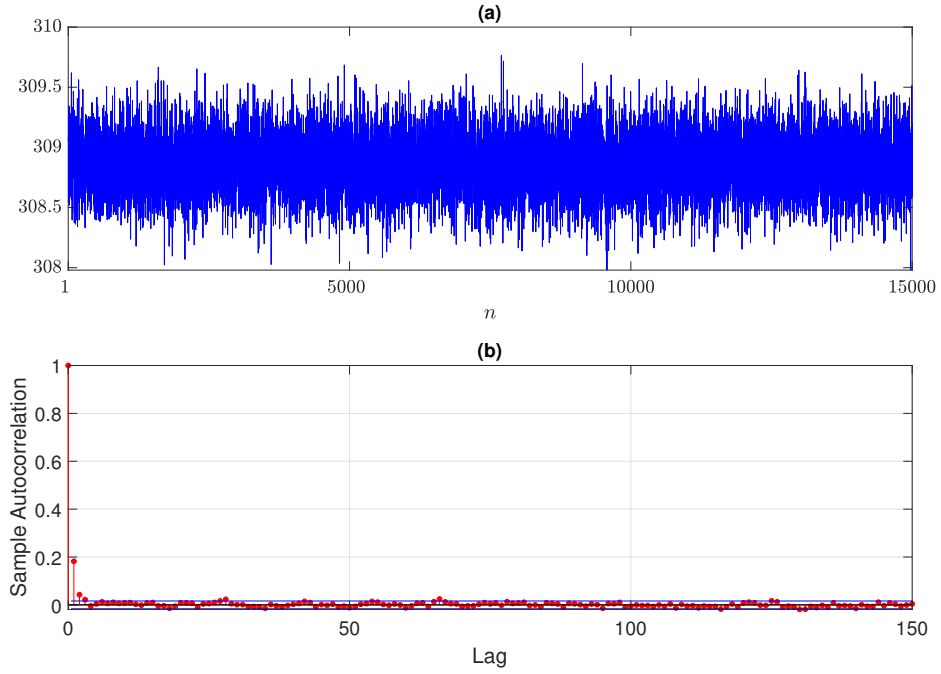
Note: The table reports the RMSE, RMSE^{v}, and RMSE^{c} losses as given in Eqs. (2.34) and (2.35).

Figure A.1.: Inefficiency factors.



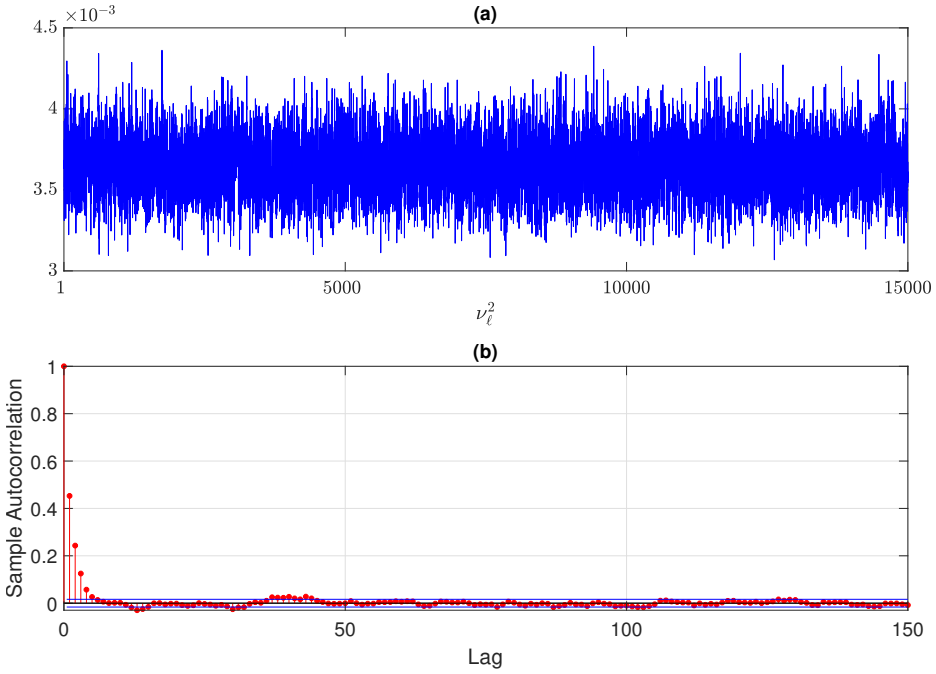
Box plots of the IF values for all parameters for the full unrestricted 9F-HAR-vβ WFSS model.

Figure A.2.: Traceplot for a parameter with small IF value.



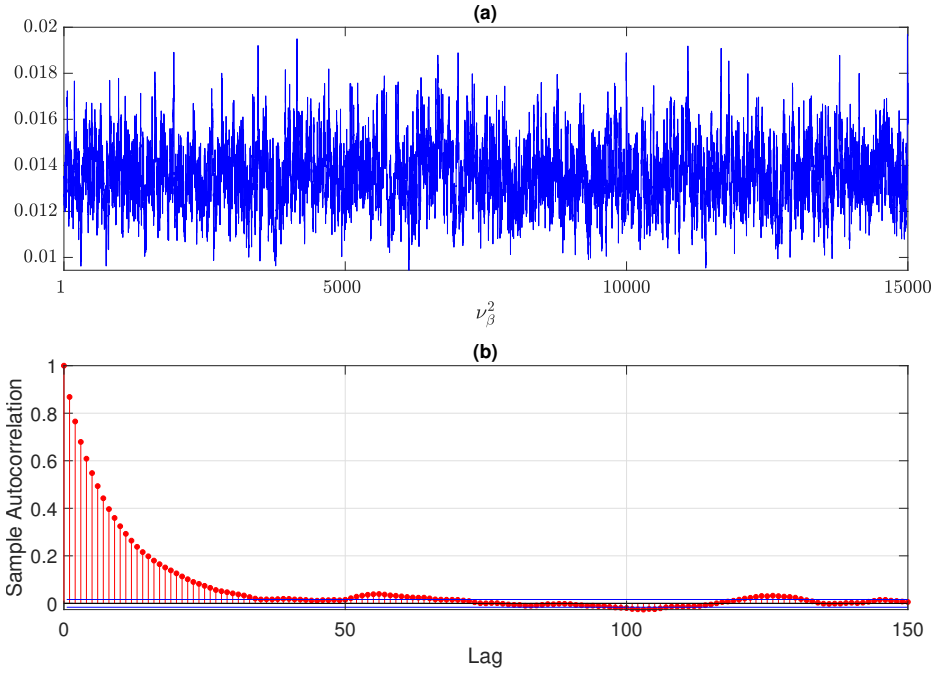
Upper panel: Trace plots of the Gibbs draws (including those of the burn-in period) for the parameter of the full unrestricted 9F-HAR-vβ WFSS model with the smallest IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Figure A.3.: Traceplot for a parameter with median IF value.



Upper panel: Trace plots of the Gibbs draws (including those of the burn-in period) for the parameter of the full unrestricted 9F-HAR-vβ WFSS model with the median IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Figure A.4.: Traceplot for a parameter with large IF value.



Upper panel: Trace plots of the Gibbs draws (including those of the burn-in period) for the parameter of the full unrestricted 9F-HAR-vβ WFSS model with the largest IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Appendix B.
Appendix for Chapter 3

B.1. Further Derivations for the UE Model

In the following we derive several results for the model at hand, which were not considered in the original work of Windle and Carvalho (2014). First, we derive a closed-form formula for the second order predictive moment, which, e.g., enables the calculation of confidence intervals for the point prediction of the covariance elements. Second, to improve the interpretability of the UE we derive the underlying dynamic process for the latent integrated covariance matrix, as implied by the Matrix-Beta transition for the integrated precision. Third, we derive the autoregressive dynamics of the logarithmic determinant of the latent precision process. E.g., Philipov and Glickman (2006) use these dynamics to deduce stationarity conditions for their matrix process.

For notational convenience reconsider the UE model without any superscript indicating correspondence to the factor or residual part. For any realized covariance C_t of dimension $d \times d$ the UE model is written as

$$C_t \mid \Sigma_t \sim \mathcal{W}_d(n, \Sigma_t/n), \tag{B.1}$$

$$\Omega_t = \mathcal{U}(\Omega_{t-1})' \Psi_t \mathcal{U}(\Omega_{t-1})/\lambda, \qquad \Psi_t \sim \mathcal{B}_d^I\left(\frac{k}{2}, \frac{n}{2}\right), \tag{B.2}$$

where $\Omega_t = \Sigma_t^{-1}$, with initial condition $\Omega_1 \sim \mathcal{W}_d(k, (nS_0)^{-1}/\lambda)$. In general the updating and filtering distributions for Ω_t are given by $\Omega_t | C_{1:t} \sim \mathcal{W}_d(n+k, (nS_t)^{-1})$ and $\Omega_{t+1} | C_{1:t} \sim \mathcal{W}_d(k, (nS_t)^{-1}/\lambda)$, respectively, where $S_t = \lambda S_{t-1} + C_t$ (see Windle and Carvalho, 2014, Prop. 1).

B.1.1. Second order predictive moments

The first order predictive moment of C_{t+1} given the information set $C_{1:t}$ is readily obtained by the law of iterated expectations (Konno, 1991; Windle and Carvalho, 2014)

$$\mathrm{E}[C_{t+1} \mid C_{1:t}] = \mathrm{E}\left[\mathrm{E}[C_{t+1} \mid \Sigma_{t+1}] \mid C_{1:t}\right] = \mathrm{E}\left[\Omega_{t+1}^{-1} \mid C_{1:t}\right] = \frac{\lambda n}{k-d-1} S_t.$$

It exists for $k > d+1$. The second order predictive moment, i.e., the covariance of $\mathrm{vec}(C_{t+1})$ given $C_{1:t}$, can be found using the law of total variance and the results in Muirhead (2005) and von Rosen (1988) for the moments of (inverted) Wishart matrices

$$\begin{aligned}
&\mathrm{Cov}[\mathrm{vec}(C_{t+1}) \mid C_{1:t}] \\
&\quad = \mathrm{E}\left[\mathrm{Cov}\left[\mathrm{vec}(C_{t+1}) \mid \Sigma_{t+1}\right] \mid C_{1:t}\right] + \mathrm{Cov}\left[\mathrm{E}\left[\mathrm{vec}(C_{t+1}) \mid \Sigma_{t+1}\right] \mid C_{1:t}\right] \\
&\quad = \mathrm{E}\left[n^{-1}(I_{d^2} + K_{dd})\left(\Omega_{t+1} \otimes \Omega_{t+1}\right) \mid C_{1:t}\right] + \mathrm{Cov}\left[\mathrm{vec}(\Omega_{t+1}) \mid C_{1:t}\right] \\
&\quad = \frac{1}{n}(I_{d^2} + K_{dd})\mathrm{E}\left[\Omega_{t+1} \otimes \Omega_{t+1} \mid C_{1:t}\right] + \mathrm{E}\left[\mathrm{vec}(\Omega_{t+1})\mathrm{vec}(\Omega_{t+1})' \mid C_{1:t}\right] \\
&\qquad\qquad - \mathrm{E}\left[\mathrm{vec}(\Omega_{t+1}) \mid C_{1:t}\right]\mathrm{E}\left[\mathrm{vec}(\Omega_{t+1}) \mid C_{1:t}\right]' \\
&\quad = \frac{(k+n-d-1)}{n(k-d)(k-d-1)(k-d-3)}\left[\lambda^2 n^2 (I_{d^2} + K_{dd})(S_t \otimes S_t) + \frac{2\lambda^2 n^2}{(k-d-1)}\mathrm{vec}(S_t)\mathrm{vec}(S_t)'\right] \\
&\quad = \frac{\lambda^2 n(k+n-d-1)}{(k-d)(k-d-1)(k-d-3)}\left[(I_{d^2} + K_{dd})(S_t \otimes S_t) + \frac{2}{(k-d-1)}\mathrm{vec}(S_t)\mathrm{vec}(S_t)'\right]
\end{aligned}$$

where K_{dd} denotes the $d^2 \times d^2$ commutation matrix with $K_{dd}\mathrm{vec}(A) = \mathrm{vec}(A')$ (see Lütkepohl, 1996, Chapter 9). The second order moment exists if $k > d+3$.

B.1.2. Underlying latent dynamic process for the integrated covariance

The Matrix-Beta type-I transition equation in Eq. (B.2) implies a shifted Matrix-F transition for the underlying latent integrated covariance matrix Σ_t itself. Defining

$\Omega_t^{-1} = \Sigma_t$, we have

$$\begin{aligned}
\Omega_t^{-1} &= \lambda(\mathcal{U}(\Omega_{t-1}))^{-1}\Psi_t^{-1}(\mathcal{U}(\Omega_{t-1})')^{-1} \\
&= \lambda(\mathcal{U}(\Omega_{t-1}))^{-1}(\Psi_t^{-1} - I_d + I_d)(\mathcal{U}(\Omega_{t-1})')^{-1} \\
&= \lambda(\mathcal{U}(\Omega_{t-1}))^{-1}(\Psi_t^{-1} - I_d)(\mathcal{U}(\Omega_{t-1})')^{-1} + \lambda(\mathcal{U}(\Omega_{t-1})'\mathcal{U}(\Omega_{t-1}))^{-1} \\
&= \lambda(\mathcal{U}(\Omega_{t-1}))^{-1}\Upsilon_t(\mathcal{U}(\Omega_{t-1})')^{-1} + \lambda\Omega_{t-1}^{-1},
\end{aligned}$$

where $\Upsilon_t = (\Psi_t^{-1} - I_d)$ follows a d-dimensional Matrix-Beta type-II distribution (also known as Matrix-F distribution), say $\Upsilon_t \sim \mathcal{B}_d^{II}(\frac{n}{2}, \frac{k}{2})$, with expectation $\mathrm{E}[\Upsilon_t] = \frac{n}{k-d-1} I_d$ (Gupta and Nagar, 2000, Theorems 5.3.5 (and below) and 5.3.20), and I_d denotes the d-dimensional identity matrix. It follows, that the conditional distribution of $\Omega_t^{-1} = \Sigma_t$ given Ω_{t-1} is of Matrix-F type shifted by the factor $\lambda\Omega_{t-1}^{-1}$, and the conditional expectation is given by

$$\mathrm{E}[\Omega_t^{-1} \mid \Omega_{t-1}] = \lambda\left(1 + \frac{n}{k-d-1}\right)\Omega_{t-1}^{-1}.$$

By restricting the parameter λ through (R1) the Ω_t^{-1} process becomes a martingale with conditional expectation $\mathrm{E}[\Omega_t^{-1}|\Omega_{t-1}] = \Omega_{t-1}^{-1}$.

