MODELLING REALIZED COVARIANCE MATRICES WITH STOCHASTIC VOLATILITY LATENT FACTORS: FILTER, LIKELIHOOD, FORECAST

Giorgio CALZOLARI[1] and Roxana HALBLEIB[2]

Preliminary Version
January 16, 2018

Abstract

This paper proposes a latent factor model with underlying Wishart distribution to capture the dynamics of daily realized covariance matrices for forecasting purposes. The long memory in the series is captured by means of aggregating latent factors with stochastic volatility structure, where the factors are extracted from the commonality in the dynamics of realized variance and covariance series. This new model accommodates the positive-definiteness and the symmetry of variance-covariance matrix forecasts within a very parsimonious framework with no parameter constraints. For estimation purposes, we implement the numerical exact maximum likelihood method on the Kitagawa state-space filtering procedure. We provide Monte Carlo evidence on the accuracy of estimates and, on real data, we show that our model outperforms existing ones when forecasting daily variance-covariance matrices.

Keywords: Realized Covariance, Latent Dynamic Factor Model, Wishart Distribution, Stochastic Volatility, State-Space Model, Kitagawa Filter, Long Memory

[1]Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", University of Firenze, Italy; email: calzolar@disia.unifi.it.

[2]Corresponding author. Department of Economics, University of Konstanz, P.O. Box 124, Universitaetstrasse 10, 78464, Konstanz, Germany; email: roxana.halbleib@uni-konstanz.de; telephone: 0049 (0) 7531 88 5373; fax: 0049 (0) 7531 88 4450.

1 Introduction

Modelling and forecasting time varying covariance matrices of financial returns has always been a challenge for the financial econometric literature. The first approaches in this direction are the Multivariate Generalized Autoregressive Heteroskedastic (MGARCH) models (see Bauwens et al. (2006) for a review) and the Multivariate Stochastic Volatility (MSV) models (see Asai and McAleer (2006) for a review), which measure the unobserved covariance matrices as a function of past observations and of latent stochastic shocks, respectively. However, both approaches are difficult to implement in practice as they suffer from the curse of dimensionality and estimation difficulties when applied to large dimensional matrices. Alternatively, the Realized Covariance (RCov) estimators introduced by Barndorff-Nielsen and Shephard (2004) provide consistent estimates of daily covariance matrices by using high-frequency information (see Zhang et al. (2005), Barndorff-Nielsen et al. (2011), Lunde et al. (2016), among others, for further developments). However, in order to make forecasts based on RCov one has to model their dynamics.

The literature on modeling and forecasting RCov matrices has been growing at a slow pace due to numerous challenges: the forecasts have to be positive definite and symmetric, the models have to account for the empirical features of the series, such as long memory and common dynamics and, in the same time, they have to remain parsimonious when applied to high dimensional matrices. All these challenges raise further problems in their practical implementation, as the estimation becomes infeasible starting already at matrices of dimension 6×6. The most prominent developments are provided by Gouriéroux et al. (2009), Chiriac and Voev (2011), Jin and Maheu (2013), Golosnoy et al. (2012) and Bauwens et al. (2017). However, each of these developments manage to tackle the challenges mentioned above either singly or by undergoing heavily parametric restrictions.

In this paper we provide a very parsimonious approach to model and forecast RCov series that solves all these challenges without parametric constraints and it is estimable by exact numerical Maximum Likelihood (ML). It consists in an additive latent factor model. The factors feature stochastic volatility structures and, conditionally on them, the RCov matrix follows a Wishart distribution. The stochastic volatility structure and the distributional assumption automatically assure the positive definiteness and symmetry of the forecasts. The latent factor structure accounts for the common dynamics in the series and the additive structure capture the long persistence in their autocorrelation function (ACF) in the spirit of Granger (1980). The estimation consists in applying a Kitagawa filtering approach to provide exact numerical values of the likelihood.

The new approach we introduce, denoted as the Factor Autoregressive (FAW) model, is very flexible as it allows that realized variances and realized covariances are driven by common, but also different factors and exhibit different type of dynamics and persistencies. Moreover, FAW model is the most parsimonious model in the existent literature as the number of parameters are of order $O(n)$, while the one of the alternatives is of order $O(n^2)$, where n gives the dimension of the RCov matrix. The model is a non-linear non-Gaussian state-space model, which might be in general very difficult to estimate. However, implementing the filter introduced by Kitagawa (1987), which is the first at-

tempt in the literature for these kind of models, allows the practical implementation of the FAW model.