B.1.3. Log determinant process for the latent precision

From Eq. (B.2) the evolution of the logarithmic determinant of Ω_t follows as

$$\begin{aligned}
\log|\Omega_t| &= \log|\mathcal{U}(\Omega_{t-1})'\Psi_t\mathcal{U}(\Omega_{t-1})/\lambda| \\
&= \log\left\{|\mathcal{U}(\Omega_{t-1})'\mathcal{U}(\Omega_{t-1})| \cdot |\Psi_t| \cdot \lambda^{-d}\right\} \\
&= \log|\Omega_{t-1}| + \log|\Psi_t| - d\log\lambda.
\end{aligned}$$

Setting $\lambda = \exp\{\mathrm{E}[\log|\Psi_t|]/d\}$ according to restriction (R3) results in a random walk process for the log determinant as

$$\log|\Omega_t| = \log|\Omega_{t-1}| + (\log|\Psi_t| - \mathrm{E}[\log|\Psi_t|]) = \log|\Omega_{t-1}| + \widetilde{\psi}_t,$$

with $\widetilde{\psi}_t = \log|\Psi_t| - \mathrm{E}[\log|\Psi_t|]$ being a white noise increment. Hence, the $\log|\Omega_t|$ process is a martingale with conditional expectation $\mathrm{E}[\log|\Omega_t| \mid \Omega_{t-1}] = \log|\Omega_{t-1}|$. The unconditional expectation for the logarithmic determinant of Ψ_t follows from $\Psi_t \sim \mathcal{B}_d^I(\frac{k}{2}, \frac{n}{2})$ as

$$\begin{aligned}
\mathrm{E}[\log|\Psi_t|] &= \left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right]^{-1} \int\limits_{0<\Psi_t<I_d} (\log|\Psi_t|)|\Psi_t|^{(k-d-1)/2}|I_d - \Psi_t|^{(n-d-1)/2}\mathrm{d}\Psi_t \\
&= 2\left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right]^{-1} \int\limits_{0<\Psi_t<I_d} \frac{\partial}{\partial k}|\Psi_t|^{(k-d-1)/2}|I_d - \Psi_t|^{(n-d-1)/2}\mathrm{d}\Psi_t \\
&= 2\left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right]^{-1} \frac{\partial}{\partial k} \int\limits_{0<\Psi_t<I_d} |\Psi_t|^{(k-d-1)/2}|I_d - \Psi_t|^{(n-d-1)/2}\mathrm{d}\Psi_t \\
&= 2\left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right]^{-1} \frac{\partial}{\partial k}\left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right] \\
&= 2\frac{\partial}{\partial k}\log\left[\frac{\Gamma_d\left(\frac{n}{2}\right)\Gamma_d\left(\frac{k}{2}\right)}{\Gamma_d\left(\frac{n+k}{2}\right)}\right] \\
&= \Psi_d^\star\left(\frac{k}{2}\right) - \Psi_d^\star\left(\frac{k+n}{2}\right),
\end{aligned}$$

with $\Psi_d^\star(a) = \partial \log\Gamma_d(a)/\partial a$ being the d-variate digamma function (see, e.g., Abramowitz and Stegun, 1972; Gupta and Nagar, 2000).

B.2. Parameter Estimation Results

This section presents selected summary results of the in-sample parameter estimates. Table B.1 reports a summary of the parameter estimates for the factor and residual components of the 12F CFSS model. The results show similar estimates for the d.o.f.

parameters n and k across the restrictions (R0) – (R3). Solely under restriction (R1) the estimates for n tend to be higher in comparison to the other restrictions. The smoothing parameter estimates also only slightly differ across the restrictions with highest (median) values for restrictions (R1) and (R2). Interestingly, the (mean and median) smoothing parameter estimates under restrictions (R0) and (R3) do not differ at all within the first four digits. This finding is confirmed by Figure B.1, which displays the cross-section of smoothing parameter estimates for the residual components in the 12F CFSS model. Panel (a) shows the results for the diagonal specification, Panel (b) for the block-diagonal specification. Here, both (R0) and (R3) are congruent throughout all blocks.

Mean and median values for the d.o.f. parameter k_i^b across the asset dimension in different factor settings are presented in Table B.2. In the out-of-sample exercise the d.o.f. are not re-estimated based on a rolling window scheme, but fixed to their (rounded) in-sample median value. The corresponding smoothing parameters are set to $\lambda_i^b = k_i^b/(k_i^b + 1)$.

B.3. Composite vs. Simulated Predictions

Since the point forecasts using Eq. (3.20) are of composite nature and neglect potential nonlinear structures in the term $B_{t+1}C_{t+1}^f B_{t+1}'$ when predicting C_{t+1}^r, we compare the ad-hoc forecasting method with the following 'unbiased' prediction based on

$$\widehat{C}_{t+1}^{r\star} \approx \frac{1}{10^3} \sum_{i=1}^{10^3} \left\{ X_b^{(i)} X_f^{(i)} X_b^{(i)\prime} + X_e^{(i)} \right\},$$

where $X_f^{(i)}$ and $X_e^{(i)}$ are draws from the respective predictive distribution in Eq. (3.11), and $X_b^{(i)}$ are draws from the predictive distribution in Eq. (3.14).

Figure B.2 shows the results of the point forecasts for the realized covariance of the nine assets from the Materials sector at the first day of the out-of-sample period ($t = 756$). The blue curves display the approximate predictive densities of C_{t+1}^r,

dashed black lines are the point forecasts using the simulation setting above, solid red lines are the point forecasts obtained by the composite method from Section 3.2. The results show a slight undershooting of the variance forecasts when using the ad-hoc method, which is negligible. For the covariance elements there is virtually no difference between both approaches. The exercise has been repeated for several days of the out-of-sample period finding similar results.

B.4. Additional Tables and Figures

Table B.1.: Summary of in-sample parameter estimates of the factor and residual components for the 12F CFSS model.

	n_i^f		k_i^f		λ_i^f	
Factor	Mean	Median	Mean	Median	Mean	Median
R0	124.30	—	52.92	—	0.6892	—
R1	137.87	—	52.17	—	0.7053	—
R2	126.20	—	52.06	—	0.7080	—
R3	123.94	—	52.93	—	0.6892	—
	n_i^e		k_i^e		λ_i^e	
Residual	Mean	Median	Mean	Median	Mean	Median
R0-D	22.90	23.02	11.75	11.71	0.6417	0.6513
R1-D	25.25	25.58	11.50	11.55	0.6621	0.6684
R2-D	23.28	23.55	11.49	11.55	0.6626	0.6698
R3-D	22.87	22.97	11.75	11.71	0.6417	0.6513
R0-S	252.31	237.72	130.41	117.86	0.6455	0.6558
R1-S	281.30	264.36	127.68	116.49	0.6650	0.6746
R2-S	258.44	244.19	127.23	116.42	0.6685	0.6786
R3-S	251.95	237.52	130.41	117.86	0.6455	0.6558

Note: This table reports the average and median across the in-sample Maximum likelihood estimates for the 12F CFSS model for the factor and residual parameters.

Table B.2.: Summary of in-sample parameter estimates for the d.o.f. parameters of the factor loading components.

	k_i^b		λ_i^b	
	Mean	Median	Mean	Median
1F	9.07	8.46	0.9007	0.8943
3F	15.68	15.00	0.9400	0.9375
12F	23.01	22.72	0.9584	0.9578

Note: This table reports the average and median across the in-sample Maximum likelihood estimates of the d.o.f. parameter k_i^b, $i = 1, \ldots, 225$, for the factor loading components in the 1F, 3F and 12F CFSS models. The corresponding smoothing parameters are calculated as $\lambda_i^b = k_i^b/(k_i^b + 1)$.

Table B.3.: List of the stocks included in the data set.

#	Symb.	Company	Sector	#	Symb.	Company	Sector
1	do	Diamond Offshore Drilling	E	29	lmt	Lockheed Martin Corporation	I
2	ne	Noble Corp	E	30	noc	Northrop Grumman Corp.	I
3	rig	Transocean	E	31	rtn	Raytheon Co.	I
4	bhi	Baker Hughes Incorporated	E	32	utx	United Technologies	I
5	hal	Halliburton Company	E	33	jci	Johnson Controls International	I
6	slb	Schlumberger Ltd.	E	34	mas	Masco Corporation	I
7	cvx	Chevron Corporation	E	35	glw	Corning Incorporated	I
8	hes	Hess Corporation	E	36	emr	Emerson Electric Company	I
9	xom	Exxon Mobil Corp.	E	37	ge	General Electric Company	I
10	apa	Apache Corporation	E	38	hon	Honeywell International Inc.	I
11	apc	Anadarko Petroleum Corporation	E	39	mmm	3M Company	I
12	cop	ConocoPhillips	E	40	cat	Caterpillar, Inc.	I
13	mro	Marathon Oil Corp.	E	41	de	Deere & Company	I
14	oxy	Occidental Petroleum Corporation	E	42	dov	Dover Corp.	I
15	vlo	Valero Energy	E	43	ir	Ingersoll-Rand PLC	I
16	wmb	Williams Cos.	E	44	itw	Illinois Tool Works	I
17	dd	E.I. du Pont de Nemours and Company	M	45	wm	Waste Management Inc.	I
18	dow	Dow Chemical Company	M	46	fdx	FedEx Corporation	I
19	apd	Air Products & Chemicals Inc	M	47	ups	United Parcel Service	I
20	px	Praxair Inc.	M	48	luv	Southwest Airlines Company	I
21	ppg	PPG Industries	M	49	csx	CSX Corp.	I
22	aa	Alcoa Inc.	M	50	nsc	Norfolk Southern Corp.	I
23	abx	Barrick Gold Corporation	M	51	unp	Union Pacific	I
24	nem	Newmont Mining Corporation	M	52	gt	Goodyear Tire & Rubber	D
25	ip	Internation Paper Company	M	53	f	Ford Motor Company DEL	D
26	ba	Boeing Company	I	54	hog	Harley-Davidson	D
27	gd	General Dynamics	I	55	dhi	D. R. Horton	D
28	lll	L-3 Communications Holdings	I	56	len	Lennar Corp.	D

Table B.3.: List of the stocks included in the data set.

#	Symb.	Company	Sector	#	Symb.	Company	Sector
57	igt	International Game Technology	D	85	bbby	Bed Bath & Beyond Inc.	D
58	anf	Abercrombie & Fitch	D	86	spls	Staples Inc.	D
59	apol	Apollo Group Inc.	D	87	cvs	CVS Caremark Corp.	S
60	nke	Nike, Inc.	D	88	wba	Wallgreens Boots Alliance	S
61	ccl	Carnival Corp.	D	89	syy	Sysco Corp.	S
62	mar	Marriott Int'l.	D	90	kr	Kroger Company	S
63	dri	Darden Restaurants	D	91	cost	Costco Wholesale Corporation	S
64	mcd	McDonald's Corporation	D	92	wmt	Wal-Mart Stores	S
65	sbux	Starbucks Corp.	D	93	cce	Coca-Cola Enterprises	S
66	yum	Yum! Brands Inc.	D	94	ko	Cocoa-Cola Company	S
67	ipg	Interpublic Group	D	95	pep	Pepsico, Inc.	S
68	omc	Omnicom Group	D	96	adm	Archer-Daniels-Midland Co	S
69	cbs	CBS Corporation	D	97	cag	ConAgra, Inc.	S
70	dis	Walt Disney Company	D	98	gis	General Mills	S
71	twx	Time Warner Inc.	D	99	k	Kellogg Co.	S
72	tgna	Tegna, Inc.	D	100	tsn	Tyson Foods	S
73	amzn	Amazon.com, Inc.	D	101	mo	Altria Group	S
74	jcp	J.C. Penney Company, Inc. Holding Co.	D	102	cl	Colgate-Palmolive Company	S
75	m	Macy's Inc.	D	103	clx	The Clorox Company	S
76	fdo	Family Dollar Stores Inc.	D	104	kmb	Kimberly-Clark	S
77	lb	L Brands Inc.	D	105	avp	Avon Products	S
78	kss	Kohl's Corporation	D	106	pg	Procter & Gamble Company	S
79	tgt	Target Corp.	D	107	bdx	Becton Dickinson	H
80	gps	Gap, Inc.	D	108	dhr	Danaher Corp.	H
81	tjx	TJX Companies Inc.	D	109	syk	Stryker Corp.	H
82	bby	Best Buy Co., Inc.	D	110	var	Varian Medical Systems	H
83	hd	Home Depot, Inc.	D	111	zbh	Zimmer Biomet Holdings	H
84	low	Lowe's Companies, Inc.	D	112	mck	McKesson Corp.	H

Table B.3.: List of the stocks included in the data set.

#	Symb.	Company	Sector	#	Symb.	Company	Sector
113	thc	Tenet Healthcare	H	141	slm	SLM Corporation	F
114	aet	Aetna Inc	H	142	axp	American Express Company	F
115	antm	Anthem Inc.	H	143	cof	Capital One Financial	F
116	hum	Humana Inc.	H	144	hrb	Block H&R	F
117	unh	United Health Group Inc.	H	145	bk	The Bank of New York Mellon Corp.	F
118	a	Agilent Technologies Inc.	H	146	ntrs	Northern Trust Corp.	F
119	abt	Abbott Laboratories	H	147	stt	State Street Corp.	F
120	bax	Baxter International Inc.	H	148	etfc	E*Trade	F
121	bmy	Bristol-Myers Squibb Company	H	149	gs	Goldman Sachs Group, Inc.	F
122	cah	Cardinal Health Inc.	H	150	ms	Morgan Stanley Dean Witter & Co	F
123	jnj	Johnson & Johnson	H	151	schw	Charles Schwab Corporation	F
124	mdt	Medtronic Inc.	H	152	pcl	Plum Creek Timber	F
125	stj	St. Jude Medical	H	153	aon	Aon plc	F
126	amgn	Amgen, Inc.	H	154	mmc	Marsh & McLennan Companies, Inc.	F
127	gild	Gilead Sciences Inc.	H	155	afl	AFLAC Inc	F
128	agn	Allergan, Plc	H	156	met	MetLife, Inc.	F
129	lly	Eli Lilly and Company	H	157	pru	Prudential Financial	F
130	mrk	Merck & Company, Inc.	H	158	lnc	Lincoln National	F
131	pfe	Pfizer, Inc.	H	159	aig	American International Group Inc.	F
132	usb	U.S. Bancorp	F	160	all	Allstate Corporation	F
133	wfc	Wells Fargo	F	161	cb	Chubb Limited	F
134	key	KeyCorp	F	162	hig	Hartford Financial Services Group, Inc.	F
135	pnc	PNC Financial Services	F	163	pgr	Progressive Corp.	F
136	sti	SunTrust Banks	F	164	trv	The Travelers Companies Inc.	F
137	bac	Bank of America Corporation	F	165	xl	XL Capital	F
138	c	Citigroup Inc.	F	166	adp	Automatic Data Processing	T
139	fitb	Fifth Third Bancorp	F	167	ctxs	Citrix Systems, Inc.	T
140	jpm	J P Morgan Chase & Co	F	168	ebay	Ebay Inc.	T

Table B.3.: List of the stocks included in the data set.

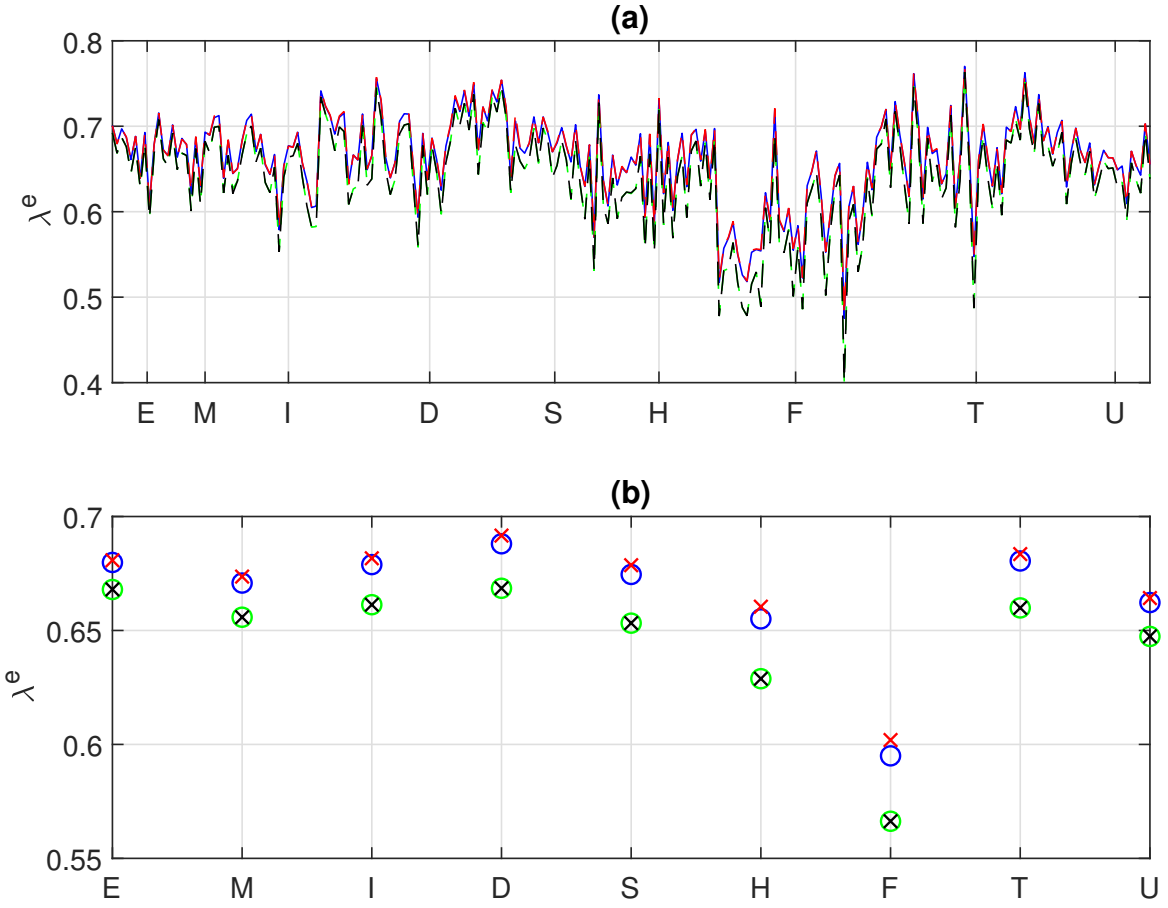
#	Symb.	Company	Sector	#	Symb.	Company	Sector
169	intu	Intuit Inc.	T	197	lltc	Linear Technology Corporation	T
170	ntap	NetApp, Inc.	T	198	mu	Motorola Solutions, Inc.	T
171	payx	Paychex, Inc.	T	199	nvda	Micron Technology, Inc.	T
172	vrsn	Verisign Inc.	T	200	qcom	NVIDIA Corporation	T
173	yhoo	Yahoo Inc.	T	201	lrcx	Lam Research	T
174	acn	Accenture plc	T	202	ter	Teradyne	T
175	csc	Computer Sciences Corp	T	203	adi	Analog Devices Inc.	T
176	emc	EMC Corporation MA	T	204	altr	Altera Corporation	T
177	ibm	International Business Machines Corp.	T	205	amd	Advanced Micro Devices Inc.	T
178	jbl	Jabil Circuit	T	206	brcm	Broadcom Corporation	T
179	adbe	Adobe Systems Incorporated	T	207	mchp	Microchip Technology	T
180	adsk	Autodesk Inc	T	208	txn	Texas Instruments	T
181	orcl	Oracle Corporation	T	209	xlnx	Xilinx Inc.	T
182	symc	Symantec Corp.	T	210	aep	American Electric Power	U
183	ca	CA, Inc.	T	211	d	Dominion Resources	U
184	msft	Microsoft Corporation	T	212	duk	Duke Energy Corporation	U
185	ea	Electronic Arts Inc.	T	213	ed	Consolidated Edison	U
186	hpq	Hewlett-Packard Company	T	214	etr	Entergy Corp.	U
187	lxk	Lexmark Int,l Inc	T	215	fe	FirstEnergy Corp	U
188	csco	Cisco Systems, Inc.	T	216	peg	Public Serv. Enterprise Inc.	U
189	jnpr	Juniper Networks, Inc.	T	217	so	Southern Co.	U
190	cien	Ciena Corporation	T	218	exc	Exelon Corp.	U
191	aapl	Apple Inc.	T	219	nee	NextEra Energy	U
192	sndk	SanDisk Corporation	T	220	ni	NiSource Inc.	U
193	xrx	Xerox Corp.	T	221	pcg	PG&E Corp.	U
194	amat	Applied Materials Inc.	T	222	sre	Sempra Energy	U
195	intc	Intel Corporation	T	223	te	TECO Energy	U
196	klac	KLA-Tencor Corporation	T	224	xel	Xcel Energy Inc	U

Table B.3.: List of the stocks included in the data set.

#	Symb.	Company	Sector	#	Symb.	Company	Sector
225	aes	AES Corp	U				

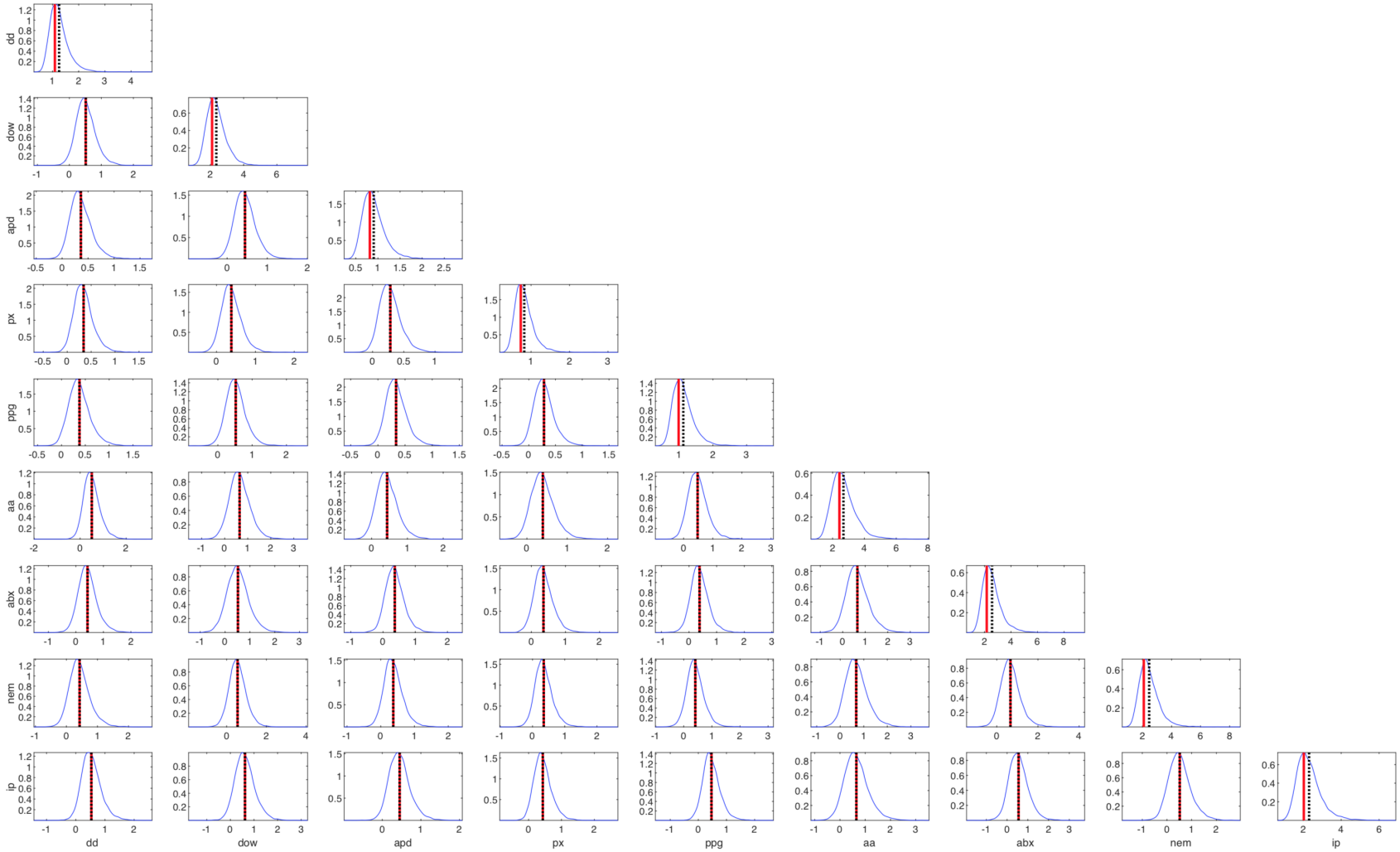
Note: Stocks are selected by liquidity from the S&P500 index and sorted by their sector and industry classification according to the Global Industry Classification Standard (GICS). Sector labels are: (E) Energy; (M) Materials; (I) Industrials; (D) Consumer Discretionary; (S) Consumer Staples; (H) Health Care; (F) Financials; (T) Information Technologies; (U) Utilities.

Figure B.1.: Smoothing parameter estimates for the residual components.



In-sample Maximum likelihood estimates of the smoothing parameters for the residual components under different restrictions. Panel (a): Diagonal residual components. Panel (b): Block-diagonal residual components based on GICS sectors. R0: black; R1: blue; R2: red; R3: green.

Figure B.2.: Simulated vs. composite predictions.



Point forecasts for the realized covariance of the nine assets from the Materials sector at the first day of the out-of-sample period ($t = 756$). Blue curve: approximate predictive densities of C^r_{t+1}; dashed black line: point forecasts using the simulation setting in B.3; solid red line: point forecasts obtained by the composite method from Section 3.2.

Appendix C.

Appendix for Chapter 4

C.1. Proof of Proposition 2

Note, whenever unambiguous in the following the subscript t and the conditioning on Σ are dropped for notational convenience.

C.1.1. Characteristic function

We obtain

$$\begin{aligned}\varphi_{\mathcal{R}}(\Theta) &= \mathrm{E}\left[\exp\left\{\iota \mathrm{tr}(\Theta S)\right\}\right] \\ &= \int_{C>0} \exp\left\{\iota \mathrm{tr}(\Theta C)\right\} f(C \mid \Sigma, \bar{n})\ \mathrm{d}C \\ &= \pi^{-k(k-1)/4} \left[\prod_{i=1}^{k} \frac{\left(\frac{n_i}{2}\right)^{\frac{n_i}{2}}}{\Gamma\left(\frac{n_i-i+1}{2}\right)\left(\sigma_{i\cdot i-1}\right)^{\frac{n_i}{2}}}\right] \\ &\qquad \times \int_{C>0} \prod_{i=1}^{k} \left(c_{i\cdot i-1}\right)^{\frac{n_i-k-1}{2}} \exp\left\{-\mathrm{tr}\left(\Xi(\Theta)^{-1}C\right)\right\}\ \mathrm{d}C,\end{aligned} \tag{C.1}$$

where $\Xi(\Theta) \equiv (\Sigma^{-\frac{\bar{n}}{2}} - \iota\Theta)^{-1}$. The integral in the latter equation is known as the well-defined *Siegel integral*, which is finite if and only if $n_i > i-1$, $i = 1, \ldots, k$, and in that case it evaluates to (see Andersson and Klein, 2010, Eq. (4))

$$\int_{C>0} \prod_{i=1}^{k} \left(c_{i\cdot i-1}\right)^{\frac{n_i-k-1}{2}} \exp\left\{-\mathrm{tr}\left(\Xi^{-1}C\right)\right\} \mathrm{d}C = \pi^{\frac{k(k-1)}{4}} \prod_{i=1}^{k} \left(\frac{\Gamma\left(\frac{n_i-i+1}{2}\right)}{\left(\xi_{i\cdot i-1}\right)^{-\frac{n_i}{2}}}\right). \tag{C.2}$$

By combining Eqs. (C.1) and (C.2) the result follows as

$$\varphi_{\mathcal{R}}(\Theta) = \prod_{i=1}^{k} \left(\frac{n_i}{2\sigma_{i \cdot i-1}} \right)^{\frac{n_i}{2}} \prod_{i=1}^{k} \left(\xi_{i \cdot i-1} \right)^{\frac{n_i}{2}}. \tag{C.3}$$

This completes the proof.

It proves helpful to rewrite the characteristic function in terms of the so-called *generalized power* or *highest weight vector* $\Delta_{\frac{\bar{n}}{2}}(\cdot)$

$$\varphi_{\mathcal{R}}(\Theta) = \Delta_{\frac{\bar{n}}{2}}(\Sigma^{-\frac{\bar{n}}{2}}) \Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta)), \tag{C.4}$$

where

$$\Delta_{\frac{\bar{n}}{2}}(A) = \prod_{i=1}^{k} (a_{i \cdot i-1})^{\frac{n_i}{2}} = |A_{[1]}|^{\frac{n_1}{2}} \prod_{i=2}^{k} \left(\frac{|A_{[i]}|}{|A_{[i-1]}|} \right)^{\frac{n_i}{2}} = \prod_{i=1}^{k} |A_{[i]}|^{\frac{n_i - n_{i+1}}{2}}, \tag{C.5}$$

with $A_{[i]}$ being the ith submatrix of $A = A_{[k]}$ and the notational convention $n_{k+1} = 0$. Note that a submatix $A_{[i]}$ of A can be equivalently expressed in terms of a $(k \times i)$ *selector matrix* $E_i = [I_i, 0_{(i \times k-i)}]'$ and $E_k = I_k$ in the way $A_{[i]} = E_i' A E_i$ (see Díaz-García, 2013).

C.1.2. Expectation

We have

$$\mathrm{E}[\mathrm{vec}(C)] = \left. \frac{\partial \varphi_{\mathcal{R}}(\Theta)}{\iota \partial \mathrm{vec}(\Theta)'} \right|_{\Theta=0}.$$

The derivative of $\varphi_{\mathcal{R}}(\Theta)$ w.r.t $\mathrm{vec}(\Theta)$ is obtained by considering the differential of $\varphi_{\mathcal{R}}$, i.e.,

$$\mathrm{d}\varphi_{\mathcal{R}}(\Theta) = \Delta_{\frac{\bar{n}}{2}}(\Sigma^{-\frac{\bar{n}}{2}}) \, \mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta)), \tag{C.6}$$

where $\Xi(\Theta) = (\Sigma^{-\frac{\bar{n}}{2}} - \iota\Theta)^{-1}$. By successively applying the product rule $\mathrm{d}(AB) = (\mathrm{d}A)B + A(\mathrm{d}B)$, $\mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta))$ obtains as[1]

$$\mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta)) = \sum_{i=1}^{k} \mathrm{d}\left(|E_i'\Xi(\Theta)E_i|^{\frac{n_i-n_{i+1}}{2}}\right) \prod_{j=1,j\neq i}^{k} |E_j'\Xi(\Theta)E_j|^{\frac{n_j-n_{j+1}}{2}}, \tag{C.7}$$

where the differential on the right hand side can be obtained from $\mathrm{d}|F(A)|^p = p|F(A)|^p\mathrm{tr}(F(A)^{-1}\mathrm{d}F(A))$ and $\mathrm{d}((F(A))^{-1}) = -(F(A))^{-1}\mathrm{d}F(A)(F(A))^{-1}$ (see, e.g., Magnus and Neudecker, 1988) as

$$\begin{aligned}&\mathrm{d}\left(|E_i'\Xi(\Theta)E_i|^{\frac{n_i-n_{i+1}}{2}}\right)\\ &= \iota\left(\frac{n_i-n_{i+1}}{2}\right)|E_i'\Xi(\Theta)E_i|^{\frac{n_i-n_{i+1}}{2}}\ \mathrm{tr}\left(\Xi(\Theta)E_i(E_i'\Xi(\Theta)E_i)^{-1}E_i'\Xi(\Theta)\mathrm{d}\Theta\right),\end{aligned} \tag{C.8}$$

since $\mathrm{d}(\Sigma^{-\frac{\bar{n}}{2}} - \iota\Theta) = -\iota\mathrm{d}\Theta$. Now combining Eqs. (C.7), (C.8) and exploiting $\mathrm{tr}(AB) = \mathrm{vec}(A)'\mathrm{vec}(B) = \mathrm{vec}(B)'\mathrm{vec}(A)$ (see Lütkepohl, 1996, Eq. 2.2.17) yields

$$\begin{aligned}&\mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta))\\ &= \iota\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta))\left[\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}\left(\Xi(\Theta)E_i(E_i'\Xi(\Theta)E_i)^{-1}E_i'\Xi(\Theta)\right)\right]'\mathrm{dvec}(\Theta).\end{aligned} \tag{C.9}$$

Finally, by plugging (C.9) in (C.6) and differentiating w.r.t. $\mathrm{vec}(\Theta)$ and dividing by ι we obtain

$$\frac{\partial\varphi_{\mathcal{R}}(\Theta)}{\iota\partial\mathrm{vec}(\Theta)'} = \underbrace{\Delta_{\frac{\bar{n}}{2}}(\Sigma^{-\frac{\bar{n}}{2}})\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta))}_{=\varphi_{\mathcal{R}}(\Theta)}\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}\left(\Xi(\Theta)E_i(E_i'\Xi(\Theta)E_i)^{-1}E_i'\Xi(\Theta)\right). \tag{C.10}$$

[1] Recall that $n_{k+1} = 0$.