The closest to our FAW model is the Conditional Autoregressive (CAW) model of Golosnoy et al. (2012), which also assumes an underlying Wishart distribution, but where the dynamics of the conditional mean of the RCov series are originally captured by a BEKK framework. In order to account for the long-memory of the series, Golosnoy et al. (2012) extend the original CAW model by replacing the BEKK structure with a HAR or MIDAS one. Differently, the FAW model allows that the dynamics exhibit stochastic components and models the long memory without any further extensions. Moreover, the CAW model becomes practically implementable only by imposing heavily parametric constraints, such as diagonality for the parameter matrices, while the FAW model does not necessitate any.

The Kitagawa filtering approach we implement provides, besides exact numerical values of the likelihood, also recursive formulas for the prediction, filtering and smoothing of the factors. Thus, the FAW approach can be applied to forecast high-dimensional RCov matrices with acceptable computation costs. In our simulation exercise, we provide evidence on the good performance of the Kitagawa filter with ML in providing accurate estimates of the parameter of interest. Moreover, we show that in contrast to standard autoregressive approaches that suffer from increasing the dimensionality of RCov matrices, the FAW approach profits in terms of efficiency of the estimates, as increasing n provides more information on the commonality in the underlying series dynamics.

In our empirical exercise we apply two types of FAW model to capture the dynamics of daily series of RCov matrices up to dimension 10×10. The in-sample goodness-of-fit improves by allowing that realized variances have their own factor component and by increasing the dimension n. Out-of-sample, we compare the performance of FAW to the RiskMetrics approach that involves no parameter estimation and to CAW models that can be applied to matrices of dimension 10×10 only under heavily parametric restrictions. The forecasting results are in line with the in-sample results and indicate that the FAW specifications outperform the alternative ones.

The rest of the paper is organized as follows. Section 2 introduces the FAW model and its estimation procedure. Section 3 presents simulation results and section 4 presents empirical results from applying it to real data as well as results from comparing its performance against existing approaches. Section 5 concludes. Details on the filtering procedure and computation are given in Appendix A.

2 Factor Wishart (FAW) Model

Let Y_t with $t = 1, \ldots, T$ be the series of positive definite daily RCov matrices of dimension $n \times n$, where $Y_t = (y_{ij,t})$ with $i, j = 1, \ldots, n$. We assume that, conditional on a set of stochastic volatility latent factors $\mathbf{h}_t$, Y_t is central Wishart distributed as it follows:

$$Y_t|\mathbf{h}_t \sim \mathcal{W}_n\left(m, \frac{\Sigma_t}{m}\right) \quad , \quad m \geq n \tag{1}$$

where Σ_t is a symmetric, positive definite matrix such that (Muirhead, 1982):

$$\mathbb{E}\left[Y_t|\mathbf{h}_t\right] = \Sigma_t \tag{2}$$

and $\mathbf{h}_t = (\mathbf{h}_{1,t}, \mathbf{h}_{2,t})'$ where $\mathbf{h}_{1,t} = (h_{1,1,t}, \dots h_{1,K_1,t})'$ and $\mathbf{h}_{2,t} = (h_{2,1,t}, \dots h_{2,K_2,t})'$ such that

$$\begin{aligned}
\Sigma_t &= \sum_{k_1=1}^{K_1} b_{k_1}\, e^{h_{1,k_1,t}}\, b_{k_1}' + \sum_{k_2=1}^{K_2} e^{h_{2,k_2,t}}\, \Gamma_{k_2} && (3)\\
h_{1,k_1,t} &= \rho_{k_1}\, h_{1,k_1,t-1} + \delta_{1,k_1}\, u_{1,k_1,t} \qquad \forall k_1 = 1, \dots, K_1 && (4)\\
h_{2,k_2,t} &= \rho_{k_2}\, h_{2,k_2,t-1} + \delta_{2,k_2}\, u_{2,k_2,t}, \qquad \forall k_2 = 1, \dots, K_2 && (5)
\end{aligned}$$

where $u_{1,k_1,t}$ and $u_{2,k_2,t}$ are independent of each other and independent and identical normally distributed with mean zero and variance 1 for all $k_1 = 1, \dots, K_1$ and $k_2 = 1, \dots, K_2$. b_{k_1} is a vector of dimension $n \times 1$ for all $k_1 = 1, \dots, K_1$, Γ_{k_2} is a diagonal matrix of dimension $n \times n$, i.e., $\Gamma_{k_2} = diag(\gamma^2_{1,k_2}, \dots, \gamma^2_{n,k_2})$ for all $k_2 = 1, \dots, K_2$. Sufficient conditions to identify the model above is that the first elements of the vector b_{k_1} with $k_1 = 1, \dots, K_1$ are set to 1.