Evaluating (C.10) at $\Theta = 0$ gives a first expression for $\mathrm{E}[\mathrm{vec}(C)]$, i.e.,

$$\left.\frac{\partial \varphi_{\mathcal{R}}(\Theta)}{\imath \partial \mathrm{vec}(\Theta)'}\right|_{\Theta=0} = \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2}\right) \mathrm{vec}\left(\Sigma^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma^{\frac{\bar{n}}{2}}\right), \tag{C.11}$$

where $\Sigma^{\frac{\bar{n}}{2}} = (\Sigma^{-\frac{\bar{n}}{2}})^{-1} = T^{\Sigma} D^{\star} T^{\Sigma'}$ with $D^{\star} = \mathrm{diag}\left(\frac{2\sigma_{1\cdot 0}}{n_1}, \ldots, \frac{2\sigma_{k\cdot k-1}}{n_k}\right)$ and

$$\Sigma^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma^{\frac{\bar{n}}{2}} = \begin{bmatrix} T^{\Sigma}_{[i]} D^{\star}_{[i]} T^{\Sigma'}_{[i]} & T^{\Sigma}_{[i]} D^{\star}_{[i]} T^{\Sigma}_{1:i,i+1:k} \\ T^{\Sigma'}_{1:i,i+1:k} D^{\star}_{[i]} T^{\Sigma'}_{[i]} & T^{\Sigma'}_{1:i,i+1:k} D^{\star}_{[i]} T^{\Sigma}_{1:i,i+1:k} \end{bmatrix}, \tag{C.12}$$

resulting from the respective decomposition of $\Sigma^{\frac{\bar{n}}{2}}$. Everything left to do now, is evaluating the sum in (C.11), here exemplarily done for $k = 2$, s.th.

$$\Sigma = T^{\Sigma} D^{\Sigma} T^{\Sigma'} = \begin{bmatrix} 1 & 0 \\ t_1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1\cdot 0} & 0 \\ 0 & \sigma_{2\cdot 1} \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{1\cdot 0} & t_1 \sigma_{1\cdot 0} \\ t_1 \sigma_{1\cdot 0} & t_1^2 \sigma_{1\cdot 0} + \sigma_{2\cdot 1} \end{bmatrix}$$

and $\Sigma^{\frac{\bar{n}}{2}}_{[1]} = T_{[1]} D^{\star}_{[1]} T'_{[1]} = \frac{2}{n_1}\sigma_{1\cdot 0}$ and $\Sigma^{\frac{\bar{n}}{2}}_{1:1,2:k} = T_{[1]} D^{\star}_{[1]} T'_{1:1,2:k} = t_1 \frac{2}{n_1}\sigma_{1\cdot 0}$, respectively. Then

$$\begin{aligned}
&\frac{n_1 - n_2}{2} \mathrm{vec}\left(\Sigma^{\frac{\bar{n}}{2}} E_1 (E_1' \Sigma^{\frac{\bar{n}}{2}} E_1)^{-1} E_1' \Sigma^{\frac{\bar{n}}{2}}\right) + \frac{n_2}{2} \mathrm{vec}\left(\Sigma^{\frac{\bar{n}}{2}} E_2 (E_2' \Sigma^{\frac{\bar{n}}{2}} E_2)^{-1} E_2' \Sigma^{\frac{\bar{n}}{2}}\right) \\
&= \frac{n_1}{2} \mathrm{vec} \begin{bmatrix} \frac{2}{n_1}\sigma_{1\cdot 0} & t_1 \frac{2}{n_1}\sigma_{1\cdot 0} \\ t_1 \frac{2}{n_1}\sigma_{1\cdot 0} & t_1^2 \frac{2}{n_1}\sigma_{1\cdot 0} \end{bmatrix} - \frac{n_2}{2} \mathrm{vec} \begin{bmatrix} \frac{2}{n_1}\sigma_{1\cdot 0} & t_1 \frac{2}{n_1}\sigma_{1\cdot 0} \\ t_1 \frac{2}{n_1}\sigma_{1\cdot 0} & t_1^2 \frac{2}{n_1}\sigma_{1\cdot 0} \end{bmatrix} \\
&\qquad + \frac{n_2}{2} \mathrm{vec} \begin{bmatrix} \frac{2}{n_1}\sigma_{1\cdot 0} & t_1 \frac{2}{n_1}\sigma_{1\cdot 0} \\ t_1 \frac{2}{n_1}\sigma_{1\cdot 0} & t_1^2 \frac{2}{n_1}\sigma_{1\cdot 0} + \frac{2}{n_2}\sigma_{2\cdot 1} \end{bmatrix} \\
&= \mathrm{vec} \begin{bmatrix} \sigma_{1\cdot 0} & t_1 \sigma_{1\cdot 0} \\ t_1 \sigma_{1\cdot 0} & t_1^2 \sigma_{1\cdot 0} + \sigma_{2\cdot 1} \end{bmatrix} \\
&= \mathrm{vec}\left(\Sigma\right).
\end{aligned}$$

This result is readily verified for $k > 2$ and completes the proof.[2]

[2] In order to obtain $\mathrm{E}[C]$ we simply reshape $\mathrm{vec}(\Sigma)$ to the symmetric squared matrix Σ.

C.1.3. Covariance structure

We have

$$\mathrm{Cov}[\mathrm{vec}(C)] = \mathrm{E}[\mathrm{vec}(C)\mathrm{vec}(C)'] - \mathrm{E}[\mathrm{vec}(C)]\mathrm{E}[\mathrm{vec}(C)]',$$

where

$$\mathrm{E}[\mathrm{vec}(C)\mathrm{vec}(C)'] = \left.\frac{\partial^2 \varphi_{\mathcal{R}}(\Theta)}{\iota \partial \mathrm{vec}(\Theta) \partial \mathrm{vec}(\Theta)'}\right|_{\Theta=0}.$$

The Hessian of $\varphi_{\mathcal{R}}(\Theta)$ is obtained by differentiating $\mathrm{d}\varphi_{\mathcal{R}}(\Theta)$ again, i.e.,

$$\begin{aligned}
&\mathrm{d}^2\varphi_{\mathcal{R}}(\Theta) \\
&\quad = \iota \Delta_{\frac{\bar{n}}{2}}(\Sigma^{-\frac{\bar{n}}{2}}) \left[\mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta)) \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) \mathrm{tr}\left(\Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta) \mathrm{d}\Theta \right) \right. \\
&\qquad \left. + \Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta)) \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) \mathrm{dtr}\left(\Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta) \mathrm{d}\Theta \right) \right], \quad \text{(C.13)}
\end{aligned}$$

where $\mathrm{d}\Delta_{\frac{\bar{n}}{2}}(\Xi(\Theta))$ is given by Eq. (C.9), and since $\mathrm{dtr}(A) = \mathrm{tr}(\mathrm{d}A)$

$$\begin{aligned}
\mathrm{dtr}\left(\Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta) \mathrm{d}\Theta \right) &= \mathrm{tr}\left((\mathrm{d}\Xi(\Theta)) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta) \mathrm{d}\Theta \right) \\
&\quad + \mathrm{tr}\left(\Xi(\Theta) E_i (\mathrm{d}(E_i' \Xi(\Theta) E_i)^{-1}) E_i' \Xi(\Theta) \mathrm{d}\Theta \right) \\
&\quad + \mathrm{tr}\left(\Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' (\mathrm{d}\Xi(\Theta)) \mathrm{d}\Theta \right). \quad \text{(C.14)}
\end{aligned}$$

Since $\mathrm{d}\Xi(\Theta) = \iota \Xi(\Theta) \mathrm{d}\Theta \Xi(\Theta)$, by exploiting $\mathrm{d}((F(A))^{-1}) = -(F(A))^{-1} \mathrm{d}F(A) (F(A))^{-1}$ it follows

$$\mathrm{d}(E_i' \Xi(\Theta) E_i)^{-1} = -\iota (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta) (\mathrm{d}\Theta) \Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1}. \quad \text{(C.15)}$$

For clarity, define

$$\Xi_i^*(\Theta) = \Xi(\Theta) E_i (E_i' \Xi(\Theta) E_i)^{-1} E_i' \Xi(\Theta).$$

Now, combining the results from Eq. (C.15) with (C.14) and inserting in (C.9) and (C.14) in (C.13) yields

$$\begin{aligned}
&\mathrm{d}^2\varphi_{\mathcal{R}}(\Theta) = \iota^2\varphi_{\mathcal{R}}(\Theta) \\
&\times\Bigg[\mathrm{vec}(\mathrm{d}\Theta)'\left(\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\Xi_i^*(\Theta))\right)\left(\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\Xi_i^*(\Theta))\right)'\mathrm{vec}(\mathrm{d}\Theta) \\
&+2\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{tr}\left(\Xi_i^*(\Theta)\mathrm{d}\Theta\Xi(\Theta)\mathrm{d}\Theta\right)-\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{tr}\left(\Xi_i^*(\Theta)\mathrm{d}\Theta\Xi_i^*(\Theta)\mathrm{d}\Theta\right)\Bigg],
\end{aligned} \tag{C.16}$$

or equivalently exploiting $\mathrm{tr}(ABCD) = \mathrm{vec}(B')'(C'\otimes A)\mathrm{vec}(D)$ (see Lütkepohl, 1996, Eq. 4.1.1.9) and for symmetric matrices $\mathrm{vec}(A) = \frac{1}{2}(I_{k^2}+K_{kk})\mathrm{vec}(A)$ (see Díaz-García, 2013, Eq. 14)

$$\begin{aligned}
&\mathrm{d}^2\varphi_{\mathcal{R}}(\Theta) = \iota^2\varphi_{\mathcal{R}}(\Theta) \\
&\times\Bigg[\mathrm{vec}(\mathrm{d}\Theta)'\left(\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\Xi_i^*(\Theta))\right)\left(\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\Xi_i^*(\Theta))\right)'\mathrm{vec}(\mathrm{d}\Theta) \\
&+\frac{1}{2}\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\mathrm{d}\Theta)'(I_{k^2}+K_{kk})\left(\Xi_i^*(\Theta)\otimes\Xi(\Theta)\right)(I_{k^2}+K_{kk})\mathrm{vec}(\mathrm{d}\Theta) \\
&-\frac{1}{4}\sum_{i=1}^{k}\left(\frac{n_i-n_{i+1}}{2}\right)\mathrm{vec}(\mathrm{d}\Theta)'(I_{k^2}+K_{kk})\left(\Xi_i^*(\Theta)\otimes\Xi_i^*(\Theta)\right)(I_{k^2}+K_{kk})\mathrm{vec}(\mathrm{d}\Theta)\Bigg],
\end{aligned} \tag{C.17}$$

since $(I_{k^2}+K_{kk})' = (I_{k^2}+K_{kk})$ and $(I_{k^2}+K_{kk})(A\otimes B)(I_{k^2}+K_{kk}) = (I_{k^2}+K_{kk})(B\otimes A)(I_{k^2}+K_{kk})$ (see Lütkepohl, 1996, Eqs. 9.2.1.16 & 9.2.2.7). Now, applying

result 7(b) in Díaz-García (2013) yields

$$\begin{aligned}
&\frac{\partial^2 \varphi_{\mathcal{R}}(\Theta)}{\partial \mathrm{vec}(\Theta) \partial \mathrm{vec}(\Theta)'} \\
&\quad = \iota^2 \varphi_{\mathcal{R}}(\Theta) \left[\left(\sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) \mathrm{vec}(\Xi_i^*(\Theta)) \right) \left(\sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) \mathrm{vec}(\Xi_i^*(\Theta)) \right)' \right. \\
&\qquad + \frac{1}{2} \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) (I_{k^2} + K_{kk}) \left(\Xi_i^*(\Theta) \otimes \Xi(\Theta) \right) (I_{k^2} + K_{kk}) \\
&\qquad \left. - \frac{1}{4} \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) (I_{k^2} + K_{kk}) \left(\Xi_i^*(\Theta) \otimes \Xi_i^*(\Theta) \right) (I_{k^2} + K_{kk}) \right]. \qquad \text{(C.18)}
\end{aligned}$$

Evaluating (C.18) at $\Theta = 0$ and dividing by ι^2 gives a first expression for $\mathrm{E}[\mathrm{vec}(C)\mathrm{vec}(C)']$, i.e.,

$$\begin{aligned}
&\left. \frac{\partial^2 \varphi_{\mathcal{R}}(\Theta)}{\iota^2 \partial \mathrm{vec}(\Theta) \partial \mathrm{vec}(\Theta)'} \right|_{\Theta=0} \\
&= \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)' \\
&\quad + \frac{1}{2} \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) (I_{k^2} + K_{kk}) \left((\Sigma^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma^{\frac{\bar{n}}{2}}) \otimes \Sigma^{\frac{\bar{n}}{2}} \right) (I_{k^2} + K_{kk}) \\
&\quad - \frac{1}{4} \sum_{i=1}^{k} \left(\frac{n_i - n_{i+1}}{2} \right) (I_{k^2} + K_{kk}) \\
&\qquad \left((\Sigma^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma^{\frac{\bar{n}}{2}}) \otimes (\Sigma^{\frac{\bar{n}}{2}} E_i (E_i' \Sigma^{\frac{\bar{n}}{2}} E_i)^{-1} E_i' \Sigma^{\frac{\bar{n}}{2}}) \right) (I_{k^2} + K_{kk}). \qquad \text{(C.19)}
\end{aligned}$$

Finally, subtracting $\mathrm{E}[\mathrm{vec}(C)]\mathrm{E}[\mathrm{vec}(C)]' = \mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)'$ (see the intermediary result in Eq. (C.11)) from Eq. (C.19) and recalling that $E_k = I_k$ and $n_{k+1} = 0$, as well as making use of $(I_{k^2} + K_{kk})(A \otimes A)(I_{k^2} + K_{kk}) = 2(I_{k^2} + K_{kk})(A \otimes A)$ (see Lütkepohl, 1996, Eq. 9.2.2.2.a) completes the proof.

C.2. Proof of Proposition 3

Recall the Riesz pdf in Eq. (4.11). We first decompose the trace in the exponent. We therefore rewrite C as follows

$$C = \begin{bmatrix} C_{[k-1]} & c'_{k,1:k-1} \\ c_{k,1:k-1} & c_{k,k} \end{bmatrix} = \begin{bmatrix} C_{[k-1]} & C_{[k-1]}b'_{k\cdot k-1} \\ b_{k\cdot k-1}C_{[k-1]} & b_{k\cdot k-1}C_{[k-1]}b'_{k\cdot k-1} + c_{k\cdot k-1} \end{bmatrix},$$

and we Cholesky decompose $\Sigma = T^{\Sigma}D^{\Sigma}T^{\Sigma\prime} = V^{-1}D^{\Sigma}V'^{-1}$, s.th.

$$\Sigma = \begin{bmatrix} \Sigma_{[k-1]} & \sigma'_{k,1:k-1} \\ \sigma_{k,1:k-1} & \sigma_{k,k} \end{bmatrix} = \begin{bmatrix} V_{[k-1]} & 0 \\ -\beta_{k\cdot k-1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} D^{\Sigma}_{[k-1]} & 0 \\ 0 & \sigma_{k\cdot k-1} \end{bmatrix} \begin{bmatrix} V'_{[k-1]} & -\beta'_{k\cdot k-1} \\ 0 & 1 \end{bmatrix}^{-1}.$$

Hence, we get the following expression for the generalized inverse

$$\begin{aligned} \Sigma^{-\frac{\bar{n}}{2}} &= \begin{bmatrix} V'_{[k-1]} & -\beta'_{k\cdot k-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D^{\star-1}_{[k-1]} & 0 \\ 0 & \frac{n_k}{2\sigma_{k\cdot k-1}} \end{bmatrix} \begin{bmatrix} V_{[k-1]} & 0 \\ -\beta_{k\cdot k-1} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{[k-1]}^{-\frac{\bar{n}_{1:k-1}}{2}} + \frac{n_k}{2\sigma_{k\cdot k-1}}\beta'_{k\cdot k-1}\beta_{k\cdot k-1} & -\frac{n_k}{2\sigma_{k\cdot k-1}}\beta'_{k\cdot k-1} \\ -\frac{n_k}{2\sigma_{k\cdot k-1}}\beta_{k\cdot k-1} & \frac{n_k}{2\sigma_{k\cdot k-1}} \end{bmatrix}, \end{aligned}$$

where $D^{\star-1}_{[k-1]} = \text{diag}\left(n_1/(2\sigma_{1\cdot 0}), \ldots, n_{k-1}/(2\sigma_{k-1\cdot k-2})\right)$ and $\Sigma_{[k-1]}^{-\frac{\bar{n}_{1:k-1}}{2}} = V'_{[k-1]}D^{\star-1}_{[k-1]}V_{[k-1]}$ denotes the generalized inverse for the $(k-1)$-dimensional upper left block of Σ with

respect to $\bar{n}_{1:k-1} = (n_1, \ldots, n_{k-1})'$. Substituting the latter results into the trace yields

$$
\begin{aligned}
&\operatorname{tr}\left(\Sigma^{-\frac{\bar{n}}{2}} C\right) \\
&\quad = \operatorname{tr}\left(\Sigma_{[k-1]}^{-\frac{\bar{n}_{1:k-1}}{2}} C_{[k-1]}\right) + \operatorname{tr}\left(\frac{n_k}{2\sigma_{k\cdot k-1}} \beta'_{k\cdot k-1}\beta_{k\cdot k-1} C_{[k-1]}\right) - \operatorname{tr}\left(\frac{n_k}{2\sigma_{k\cdot k-1}} \beta'_{k\cdot k-1} b_{k\cdot k-1} C_{[k-1]}\right) \\
&\qquad - \operatorname{tr}\left(\frac{n_k}{2\sigma_{k\cdot k-1}} b'_{k\cdot k-1}\beta_{k\cdot k-1} C_{[k-1]}\right) + \frac{n_k}{2\sigma_{k\cdot k-1}} b_{k\cdot k-1} C_{[k-1]} b'_{k\cdot k-1} + \frac{n_k}{2\sigma_{k\cdot k-1}} c_{k\cdot k-1} \\
&\quad = \operatorname{tr}\left(\Sigma_{[k-1]}^{-\frac{\bar{n}_{1:k-1}}{2}} C_{[k-1]}\right) + \frac{n_k}{2\sigma_{k\cdot k-1}} \beta_{k\cdot k-1} C_{[k-1]} \beta'_{k\cdot k-1} - \frac{n_k}{2\sigma_{k\cdot k-1}} b_{k\cdot k-1} C_{[k-1]} \beta'_{k\cdot k-1} \\
&\qquad - \frac{n_k}{2\sigma_{k\cdot k-1}} \beta_{k\cdot k-1} C_{[k-1]} b'_{k\cdot k-1} + \frac{n_k}{2\sigma_{k\cdot k-1}} b_{k\cdot k-1} C_{[k-1]} b'_{k\cdot k-1} + \frac{n_k}{2\sigma_{k\cdot k-1}} c_{k\cdot k-1} \\
&\quad = \operatorname{tr}\left(\Sigma_{[k-1]}^{-\frac{\bar{n}_{1:k-1}}{2}} C_{[k-1]}\right) + \frac{1}{2}(b_{k\cdot k-1} - \beta_{k\cdot k-1}) \frac{n_k}{\sigma_{k\cdot k-1}} C_{[k-1]} (b_{k\cdot k-1} - \beta_{k\cdot k-1})' + \frac{n_k}{2\sigma_{k\cdot k-1}} c_{k\cdot k-1} \\
&\quad = \sum_{i=1}^{k} \frac{n_i}{2\sigma_{i\cdot i-1}} c_{i\cdot i-1} + \frac{1}{2} \sum_{i=2}^{k} (b_{i\cdot i-1} - \beta_{i\cdot i-1}) \frac{n_i}{\sigma_{i\cdot i-1}} C_{[i-1]} (b_{i\cdot i-1} - \beta_{i\cdot i-1})',
\end{aligned}
$$

where the last line follows by induction.

Now, we have to rearrange the expression in front of the exponential function, i.e.

$$
\begin{aligned}
\pi^{-\frac{k(k-1)}{4}} \prod_{i=1}^{k} \frac{\left(\frac{n_i}{2}\right)^{\frac{n_i}{2}} \left[c_{i\cdot i-1}\right]^{\frac{n_i-k-1}{2}}}{\Gamma\left(\frac{n_i-i+1}{2}\right) \left[\sigma_{i\cdot i-1}\right]^{\frac{n_i}{2}}} &= \prod_{i=1}^{k} \frac{\pi^{-\frac{i-1}{2}} \left[c_{i\cdot i-1}\right]^{\frac{n_i-k-1}{2}}}{\Gamma\left(\frac{n_i-i+1}{2}\right) \left[\frac{2}{n_i}\sigma_{i\cdot i-1}\right]^{\frac{n_i}{2}}} \\
&= \prod_{i=1}^{k} \frac{\pi^{-\frac{i-1}{2}} \left[c_{i\cdot i-1}\right]^{\frac{n_i-i+1}{2}-1} \left[c_{i\cdot i-1}\right]^{-\frac{k-i}{2}}}{\Gamma\left(\frac{n_i-i+1}{2}\right) \left[\frac{2}{n_i}\sigma_{i\cdot i-1}\right]^{\frac{n_i-i+1}{2}} \left[\frac{2}{n_i}\sigma_{i\cdot i-1}\right]^{\frac{i-1}{2}}} \\
&= \prod_{i=1}^{k} \frac{\left[c_{i\cdot i-1}\right]^{\frac{n_i-i+1}{2}-1}}{\Gamma\left(\frac{n_i-i+1}{2}\right) \left[\frac{2}{n_i}\sigma_{i\cdot i-1}\right]^{\frac{n_i-i+1}{2}}} \frac{(2\pi)^{-\frac{i-1}{2}} |C_{[i-1]}^{-1}|^{\frac{1}{2}}}{\left[\frac{\sigma_{i\cdot i-1}}{n_i}\right]^{\frac{i-1}{2}}}.
\end{aligned}
$$

Finally, taking into account the Jacobian $\prod_{i=1}^{k} \mathcal{J}(c_{i,1:i-1} \to b_{i\cdot i-1}) = \prod_{i=1}^{k} |C_{[i-1]}^{-1}|^{-1}$ the proof is completed.

C.3. Details on the MCMC Algorithm

C.3.1. Prior assumptions

The prior assumptions we use for the parameters are fairly uninformative. For the $(\gamma, \phi_1, \phi_2, \phi_3, \nu^2)$ parameters in each of the Gaussian HAR processes $\{\log \sigma_{i\cdot i-1,t} =$

$x_{it}\}$ we assume independent conjugate Normal-inverted-Gamma priors with hyper-parameters selected such that $\mathrm{E}(\nu^2) = 0.2$ and $\mathrm{Var}(\nu^2) = 0.0156$, $\mathrm{E}(\phi_1, \phi_2, \phi_3) = (0.3, 0.3, 0.3)$ and $\mathrm{Cov}(\phi_1, \phi_2, \phi_3) = \mathrm{diag}(0.1, 0.1, 0.1)$, $\mathrm{E}(\gamma^*) = 0$ and $\mathrm{Var}(\gamma^*) = 20$, where $\gamma^* = \gamma(1 - \phi_1 - \phi_2 - \phi_3)$. In the basic RSS model we apply the standard conditional conjugate priors $\tau_{ji}^2 \sim \mathcal{IG}(s_0, S_0)$ with $s_0 = 0.1$ and $S_0 = 0.001$, and $\bar{\beta}_{ji} \sim \mathcal{N}(0, A_0)$ with $A_0 = 10$. The prior hyper-parameters for the shrinkage priors are set to $d_1 = 10$ and $d_2 = d_3 = 0.001$, as proposed by Bitto and Frühwirth-Schnatter (2019). The prior for p_{ji0} is chosen as $\mathcal{IG}(v_p, (v_p - 1)c_p)$, with $v_p = 20$ and $c_p = 1$, s.th. no prior moment exists.

C.3.2. Sampling of $\beta_{ji,0:T}$

Step 1.) of the proposed MCMC sampling scheme in Section 4.4.1 samples the latent states $\beta_{ji,0:T}$ conditional on static parameters and $\sigma_{i\cdot i-1,1:T}$ using the *precision sampler*, as discussed in Chan and Jeliazkov (2009).[3] Subsequently, we provide details how the precision sampling algorithm is implemented.

The implementation exploits the non-centered and adjusted parameterization of the state-space model for $b_{i\cdot i-1,t}$ as stated in Eq. (4.23). Recall the measurement and transition equations given by

$$b^{\star}_{ji,t} = \tau_{ji}\tilde{\beta}_{ji,t} + \tilde{\varepsilon}_{ji,t}, \qquad \tilde{\varepsilon}_{ji,t} \sim \mathcal{N}(0, \varpi_{ji,t}), \tag{C.20}$$

$$\tilde{\beta}_{ji,t} = \tilde{\beta}_{ji,t-1} + \tilde{\zeta}_{ji,t}, \qquad \tilde{\zeta}_{ji,t} \sim \mathcal{N}(0, 1), \tag{C.21}$$

with initial condition $\tilde{\beta}_{ji,0} \sim \mathcal{N}(0, \tilde{p}_{ji0})$. Here, $b^{\star}_{ji,t} = b_{ji,t} - \bar{\beta}_{ji} - \sum_{l=1}^{j-1} h_{lj,t}\tilde{\varepsilon}_{li,t}$ and $\varpi_{ji,t} = \sigma_{i\cdot i-1,t}\delta_{j\cdot j-1,t}/n_i$, for $j = 1, \ldots, i-1$ and $t = 1, \ldots, T$. The components $h_{lj,t}$ and $\delta_{j\cdot j-1,t}$ (uniquely found from the LDL-decomposition of $C^{-1}_{[i-1]t} = H_t\Delta_t H_t'$) are readily pre-calculated once before running the MCMC algorithm.

Conditional on all other variables, the joint posterior for the state process $\tilde{\beta}_{ji,0:T}$ is

[3] Note, that the precision sampler is equivalent to the *all without a loop* sampler used by Bitto and Frühwirth-Schnatter (2019).

multivariate Gaussian, i.e.,

$$\tilde{\beta}_{ji,0:T} \mid b_{i\cdot i-1,1:T}, \{\tilde{\beta}_{li,1:T}\}_{l=1}^{j-1}, \bar{\beta}_{ji}, \tau_{ji}, \tilde{p}_{ji0} \sim \mathcal{N}_{T+1}(\Pi_\star^{-1} m_\star, \Pi_\star^{-1}), \tag{C.22}$$

where

$$\Pi_\star^{-1} = K' \Pi_0^{-1} K + \Pi_1^{-1}, \qquad m_\star = \tau_{ji}(0, b^\star_{ji,1}/\varpi_{ji,1}, \cdots, b^\star_{ji,T}/\varpi_{ji,T})', \tag{C.23}$$

with $\Pi_0^{-1} = \text{diag}(1/\tilde{p}_{ji0}, 1, \ldots, 1)$, $\Pi_1^{-1} = \text{diag}(0, \tau_{ji}^2/\varpi_{ji,1}, \ldots, \tau_{ji}^2/\varpi_{ji,T})$, and K is a $(T+1) \times (T+1)$ banded matrix with ones on its main diagonal and -1 on its second lower diagonal, i.e.,

$$K = \begin{bmatrix} 1 & 0 & & & \\ -1 & 1 & \ddots & & \\ 0 & -1 & \ddots & & \\ & \ddots & \ddots & 1 & 0 \\ & & & -1 & 1 \end{bmatrix}, \quad \text{s.th } K'\Pi_0^{-1}K = \begin{bmatrix} \tilde{p}_{ji0}^{-1} + 1 & -1 & & & \\ -1 & 2 & \ddots & & \\ 0 & -1 & \ddots & & \\ & \ddots & \ddots & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}.$$

The posterior in Eq. (C.22) is $(T+1)$-dimensional and sampling involves the inversion of $\Pi_\star$. However, as all matrices involved in the precision sampler are either sparse diagonal, or sparse banded, $\Pi_\star$ results as a sparse banded matrix, too. This implies that its Cholesky decomposition $\Pi_\star = P_\star P_\star'$ is computationally inexpensive. Efficient sampling from the posterior can be conducted by first drawing $e \sim \mathcal{N}(0, I_{T+1})$ and then consecutively solving $P_\star \hat{m}_\star = m_\star$ for $\hat{m}_\star$ and $P_\star' \tilde{\beta} = \hat{m}_\star + e$ for $\tilde{\beta} = (\tilde{\beta}_{ji,0}, \ldots, \tilde{\beta}_{ji,T})$, by making use of back-band substitution instead of calculating $P_\star^{-1}$ (see Chan and Jeliazkov, 2009; McCausland et al., 2011; and the Webappendix of Bitto and Frühwirth-Schnatter, 2019).

Finally, after sampling $\tilde{\beta}_{ji,t}$, we obtain a draw for $\beta_{ji,t}$ by the transformation $\beta_{ji,t} = \bar{\beta}_{ji} + \tau_{ji}\tilde{\beta}_{ji,t}$, $t = 0, \ldots, T$.