We denote Equation (1) the observation equation and equations (4) and (5) the state equations. As a consequence, the model in equations (1)-(5) is a state-space model with underlying Wishart distribution, which we denote as the Factor Wishart model with K_1 and K_2 factors: FAW(K_1,K_2).

In Equation (1), m is scalar and denotes the degrees of freedom of the Wishart distribution. Equation (3) captures the common dynamics of realized variances and covariances by means of latent factors with stochastic volatility structures as defined in the equations (4) and (5). More precisely, the structure given in Equation(3) is very flexible as it allows that the diagonal elements of Σ_t that correspond to realized variances are driven by $K_1 + K_2$ factors, while the off-diagonal elements of Σ_t that correspond to realized covariances are driven by the K_1 factors that are also common to realized variances. This is in line with the empirical features of the observed realized variances and covariances that indicate that both type of series exhibit common dynamics, especially during turbulent times (e.g., Calzolari et al. (2017)) with the variances showing more dependency in time than the covariances.

The framework given in equations (3)-(5) allows matrix Σ_t to have a stochastic structure, which is different from the CAW model of Golosnoy et al. (2012), where Σ_t follows a conditionally deterministic path that resembles a BEKK-GARCH specification. More precisely, we allow that both sets of factors have stochastic volatility structure, which automatically assures the positive definiteness and symmetry of Σ_t. Σ_t and $\mathbf{h}_t$ are stationary when $|\rho_{k_1}| < 1$ for all $k_1 = 1, \dots, K_1$ and $|\rho_{k_2}| < 1$ for all $k_2 = 1, \dots, K_2$.

Particularly appealing for our FAW model is that it captures the main empirical feature of realized (co)variance series, namely the slowly decaying autocorrelation function or long-memory type of persistence by using a very simple framework, namely an additive structure of stochastic volatility latent factors with short memory dynamics in the spirit of Granger (1980). Thus, our approach does not need further extensions as the CAW

model to account for the long-memory of the underlying series, i.e. CAW-MIDAS and CAW-HAR, but it automatically accommodates it.

Thus, FAW model manages to accommodate the two main features of realized covariances: common dynamics and long-memory within a very parsimonious framework that automatically guarantees the symmetry and positive definiteness of the covariance matrices. The total number of parameters involved is equal to $(K_1 + K_2)n - K_1 + 1$. Thus, for $K_1 = K_2 = 1$, the total number of parameters is equal to $2n+4$. This makes our approach the most parsimonious one in the literature so far, which accounts for the rich dynamic structures of the underlying series without parameter constraints, besides stationary: all existing approaches that aim at modelling the covariance matrices with $n(n-1)/2$ different elements involve a total number of parameters of order n^2: e.g., the simplest CAW model, namely the diagonal CAW(1,1) has already $n(n+1)/2 + 2n + 1$ parameters.

For our purposes, in what follows, we focus on providing empirical evidence from applying two types of FAW models: 1) one with only one factor that is common to both variances and covariances, i.e., $\mathbf{h}_1 = \mathbf{h}_2$ and $K_1 = K_2 = 1$, which we denote FAW(1) and 2) one with two factors, one for the variances and covariances and one typical only to the variances, i.e. $\mathbf{h}_1 \neq \mathbf{h}_2$ and $K_1 = K_2 = 1$, which we denote FAW(1,1). Thus, the dynamics of Σ_t defined in equations (3)-(5) can be written for FAW(1) as:

$$\Sigma_t = (bb' + \Gamma)\, e^{h_t}, \tag{6}$$
$$h_t = \rho\, h_{t-1} + \delta\, u_t, \tag{7}$$

and for FAW(1,1) as:

$$\Sigma_t = (b\, e^{h_{1,t}}\, b') + e^{h_{2,t}}\, \Gamma, \tag{8}$$
$$h_{1,t} = \rho_1\, h_{1,t-1} + \delta_1\, u_{1,t}, \tag{9}$$
$$h_{2,t} = \rho_2\, h_{2,t-1} + \delta_2\, u_{2,t}, \tag{10}$$

where $\Gamma = diag(\gamma_1^2, \ldots, \gamma_n^2)$, $|\rho|, |\rho_1|, |\rho_2|$ are strictly smaller than 1, $u_t, u_{1,t}, u_{2,t}$ are $i.i.d.\ N(0,1)$ and the first element of b is fixed to 1.