C.3.3. Sampling $\bar{\beta}$, τ and corresponding shrinkage prior hyperparameters

The MCMC sampling scheme proposed by Bitto and Frühwirth-Schnatter (2019) for the parameters of the latent beta processes consists of full conditional updating steps for $\bar{\beta}_{ji}$ and τ_{ji} (and optionally an interweaving step), their respective prior variances $s^{\bullet}_{ji}$ and the hyper-parameters $a^{\bullet}_i$, $b^{\bullet}_i$, $\bullet \in \{\beta, \tau\}$. The sampler cycles through the following steps, where steps ii) to iv) are done consecutively for $\bullet \in \{\beta, \tau\}$:

i) Conditional on $\tilde{\beta}_{i\cdot i-1,0:T}$, $x_{i,1:T}$, n_i and the prior variances $\{s^{\beta}_{ji}\}$ and $\{s^{\tau}_{ji}\}$, joint updating of $\{\bar{\beta}_{ji}\}$ and $\{\tau_{ji}\}$ is done by sampling $\alpha_i = (\bar{\beta}_{1i}, \ldots, \bar{\beta}_{i-1i}, \tau_{1i}, \ldots, \tau_{i-1i})'$ from

$$\alpha_i \sim \mathcal{N}_{2(i-1)}(m_1, P_1^{-1}) \tag{C.24}$$

with

$$P_1^{-1} = P_0^{-\frac{1}{2}} \left(I_{2(i-1)} + P_0^{-\frac{1}{2}} X' \Omega X P_0^{-\frac{1}{2}} \right)^{-1} P_0^{-\frac{1}{2}}, \qquad m_1 = P_1^{-1} X' \Omega b, \tag{C.25}$$

where $P_0^{-1} = \text{diag}(s^{\beta}_{1i}, \ldots, s^{\beta}_{i-1i}, s^{\tau}_{1i}, \ldots, s^{\tau}_{i-1i})$, $b = (b'_{i\cdot i-1,1}, \ldots, b'_{i\cdot i-1,T})'$ is a $T(i-1)\times 1$ vector with stacked realized beta coefficients, X is a sparse $T(i-1)\times 2(i-1)$ matrix which results from stacking $X_t = [I_{i-1}, \text{diag}(\tilde{\beta}_{i\cdot i-1,t})]$ for $t = 1, \ldots, T$, and $\Omega = \text{diag}\left(\frac{n_i}{\sigma_{i\cdot i-1,1}} C_{[i-1]1}, \ldots, \frac{n_i}{\sigma_{i\cdot i-1,T}} C_{[i-1]T} \right)$ is a sparse $T(i-1) \times T(i-1)$ block diagonal covariance matrix.[4]

i*) *Interweaving step:* After a draw for $(\bar{\beta}_{1i}, \ldots, \bar{\beta}_{i-1i}, \tau_{1i}, \ldots, \tau_{i-1i})'$ is obtained, each pair $(\bar{\beta}_{ji}, \tau_{ji})$, $j = 1, \ldots, i-1$, is redrawn through interweaving into the state equation of the centered parameterization. Therefore, we update $\beta_{ji,0:T}$ via the deterministic transformation $\beta_{ji,t} = \bar{\beta}_{ji} + \tau_{ji}\tilde{\beta}_{ji,t}$, for $t = 0, \ldots, T$, and

[4]Note, that the matrix Ω can be expressed as the product of two sparse (block)diagonal matrices Ω_C and Ω_σ, where $\Omega_C = \text{diag}(C_{[i-1]1}, \ldots, C_{[i-1]T})$ and $\Omega_\sigma = [\text{diag}(n_i/\sigma_{i\cdot i-1,1}, \ldots, n_i/\sigma_{i\cdot i-1,T},) \otimes I_{i-1}]$. The data-dependent covariance part Ω_C is readily prepared once before running the MCMC algorithm.

sample from

$$\tau_{ji}^2 \mid \beta_{ji,0:T}, \bar{\beta}_{ji,}, s_{ji}^{\tau}, \tilde{p}_{ji0} \sim \mathcal{GIG}\left(-\frac{T}{2}, \frac{1}{s_{ji}^{\tau}}, \sum_{t=1}^{T}(\beta_{ji,t} - \beta_{ji,t-1})^2 + \frac{(\beta_{ji,0} - \bar{\beta}_{ji})^2}{\tilde{p}_{ji0}}\right),$$

$$\bar{\beta}_{ji} \mid \beta_{ji,0}, \tau_{ji}, s_{ji}^{\beta}, \tilde{p}_{ji0} \sim \mathcal{N}\left(\frac{s_{ji}^{\beta}\beta_{ji,0}}{s_{ji}^{\beta} + \tau_{ji}^2\tilde{p}_{ji0}}, \frac{s_{ji}^{\beta}\tau_{ji}^2\tilde{p}_{ji0}}{s_{ji}^{\beta} + \tau_{ji}^2\tilde{p}_{ji0}}\right).$$

The new draws $(\bar{\beta}_{1i}, \ldots, \bar{\beta}_{i-1i}, \tau_{1i}, \ldots, \tau_{i-1i})'$ replace those obtained in step i). Note, that after taking the square root of τ_{ji}^2, the sign of the old τ_{ji} value is assigned to the new draw. Finally, $\tilde{\beta}_{ji,0:T}$ is updated via $\tilde{\beta}_{ji,t} = (\beta_{ji,t} - \bar{\beta}_{ji})/\tau_{ji}$, for $t = 0, \ldots, T$.

ii) Full conditional sampling of $a_i^{\bullet}$ is conducted using a random walk Metropolis-Hastings step based on proposing $\log(a_i^{\bullet,new}) \sim \mathcal{N}(\log(a_i^{\bullet}), d_4)$ and accepting $a_i^{\bullet,new}$ as the new draw with probability

$$\min\left\{\frac{\exp(-d_1 a_i^{\bullet,new})a_i^{\bullet,new}}{\exp(-d_1 a_i^{\bullet})a_i^{\bullet}} \prod_{j=1}^{i-1} \frac{p(\bullet_{ji}|a_i^{\bullet,new}, b_i^{\bullet})}{p(\bullet_{ji}|a_i^{\bullet}, b_i^{\bullet})},\ 1\right\}, \tag{C.26}$$

where

$$\begin{aligned} p(\bullet_{ji}|a_i^{\bullet}, b_i^{\bullet}) &= \int p(\bullet_{ji}|s_{ji}^{\bullet})p(s_{ji}^{\bullet}|a_i^{\bullet}, b_i^{\bullet})\mathrm{d}s_{ji}^{\bullet} \\ &= \frac{(\sqrt{a_i^{\bullet}b_i^{\bullet}})^{a_i^{\bullet}+1/2}}{\sqrt{\pi}2^{a_i^{\bullet}-1/2}\Gamma(a_i^{\bullet})} \mid \bullet_{ji} \mid^{a_i^{\bullet}-1/2} K_{a_i^{\bullet}-1/2}(\mid \bullet_{ji} \mid \sqrt{a_i^{\bullet}b_i^{\bullet}}), \end{aligned}$$

and $K_{a_i^{\bullet}-1/2}(\cdot)$ is the modified Bessel function of the second kind with index $(a_i^{\bullet} - 1/2)$.[5]

iii) For the Normal-Gamma hierarchical priors, it follows that the conditionally Normal priors $\bullet_{ij}|s_{ji}^{\bullet}$ lead to posteriors for the variances $s_{ji}^{\bullet}$ where the respective likelihoods are the kernel of an inverse Gamma density in $s_{ji}^{\bullet}$. In combination with the Gamma priors $s_{ji}^{\bullet} \mid a_i^{\bullet}, b_i^{\bullet}$ this leads to Generalized Inverse Gaussian

[5] $K_{a_i^{\bullet}-1/2}(\cdot)$ is e.g. implemented in MATLAB in the function `besselk`.

distributions

$$s_{ji}^{\bullet} \mid \bullet_{ji}, a_i^{\bullet}, b_i^{\bullet} \sim \mathcal{GIG}\left(a_i^{\bullet} - \frac{1}{2}, a_i^{\bullet} b_i^{\bullet}, (\bullet_{ji})^2\right). \tag{C.27}$$

iv) Finally, full conditional updating of $b_i^{\bullet}$ is done by sampling

$$b_i^{\bullet} \mid a_i^{\bullet}, \{s_{ji}^{\bullet}\}_{j=1}^{i-1} \sim \mathcal{G}\left(d_2 + (i-1)a_i^{\bullet}, d_3 + \frac{a_i^{\bullet}}{2}\sum_{j=1}^{i-1} s_{ji}^{\bullet}\right), \tag{C.28}$$

where $\mathcal{G}$ denotes the Gamma distribution in shape-rate parameterization.

Efficient sampling from the GIG distribution in steps i*) and iii) is done based on the random number generator proposed by Hörmann and Leydold (2014). Hartkopf (2020) translated their R-function `GIGrvg` (see Hörmann and Leydold, 2015) to MATLAB and included proper handling of limiting cases (see also Appendix C.6).

C.3.4. Sampling of the scale parameters $\tilde{p}_0$

For $\{\tilde{p}_{ji0}\}$ we impose independent conjugate inverse gamma priors. Samples from the corresponding full conditional posterior distributions are obtained by sampling

$$\tilde{p}_{ji0} \mid \tilde{\beta}_{ji,0} \sim \mathcal{IG}\left(v_p + \frac{1}{2}, (v_p - 1)c_p + \frac{1}{2}\tilde{\beta}_{ji,0}^2\right), \tag{C.29}$$

for $j = 1, \ldots, i-1$.

C.3.5. Sampling of $\log \sigma_{i\cdot i-1,1:T}$

In step 2.) of the proposed MCMC sampling scheme, we sample the latent state series $\{x_{i,1:T}\}$ full conditional on the static model parameters $(\gamma_i, \phi_{i1}, \phi_{i2}, \phi_{i3}, \nu_i^2)$ and latent states $\{\beta_{i\cdot i-1,1:T}\}$ using particle Gibbs with ancestor sampling (PG-AS, Andrieu et al., 2010; Lindsten et al., 2014). A full description for the PG-AS algorithm is already given in Appendix A.1.1. Subsequently, we provide details on how the PG-AS algorithm is implemented for the RSS model.

The implementation of the PG-AS algorithm relies on a vector representation of the latent HAR process. Therefore, we define $Z_{it} = (x_{it}, x_{it-1}, \ldots, x_{it-21})'$ and we re-write the state equation in (4.18) as follows

$$Z_{it} = \Gamma_i + \Phi_i Z_{it-1} + F\eta_{it}, \qquad \eta_{it} \sim \mathcal{N}(0, \nu_i^2), \tag{C.30}$$

where $F = (1, 0_{21}')'$, 0_d is a $d \times 1$ vector of zeros, and

$$\Gamma_i = \begin{pmatrix} \gamma_i(1 - \phi_{1i} - \phi_{2i} - \phi_{3i}) \\ 0_{21} \end{pmatrix}, \qquad \Phi_i = \begin{bmatrix} \phi_i^\star \\ I_{21} \mid 0_{21} \end{bmatrix},$$

with $\phi_i^\star = (\phi_{i1}^\star, \phi_{i2}^\star \iota_4', \phi_{i3}^\star \iota_{17}')$, $\phi_{i1}^\star = \phi_{i1} + \frac{1}{5}\phi_{i2} + \frac{1}{22}\phi_{i3}$, $\phi_{i2}^\star = \frac{1}{5}\phi_{i2} + \frac{1}{22}\phi_{i3}$, $\phi_{i3}^\star = \frac{1}{22}\phi_{i3}$, and ι_d is a $d \times 1$ vector of ones.

Now, the PG-AS algorithm proceeds in the same steps as described in Appendix A.1.1 to produce a trajectory $Z_{i,1:T}$. For the RSS model, the period-t weights are calculated as

$$w_{it}^{(j)} = \exp\left\{-\frac{n_i}{2}\left(F' Z_{it}^{(j)} + e^{-F' Z_{it}^{(j)}} y_{it}\right)\right\}, \tag{C.31}$$

with $y_{it} = c_{i\cdot i-1,t} + (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})' C_{[i-1],t} (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})$. A sample for $x_{i,1:T}$ is obtained as $x_{it} = F' Z_{it}$, for $t = 1, \ldots, T$.

C.3.6. Sampling of HAR parameters γ, ϕ_1, ϕ_2, ϕ_3, ν^2

See Appendix A.1.3 for a description of how to obtain a sample from the posterior distribution of the parameters γ_i, ϕ_{i1}, ϕ_{i2}, ϕ_{i3} and ν_i^2, for $i = 1, \ldots, k$.

C.3.7. Sampling of d.o.f. parameters n

Draws from the posterior distributions of the k Riesz degrees of freedom parameters are obtained by the Metropolis-Hastings (MH) algorithm. Therefore we select independent uniform priors $p(n_i)$ on the grid $i - 1 < n_i < \infty$. The (indepen-

dent) target density for each n_i, $i = 1, \ldots, k$, is then given by $\pi(n_i \mid \cdot) \equiv \pi(n_i \mid \{x_{i,1:T}\}, \{\beta_{i\cdot i-1,1:T}\}, C_{1:T})$, i.e.,

$$\pi(n_i \mid \cdot) \propto \left[\frac{(n_i/2)^{n_i/2}}{\Gamma\left(\frac{n_i-i+1}{2}\right)}\right]^T \left(\prod_{t=1}^{T} y_{it}^c\right)^{n_i/2} \exp\left\{-\frac{n_i}{2}\sum_{t=1}^{T}(y_{it}^c + y_{it}^b)\right\}, \tag{C.32}$$

with $y_{it}^c = c_{i\cdot i-1,t}/\sigma_{i\cdot i-1,t}$, and $y_{it}^b = (b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})' C_{[i-1],t}(b_{i\cdot i-1,t} - \beta_{i\cdot i-1,t})/\sigma_{i\cdot i-1,t}$.

The proposal density for the MH step is chosen to be a tailored t distribution with ξ degrees of freedom, mean m_{n_i} and variance $v_{n_i}^2$, say $f_T(n_i|\xi, m_{n_i}, v_{n_i}^2)$. We find m_{n_i} and $v_{n_i}^2$ by numerical optimization of the logarithmic target density as

$$m_{n_i} = \underset{n_i}{\text{argmax}}\ \log \pi(n_i \mid \cdot), \qquad v_{n_i}^2 = \{-\partial^2 \log \pi(n_i \mid \cdot)/\partial n_i^2\}^{-1}. \tag{C.33}$$

The MH step is completed by proposing the value $n_i' > i-1$ from $f_T(n_i|\xi, m_{n_i}, v_{n_i}^2)$ and accepting it with probability

$$\alpha(n_i, n_i') = \min\left\{\frac{\pi(n_i' \mid \cdot) f_T(n_i|\xi, m_{n_i}, v_{n_i}^2)}{\pi(n_i \mid \cdot) f_T(n_i'|\xi, m_{n_i}, v_{n_i}^2)}, 1\right\}. \tag{C.34}$$

If the proposal value is rejected, the current value n_i is retained as the next draw. Numerical optimization of the target density is readily speeded up by pre-calculating the terms depending on y_{it}^c and y_{it}^b.

C.4. RBPF Approximation of the Likelihood in Equation (4.29)

The Deviance Information Criteria (DIC) calculations in Eq. (4.27) are based on the likelihood in Eq. (4.29). We evaluate the integral in Eq. (4.29) using a Rao-Blackwellized version of the Bootstrap particle filter (RBPF) algorithm of Gordon et al. (1993). Based on the factorization in Eq. (4.30), the RBPF uses Bayes' theorem to marginalize out the linear state variables $\beta_{i\cdot i-1,1:T}$ based on standard Kalman filter

recursions and then applies a standard BPF to the simplified integrals

$$p_i(C_{1:T}) \propto \int \left[\prod_{t=1}^{T} f(c_{i\cdot i-1,t}, b_{i\cdot i-1,t} \mid C_{[i-1],1:t}, c_{i\cdot i-1,1:t-1}, b_{i\cdot i-1,1:t-1}, x_{i,1:t}, \theta) \times f(x_{i,t}|x_{i,1:t-1}\theta) \right] \mathrm{d}x_{i,1:T}, \tag{C.35}$$

for $i = 1, \ldots, k$.