In order to estimate the parameters of FAW model, we implement the maximum likelihood approach with the filtering procedure of Kitagawa (1987). Let $\boldsymbol{\theta}$ be the vector of parameters of the FAW model given in equations (1)-(5). Then, the maximum likelihood estimator of $\boldsymbol{\theta}$ is obtained by:

$$\widehat{\boldsymbol{\theta}_{ML}} = \arg\max_{\theta} \ln f(Y_1, Y_2, \ldots, Y_T; \boldsymbol{\theta}) = \arg\max_{\theta} \sum_{t=2}^{T} \ln f(Y_t|\overline{Y_{t-1}}; \boldsymbol{\theta}) + f(Y_1|; \boldsymbol{\theta}), \tag{11}$$

where $f(\cdot)$ denotes the probability density function (pdf) and $\overline{Y_t}$ contains , $Y_1, Y_2, ..., Y_t$.

The idea of the filtering approach is to compute $f(Y_t|\overline{Y_{t-1}}; \boldsymbol{\theta})$ by integrating out $\mathbf{h_t}$ from the joint conditional pdf $f(Y_t, \mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta})$:

$$f(Y_t|\overline{Y_{t-1}}; \boldsymbol{\theta}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(Y_t, \mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta}) \mathrm{d}h_{1,1,t}, \ldots, \mathrm{d}h_{1,K_1,t}, \mathrm{d}h_{2,1,t}, \ldots, \mathrm{d}h_{2,K_2,t}. \tag{12}$$

The filtering approach consists further in decomposing $f(Y_t, \mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta})$ as it follows:

$$f(Y_t, \mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta}) = f(Y_t|\mathbf{h_t}, \overline{Y_{t-1}}; \boldsymbol{\theta}) f(\mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta}) = f(Y_t|\mathbf{h_t}; \boldsymbol{\theta}) f(\mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta}) \quad (13)$$

as $\overline{Y_{t-1}}$ plays no role in the $f(Y_t|\mathbf{h_t}, \overline{Y_{t-1}}; \boldsymbol{\theta})$ as given in Equation (1) and

$$f\left(Y_t|\mathbf{h_t}; \boldsymbol{\theta}\right) = \frac{\left|\frac{\Sigma_t}{m}\right|^{-\frac{m}{2}} |Y_t|^{\frac{(m-n-1)}{2}}}{2^{\frac{mn}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(\frac{m+1-i}{2}\right)} e^{-\frac{1}{2}\,\mathrm{tr}\left(m\Sigma_t^{-1} Y_t\right)}, \quad (14)$$

where $\Gamma(\cdot)$ denotes the Gamma function. $f(\mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta})$ in Equation (2) is obtained recursively. At $t = 1$, we use as initial $f(h_1)$ its unconditional density. Appendix A provides a detailed description of the filtering approach of Kitagawa (1987) for FAW(1) and FAW(1,1).

This filtering procedure provides the exact numerical values of the likelihood by numerically integrating over a fixed number of nodes the function $f(Y_t, \mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta})$.

Besides the exact values of the likelihood, the Kitagawa filter also provides (1) the numerical values of the conditional density of the latent variable(s) at time t given observations till time $t-1$, namely $f(\mathbf{h_t}|\overline{Y_{t-1}}; \boldsymbol{\theta})$; (2) the numerical values of the conditional density(s) and conditional expectation(s) of the latent factor(s) at time t given the observed variables till time t, that is(are) the filtered factor(s), i.e., $f(\mathbf{h_t}|\overline{Y_t}; \boldsymbol{\theta})$ and $E(\mathbf{h_t}|\overline{Y_t})$ and (3) the numerical values of the conditional density(s) of the latent factor(s) at time $t+1$ given the observed variables till time t, i.e., $f(\mathbf{h_{t+1}}|\overline{Y_t}; \boldsymbol{\theta})$ from which one can easily compute the forecasts given by $E(\mathbf{h_{t+1}}|\overline{Y_t})$.