The RBPF algorithm proceeds in the same steps as described in Appendix A.2 to estimate $p_i(C_{1:T}) = p_i(C_1) \prod_{t=2}^{T} p_i(C_t|C_{1:t-1})$. Here, the period-$t$ importance weight (ignoring the factor $K_t(n_i, c_{i\cdot i-1,t}) = (n_i/2)^{n_i/2} \left[\Gamma((n-i+1)/2)\right]^{-1} c_{i\cdot i-1,t}^{(n_i-k-1)/2}$) is given by

$$w_{it}^{(j)} = \frac{|V_{it}^{(j)}|^{1/2}}{|V_{it}^{\star(j)}|^{1/2}} \exp\left\{ -\frac{1}{2} \left[n_i \left(x_{it}^{(j)} + c_{ii,t} e^{-x_{it}^{(j)}} \right) - m_{it}^{(j)\prime} V_{it}^{(j)} m_{it}^{(j)} + m_{it}^{\star(j)\prime} V_{it}^{\star(j)} m_{it}^{\star(j)} \right] \right\}, \tag{C.36}$$

where $m_{it}^{\star}, V_{it}^{\star}$ and m_{it}, V_{it} denote the mean and variance of the predictive and filtering distributions for $\beta_{i\cdot i-1,t}$, found through the Kalman filter recursions. Given the initialization $m_0 = (\bar{\beta}_{1i}, \ldots, \bar{\beta}_{i-1i})'$ and $V_0 = \mathrm{diag}(p_{1i0}, \ldots, p_{i-1i0})$, the moments are updated via

$$m_{it}^{\star(j)} = m_{it-1}^{(j)}, \tag{C.37}$$

$$V_{it}^{\star(j)} = V_{it-1}^{(j)} + \mathrm{diag}(\tau_{1i}^2, \ldots, \tau_{i-1i}^2), \tag{C.38}$$

$$m_{it}^{(j)} = V_{it}^{(j)} \left((V_{it}^{\star(j)})^{-1} m_{it}^{\star(j)} + n_i e^{-x_{it}^{(j)}} C_{[i-1],t} b_{i\cdot i-1,t} \right), \tag{C.39}$$

$$V_{it}^{(j)} = \left((V_{it}^{\star(j)})^{-1} + n_i e^{-x_{it}^{(j)}} C_{[i-1],t} \right)^{-1}, \tag{C.40}$$

for $t = 1, \ldots, T$, while running the particle filter algorithm.

C.5. Additional Posterior Summary Results

We carefully checked the convergence of the proposed MCMC algorithm as well as its mixing rate (in terms of the correlation of the Gibbs draws). For this we monitored the trace plots of the Gibbs draws for the parameters and their sample autocorrelation function (ACF). Following the standard practice in the literature, we report in Section 4.5.2 as diagnostic summary statistics for the mixing rate the inefficiency factor (IF). For the RSS model incorporating shrinkage priors the IF values across all of its parameters (excluding the shrinkage prior parameters) range from 1.01 to 74.02 with a median value of 20.74 indicating a high sampling efficiency with a fast mixing rate of the MCMC algorithm. In Figure C.1 we provide the box plot of the IF values for all the parameters (excluding the shrinkage prior parameters) for the RSS shrinkage model and in Figures C.2 – C.4 the trace plots of the Gibbs draws together with the respective ACFs for a parameter with small IF value (Figure C.2), median IF value (Figure C.3) and large IF value (Figure C.4). Those plots corroborate the fast mixing rate of the MCMC algorithm.

C.6. A Note on Sampling from the GIG Distribution

Hörmann and Leydold (2014, 2015) have implemented a stable random number generator for the $\mathcal{GIG}$ in their R package `GIGrvg`. Originally written in `C/C++` the package `GIGrvg` has been translated to MATLAB and made publicly available by Hartkopf (2020, `gigrnd`).

Note, that the sampling scheme of Hörmann and Leydold (2014, 2015) seems to lack proper handling of the limiting cases when the $\mathcal{GIG}$ parameters are close to zero. In the following, useful results for those limiting cases are derived. The MATLAB function `gigrnd` has been expanded to handle these limiting cases more properly.

C.6.1. Some properties of the Generalized Inverse Gaussian distribution

A random variable X follows a Generalized Inverse Gaussian distribution with parameters λ, χ and ψ, say $X \sim \mathcal{GIG}(\lambda, \chi, \psi)$, if its pdf is given by

$$f_X(x) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\chi x^{-1} + \psi x)\right\}, \tag{C.41}$$

where $K_\lambda(z)$ denotes the modified Bessel function of the second kind. The domain of variation of the parameters in (C.41) is given by

$$\lambda \in \mathbb{R}, \quad (\chi, \psi) \in \begin{cases} \{(\chi, \psi) \colon \chi \geq 0, \psi > 0\}, & \lambda > 0 \\ \{(\chi, \psi) \colon \chi > 0, \psi > 0\}, & \lambda = 0 \\ \{(\chi, \psi) \colon \chi > 0, \psi \geq 0\}, & \lambda < 0, \end{cases} \tag{C.42}$$

where parameterizations with $\chi = 0$ ($\lambda > 0$) or $\psi = 0$ ($\lambda < 0$) have to be understood as limiting cases (Jørgensen, 1982; von Hammerstein, 2010).

If the random variable X has distribution $\mathcal{GIG}(\lambda, \chi, \psi)$ we have the following reciprocal and rescaling properties (see Jørgensen, 1982, pp. 7–12)

$$X^{-1} \sim \mathcal{GIG}(-\lambda, \psi, \chi) \tag{C.43}$$

$$cX \sim \mathcal{GIG}(\lambda, c\chi, c^{-1}\psi), \quad \text{for } c > 0, \tag{C.44}$$

and its rth moment is given by

$$\mathrm{E}[X^r] = \frac{K_{\lambda+r}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \left(\frac{\chi}{\psi}\right)^{\frac{r}{2}}. \tag{C.45}$$

C.6.2. Random number generation for limiting cases

Hörmann and Leydold (2014, 2015) have implemented a stable random number generator for the $\mathcal{GIG}(\lambda, \chi, \psi)$ distribution in the case that $\chi, \psi > 0$ in their R package `GIGrvg`. However, their sampling algorithm seems to lack proper handling of the limiting cases $\chi = 0(\lambda > 0)$ or $\psi = 0(\lambda < 0)$, see Eq. (C.42). In the following, few re-

sults for $\mathcal{GIG}(\lambda, 0, \psi)$ and $\mathcal{GIG}(-|\lambda|, \chi, 0)$ are derived, and also the problematic cases $\lambda < 0, \chi \to 0$ and $\lambda > 0, \psi \to 0$ are considered.

The derivations of the limiting cases rely on the following asymptotically equivalent representation of the modified Bessel function of the second kind (see Abramowitz and Stegun, 1972, Eq. (9.6.9), and Greiner and Reinhardt, 2009, p.74)

$$K_\lambda(z) \sim \Gamma(|\lambda|) 2^{|\lambda|-1} z^{-|\lambda|}, \quad \text{for} |\lambda| > 0, \tag{C.46}$$

with $\Gamma(z)$ being the Gamma function.[6]

Case $\lambda > 0, \chi = 0$

In the case $\lambda > 0, \chi = 0$ considering the limit $\lim_{\chi \to 0} f_X(x)$ yields

$$\begin{aligned}
\lim_{\chi \to 0} f_X(x) &= \lim_{\chi \to 0} \left[\frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \lim_{\chi \to 0} \left[\frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} \right] \lim_{\chi \to 0} \left[x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&\overset{\text{by (C.46)}}{=} \lim_{\chi \to 0} \left[\frac{(\psi/\chi)^{\lambda/2}}{2\Gamma(\lambda) 2^{\lambda-1} (\chi\psi)^{-\lambda/2}} \right] \lim_{\chi \to 0} \left[x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \lim_{\chi \to 0} \left[\frac{\psi^\lambda}{2^\lambda \Gamma(\lambda)} \right] \lim_{\chi \to 0} \left[x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \frac{(\psi/2)^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \exp\left\{ -\frac{\psi}{2} x \right\}.
\end{aligned} \tag{C.47}$$

The latter density (C.47) corresponds to that of a Gamma distribution with degrees of freedom parameter λ and scale $2/\psi$. Hence, for $\chi < \epsilon$, with ϵ being sufficiently small, one obtains a random sample from $\mathcal{GIG}(\lambda, 0, \psi)$ by sampling $X \sim \mathcal{G}(\lambda, 2/\psi)$.[7]

[6] In the case $\lambda = 0$ one obtains $K_0(z) \sim -\ln(z)$ implying that the integrating constant of the $\mathcal{GIG}$ in (C.41) goes to zero for $\chi\psi \to 0$ (see von Hammerstein, 2010, p. 9).

[7] In MATLAB one can generate a random sample from this distribution by calling `x = 2*randg(lambda)/psi;`

Case $\lambda < 0, \psi = 0$

In the case $\lambda < 0, \psi = 0$ considering the limit $\lim_{\psi \to 0} f_X(x)$ yields

$$\begin{aligned}
\lim_{\psi \to 0} f_X(x) &= \lim_{\psi \to 0} \left[\frac{(\psi/\chi)^{-|\lambda|/2}}{2K_{|\lambda|}(\sqrt{\chi\psi})} x^{-|\lambda|-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \lim_{\psi \to 0} \left[\frac{(\psi/\chi)^{-|\lambda|/2}}{2K_{|\lambda|}(\sqrt{\chi\psi})} \right] \lim_{\psi \to 0} \left[x^{-|\lambda|-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&\overset{\text{by (C.46)}}{=} \lim_{\psi \to 0} \left[\frac{(\psi/\chi)^{-|\lambda|/2}}{2\Gamma(|\lambda|) 2^{|\lambda|-1} (\chi\psi)^{-|\lambda|/2}} \right] \lim_{\psi \to 0} \left[x^{-|\lambda|-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \lim_{\psi \to 0} \left[\frac{\chi^{|\lambda|}}{2^{|\lambda|}\Gamma(|\lambda|)} \right] \lim_{\psi \to 0} \left[x^{-|\lambda|-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \frac{(\chi/2)^{|\lambda|}}{\Gamma(|\lambda|)} x^{-|\lambda|-1} \exp\left\{ -\frac{\chi}{2} x^{-1} \right\}. \qquad \text{(C.48)}
\end{aligned}$$

The latter density (C.48) corresponds to that of an Inverse-Gamma distribution with degrees of freedom parameter $|\lambda|$ and scale $\chi/2$. Hence, for $\psi < \epsilon$, with ϵ being sufficiently small, one obtains a random sample from $\mathcal{GIG}(-|\lambda|, \chi, 0)$ by sampling $X \sim \mathcal{IG}(|\lambda|, \chi/2)$.[8]

Case $\lambda > 0, \psi \to 0$

In the *undefined* case $\lambda > 0, \psi = 0$ considering the limit $\lim_{\psi \to 0} f_X(x)$ yields

$$\begin{aligned}
\lim_{\psi \to 0} f_X(x) &= \lim_{\psi \to 0} \left[\frac{(\psi/\chi)^{\lambda/2}}{2K_{\lambda}(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \\
&= \lim_{\psi \to 0} \left[\left(\frac{\psi}{2} \right)^{\lambda} \frac{1}{\Gamma(\lambda)} x^{\lambda-1} \exp\left\{ -\frac{1}{2}(\chi x^{-1} + \psi x) \right\} \right] \to 0, \qquad \text{(C.49)}
\end{aligned}$$

implying that the density of $\mathcal{GIG}(\lambda, \chi, 0)$ has no positive mass. However, it may occur in applications that one faces the case $\lambda > 0$ and $0 < \psi < \epsilon$, with $\epsilon > 0$ being some small non-zero upper bound. In this problematic case the rescaling property of the GIG might prove itself as helpful. Using (C.44) one could generate a random sample

[8] In MATLAB one can generate a random sample from this distribution by calling `x = 0.5*chi/randg(abs(lambda));`

from

$$Y = \psi X \sim \mathcal{GIG}(\lambda, \psi\chi, 1), \tag{C.50}$$

and do rescaling afterwards to obtain a draw for X as $X = Y/\psi$, assuming that ψ is large enough to assure $\psi^{-1} \ll \infty$. Now, if $\chi\psi > 0$ the random variate Y can be generated using the sampling procedure of Hörmann and Leydold (2014, 2015), or else if $\psi\chi \to 0$ by sampling from a Gamma distribution with degrees of freedom λ and scale 2, i.e., $Y \sim \mathcal{G}(\lambda, 2)$.

Case $\lambda < 0, \chi \to 0$

In the *undefined* case $\lambda < 0, \chi = 0$ considering the limit $\lim_{\chi\to 0} f_X(x)$ yields

$$\begin{aligned}\lim_{\chi\to 0} f_X(x) &= \lim_{\chi\to 0}\left[\frac{(\psi/\chi)^{-|\lambda|/2}}{2K_{|\lambda|}(\sqrt{\chi\psi})}x^{-|\lambda|-1}\exp\left\{-\frac{1}{2}(\chi x^{-1}+\psi x)\right\}\right] \\ &= \lim_{\chi\to 0}\left[\left(\frac{\chi}{2}\right)^{|\lambda|}\frac{1}{\Gamma(|\lambda|)}x^{-|\lambda|-1}\exp\left\{-\frac{1}{2}(\chi x^{-1}+\psi x)\right\}\right] \to 0,\end{aligned} \tag{C.51}$$

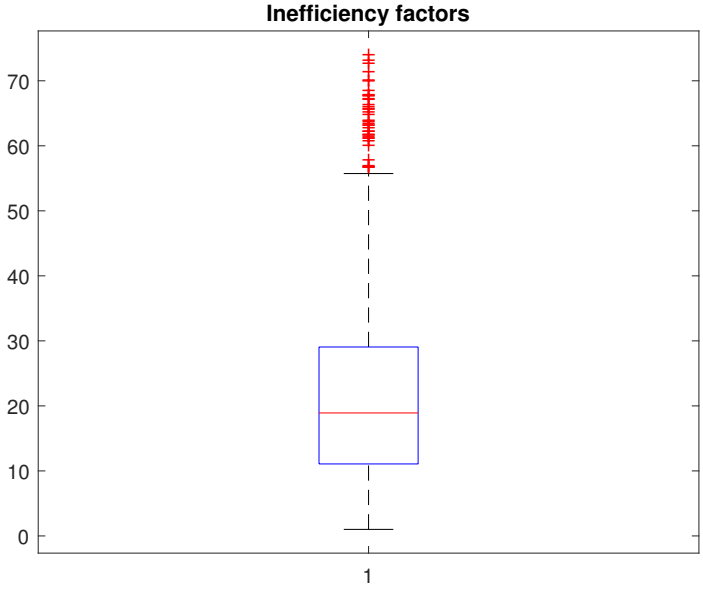
implying that the density of $\mathcal{GIG}(\lambda, 0, \psi)$ has no positive mass. However, it may occur in applications that one faces the case $\lambda < 0$ and $0 < \chi < \epsilon$, with $\epsilon > 0$ being some small non-zero upper bound. In this problematic case the rescaling property of the GIG might prove itself as helpful. Using (C.44) one could generate a random sample from

$$Y = \chi^{-1}X \sim \mathcal{GIG}(\lambda, 1, \psi\chi), \tag{C.52}$$

and do rescaling afterwards to obtain a draw for X as $X = \chi Y$, assuming that χ is large enough to assure $\chi^{-1} \ll \infty$. Now, if $\chi\psi > 0$ the random variate Y can be generated using the sampling procedure of Hörmann and Leydold (2014, 2015), or else if $\psi\chi \to 0$ by sampling from an Inverse-Gamma distribution with degrees of freedom $|\lambda|$ and scale $1/2$, i.e., $Y \sim \mathcal{IG}(|\lambda|, 1/2)$.

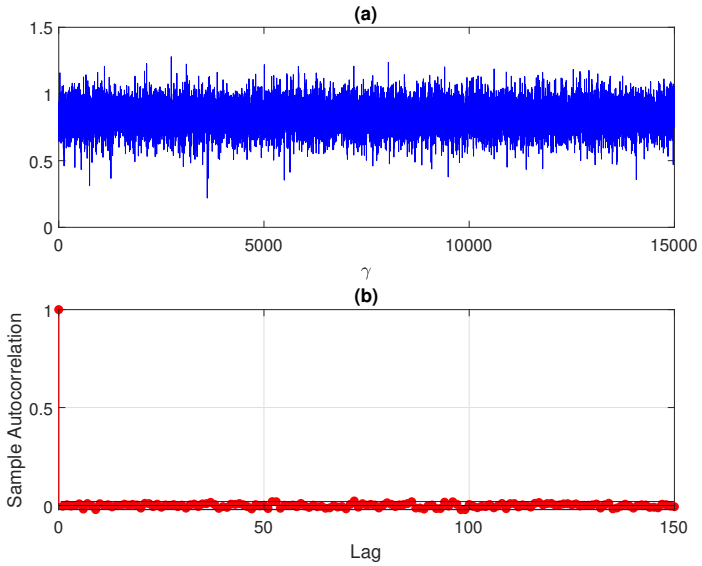
C.7. Additional Figures

Figure C.1.: Inefficiency factors.



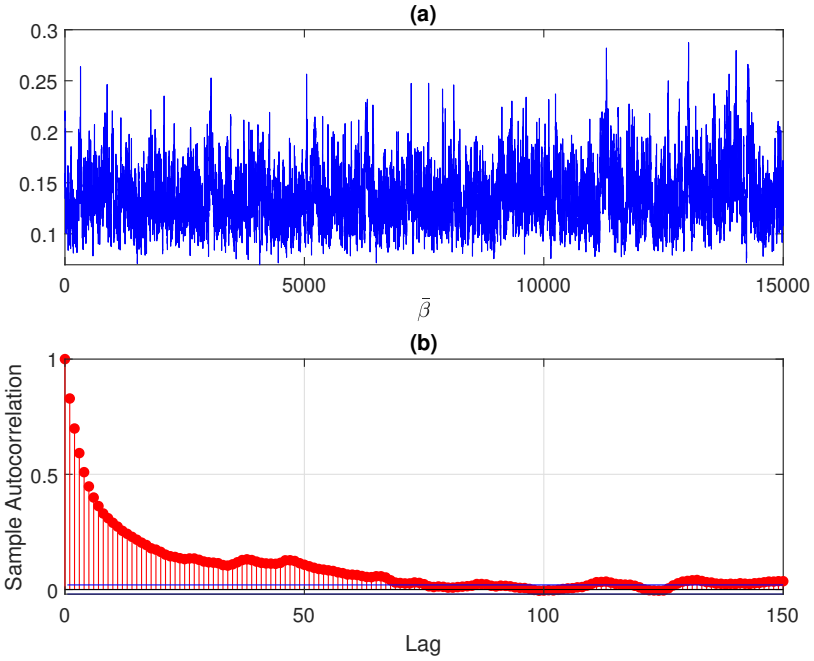
Box plots of the IF values for all parameters (excluding shrinkage prior parameters) for the RSS shrinkage model.

Figure C.2.: Traceplot for a parameter with small IF value.



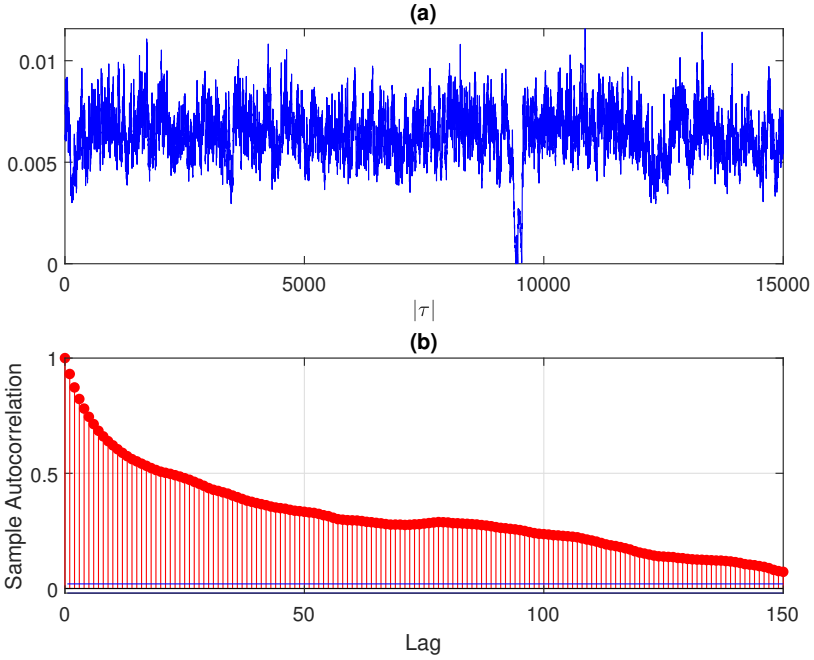
Upper panel: Trace plots of the Gibbs draws for the parameter of the RSS shrinkage model with small IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Figure C.3.: Traceplot for a parameter with median IF value.



Upper panel: Trace plots of the Gibbs draws for the parameter of the RSS shrinkage model with median IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Figure C.4.: Traceplot for a parameter with large IF value.



Upper panel: Trace plots of the Gibbs draws for the parameter of the the RSS shrinkage model with large IF value; Lower panel: The sample ACF for the Gibbs draws (excluding those of the burn-in period).

Bibliography

Abramowitz, M. and Stegun, I., editors (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Number 55 in National Bureau of Standards Applied Mathematics. Dover Publications Inc., New York, 9 edition.

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high-dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realised volatility. *Econometrica*, 71:579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, J. (2005). A framework for exploring the macroeconomic determinants of systematic risk. *The American Economic Review (Papers and Proceedings)*, 95(2):398–404.

Andersson, S. A. and Klein, T. (2010). On Riesz and Wishart distributions associated with decomposable undirected graphs. *Journal of Multivariate Analysis*, 101:789–810.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society - Series B*, 72:269–342.

Ang, A. and Chen, J. (2007). CAPM over the long run: 1926–2001. *Journal of Empirical Finance*, 14(1):1–40.

Asai, M. and McAleer, M. (2015). Forecasting co-volatilites via factor models with asymmetry and long memory in realized covariance. *Journal of Econometrics*, 189(2):251–262.

Asai, M., McAleer, M., and Medeiros, M. C. (2012a). Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics*, 189(2):495–512.

Asai, M., McAleer, M., and Medeiros, M. C. (2012b). Modeling and forecasting noisy realized volatility. *Computational Statistics and Data Analysis*, 56:217–230.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 64(2):253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica*, 72(3):885–925.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1):93–101.

Bauwens, L., Braione, M., and Storti, G. (2014). Forecasting comparison of long term component dynamic models for realized covariance matrices. *CORE working paper*, 2014/53.

Bauwens, L., Braione, M., and Storti, G. (2016). Multiplicative conditional correlation models for realized covariance matrices. *CORE working paper*, 2016/41.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21:79–109.

Bauwens, L., Storti, G., and Violante, F. (2012). Dynamic conditional correlation models for realized covariance matrices. *CORE working paper*.

Bekierman, J. and Gribisch, B. (2016). Estimating stochastic volatility models using realized measures. *Studies in Nonlinear Dynamics & Econometrics*, 20:279–300.

Bekierman, J. and Manner, H. (2018). Forecasting realized variance measures using time-varying coefficient models. *International Journal of Forecasting*, 34(2):276–287.

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97.

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

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

Bollerslev, T. and Zhang, B. Y. B. (2003). Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10(5):533–558.

Boudt, K., Laurent, S., Lunde, A., Quaedvlieg, R., and Sauri, O. (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics*, 196:347–367.

Brito, D., Medeiros, M. C., and Ribeiro, R. (2018). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage. *SSRN Working Paper*.

Callot, L. A. F., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

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

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

Chan, J. C. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state-space models. *International Journal of Mathematical Modeling and Numerical Optimization*, 1:101–120.

Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high-dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371.

Chiriac, R. and Voev, V. (2011). Modeling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Christensen, K. and Kinnebrock, S. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Corsi, F. (2009). A simple approximative long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Audrino, F., and Reno, R. (2012). *Handbook of Volatility Models and Their Applications*, chapter HAR modeling for realized volatility forecasting, pages 363–382. John Wiley & Sons, Hoboken, New Jersey.

Darolles, S., Francq, C., and Laurent, S. (2018). Asymptotics of Cholesky GARCH models and time-varying conditional betas. *Journal of Econometrics*, 204:223–247.

de Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrica*, 82:339–350.

Díaz-García, J. A. (2013). A note on the moments of the Riesz distribution. *Journal of Statistical Planning and Inference*, 143:1880–1886.

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business and Economic Statistics*, 30(2):212–228.

Engle, R., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business and Economic Statistics*, 38(2):363–375.

Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics*, 14(4):643–667.

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

Fan, J., Furger, A., and Xiu, D. (2016). Incorporating Global Industrial Classification Standard into portfolio allocation: a simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business and Economic Statistics*, 34(4):489–503.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606.

Frost, P. A. and Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Frühwirth-Schnatter, S. and Wagner, H. (2011). *Bayesian Statistics 9*, chapter Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data, pages 165–200. Oxford University Press.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Hoboken, New Jersey.

Ghysels, E. and Jacquier, E. (2006). Market beta dynamics and portfolio efficiency. *Working paper, University of North Carolina*.

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211–223.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEEE Proceedings-F*, volume 140, pages 107–113.

Gorgi, P., Hansen, P. R., and Koopman, S. J. (2019). Realized Wishart-GARCH: a score-driven multi-asset volatility model. *Journal of Financial Econometrics*, 17(1):1–32.

Gourieroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

Greiner, W. and Reinhardt, J. (2009). *Quantum Electrodynamics*. Springer, 4 edition.

Gribisch, B. (2018). A latent dynamic factor approach to forecasting multivariate stock market volatility. *Empirical Economics*, 55(2):621–651.

Gribisch, B. and Hartkopf, J. P. (2020). Modeling realized covariance measures with heterogeneous liquidity: A generalized matrix-variate Wishart state-space model. *SSRN Working paper*. http://dx.doi.org/10.2139/ssrn.3676140.

Gribisch, B., Hartkopf, J. P., and Liesenfeld, R. (2020). Factor state-space models for high-dimensional realized covariance matrices of asset returns. *Journal of Empirical Finance*, 55(1):1–20.

Grothe, O., Kleppe, T. S., and Liesenfeld, R. (2019). The Gibbs sampler with particle efficient importance sampling for state-space models. *Econometric Reviews*, 38(10):1152–1175.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix variate distributions*. Monographs and Surveys in Pure and Applied Mathematics 104. Chapman & Hall, Boca Raton.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

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

Hartkopf, J. P. (2020). gigrnd. A MATLAB function for the generation of Generalized Inverse Gaussian distributed random variates. MATLAB Central File Exchange, https://www.mathworks.com/matlabcentral/fileexchange/78805-gigrnd).

Hassairi, A. and Lajmi, S. (2001). Riesz exponential families on symmetric cones. *Journal of Theoretical Probability*, 14(4):927–948.

Hautsch, N., Kyj, L., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.

Hautsch, N., Kyj, L. M., and Malec, P. (2011). The merit of high-frequency data in portfolio allocation. *SSRN Working paper*. http://dx.doi.org/10.2139/ssrn.1926098.

Hörmann, W. and Leydold, J. (2014). Generating Generalized Inverse Gaussian random variates. *Statistical Computing*, 24:547–557.

Hörmann, W. and Leydold, J. (2015). GIGrvg: Random variate generator for the GIG distribution. R package version 0.5. URL: http://CRAN.R-project.org/package=GIGrvg.

Huber, F., Koop, G., and Onorante, L. (2020). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of the Royal Statistical Society - Series B*, to appear.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1683.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jin, X. and Maheu, J. M. (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics*, 11(2):335–369.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics*, 192(1):19–39.

Jin, X., Maheu, J. M., and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics*, 34:641–660.

Jørgensen, B. (1982). *Statistical properties of the Generalized Inverse Gaussian distribution*. Number 9 in Lecture Notes in Statistics. Springer, New York, Berlin.

Kalnina, I. (2015). Inference for nonparametric high-frequency estimators with an application to time-variation in betas. *Working paper, Université de Montréal*.

Kastner, G. (2019). Sparse Bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics*, 210:98–115.

Kastner, G. and Huber, F. (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, pages 1–24.

Kim, C. and Nelson, C. (1999). *State-Space Models with Regime Switching*. MIT Press, Cambridge, Massachusetts, London, England.

Kim, D. (2014). Maximum Likelihood estimation for vector autoregressions with multivariate stochastic volatility. *Economics Letters*, 123:282–286.

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

Konno, Y. (1988). Exact moments of the multivariate F and Beta distributions. *Journal of the Japanese Statistical Society*, 18(2):123–130.

Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution. *Annals of the institute of statistical mathematics*, 43(1):157–165.

Koop, G., Korobilis, D., and Pettenuzo, D. (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics*, 210:135–154.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3(2):73–84.

Laurent, S., Rombouts, J. V. K., and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173(1):1–10.

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003). Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics*, 85(3):735–747.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621.

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

Lindsten, F., Jordan, M. I., and Schön, T. B. (2014). Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15:2145–2184.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37.

Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2016). Parsimony inducing priors for large scale state-space models. *Bayesian Analysis*, pages 116–133.

Lunde, A., Shephard, N., and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, 34(4):504–518.

Luo, J. and Chen, L. (2020). Realized volatility forecast with Bayesian random compressed multivariate HAR model. *International Journal of Forecasting*, 36:781–799.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Chichester, West Sussex, England.

Magnus, J. R. and Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, New York.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135:499–526.

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55:199–212.

Morgan, J. P. (1996). Riskmetrics. techreport, J. P. Morgan, New York. Fourth ed.

Moura, G. V. and Noriller, M. R. (2019). Maximum Likelihood estimation of a TVP-VAR. *Economics Letters*, 174:78–83.

Moura, G. V., Santos, A., and Ruiz, E. (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking and Finance*, 118:1–13.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Hoboken, New Jersey.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics*, 27(6):907–933.

Opschoor, A., Janus, P., Lucas, A., and van Dijk, D. (2017). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36(4):643–657.

Opschoor, A. and Lucas, A. (2019). Fractional integration and fat tails for realized covariance kernels. *Journal of Financial Econometrics*, 17(1):66–90.

Park, S. and Linton, O. (2012). *Handbook of Volatility Models and Their Applications*, chapter Realized volatility: Theory and applications, pages 317–345. John Wiley & Sons, Hoboken, New Jersey.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103:681–686.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1):23–42.

Philipov, A. and Glickman, M. E. (2006). Factor stochastic volatility via Wishart processes. *Econometric Reviews*, 25(2–3):311–334.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Schön, T., Gustafsson, F., and Nordlund, P. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7):2279–2289.

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

Shephard, N. (1994). Local scale models: State space alternative to integrated GARCH processes. *Journal of Econometrics*, 60:181–202.

Sheppard, K. and Xu, W. (2019). Factor high-frequency based volatility (HEAVY) models. *Journal of Financial Econometrics*, 17(1):33–65.

Shirota, S., Omori, Y., Lopes, H. F., and Piao, H. (2017). Cholesky realized stochastic volatility model. *Econometrics and Statistics*, 3:34–59.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society - Series B*, 64(4):583–639.

Takahashi, M., Omori, Y., and Watanabe, T. (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis*, 53:2404–2426.

Tao, M., Wang, Y., Yao, Q., and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association*, 106:1025–1040.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society - Series B*, 58:267–288.

Uhlig, H. (1994). On singular Wishart and singular multivariate Beta distributions. *The Annals of Statistics*, 22:395–405.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, 65:59–73.

Vassallo, D., Buccheri, G., and Corsi, F. (2019). A DCC-type approach for realized covariance modeling with score-driven dynamics. *Working paper*.

Veleva, E. (2009). Testing a normal covariance matrix for small samples with monotone missing data. *Applied Mathematical Sciences*, 54(3):2695–2702.

von Hammerstein, E. A. (2010). *Generalized hyperbolic distributions: Theory and applications to CDO pricing*. PhD thesis, Fakultät für Mathematik und Physik der Albert-Ludwigs-Universität Freiburg im Breisgau.

von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scandinavian Journal of Statistics*, 15(2):97–109.

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

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

Zhang, L., Aït-Sahalia, Y., and Mykland, P. A. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100:1394–1411.