3 Monte Carlo

This section provides simulation evidence on the performance of ML to accurately estimate the parameters of FAW(1) and FAW(1,1) on a series of length $T = 4000$ of covariance matrices computed as the sample variance-covariance of m vectors of simulated multivariate normal variables with variance-covariance that changes for every t and is equal to Σ_t, where Σ_t is obtained from equations (6)-(7) for FAW(1) and from equations (9)-(10) for FAW(1,1). The choice of the parameters and length of the series is given by the values of the parameter estimates and the length of the real data we use in our empirical application described in Section 4. Tables B.1 and B.2 provides the average and standard deviation of the parameter estimates over $R = 1000$ replications for FAW(1) and FAW(1,1), respectively. We present here results for $n = 5, 8$ and 10 series, while results for larger n can be obtained from the authors upon request. As one may see from the tables, most of the estimates *seem* to be unbiased, as the difference between the true value and the average one is observable starting with the third digit after the comma. In particular, δ's and b's are estimated more accurately than ρ's, m and γ's. Increasing n increases the precision of the estimates as it provides more information on the common dynamics of the underlying series. This effect is particularly evident for FAW(1), where the commonality in the dynamics of both realized variances and realized covariances is captured by a single common factor. As expected, increasing the structure of FAW by

allowing that the dynamics of realized variances are driven by an additional factor compared to realized covariances (FAW(1,1)) worsens in general the quality of the estimates: while the parameter estimates corresponding to the factor typical to realized variances (ρ_2, δ_2, Γ) become less accurate, the other ones are similar to the ones of FAW(1) in terms of accuracy. The efficiency of the estimates also worsens in general, except for the one of the estimates of the common factor of realized variances and covariances, namely, b, ρ_1 and δ_1, for which the standard deviations decrease or are similar to the ones of FAW(1). In what follows we provide empirical evidence from applying the two models to real data.

4 Empirical Application

This section provides empirical results from applying FAW(1) and FAW(1,1) to daily series of RCov matrices of maximum $n = 10$ stocks composing the Dow Jones Industrial Average Index, namely of Alcoa Inc. (AA), American Express Company (AXP), Boeing Corporation (BA), Bank of America Corporation (BAC), Citigroup Inc. (C), Caterpillar Inc. (CAT), Chevron Corporation (CVX), Dupont (DD), Walt Disney Company (DIS), General Electric Company (GE), traded on NYSE from 01.11.2001 until 08.08.2017 ($T = 3931$). The data is obtained from NYSE Trade and Quote (TAQ) database [1] and the RCov matrix estimators are computed from returns sampled at 5-minute frequencies (up-scaled here 100) by implementing the subsample approach of Zhang et al. (2005). Figures C.1 and C.2 plot the line graphs of the realized variances and realized covariances, while figures C.3 and C.4 plot their autocorrelation function (ACF) up to lag 200 [2]. As one may observe from the first two graphs, the realized variances and the realized covariances exhibit similar dynamics, especially during turbulent financial times, such as the previous financial crises, with the commonality in the dynamics of realized variances being more pronounced. This indicates that there may be two type of factors driving these common dynamics: one common to both type of series and one common solely to realized variances. Thus, we expect that at least in-sample the FAW(1,1) is preferred to FAW(1) in terms of goodness-of-fit. From the last two graphs, we observe that both realized variances and realized covariances exhibit strong and slowly decaying ACF's typical to long memory processes. This feature is captured by our FAW representation through aggregating dynamic factors with short memory stochastic volatility structure.

Tables B.3 and B.4 present results from estimating the parameters of the models FAW(1) and FAW(1,1), respectively, from the realized covariance matrices of the first $n = 5$, $n = 8$ and $n = 10$ stocks mentioned above as well as the number of parameters involved. As one may observe, all estimated parameters are significantly different from zero. Increasing n reduces the persistence in the factors, as the values of the estimated ρ's decrease. This can be due to the fact that adding further stocks, one adds realized variance and covariance series with different persistencies (smaller and/or faster decaying ACF patters) as shown in figures C.3-C.4. The factor common to realized variances and covariances seem to be more persistent than the one typical only to realized variances ($\hat{\rho}_1 > \hat{\rho}_2$). Increasing n also has an effect on the estimated degrees of freedom, which

[1] We would like to thank Sebastian Bayer for preparing the data.

[2] Graphs for the rest of 405 realized covariances can be obtained from the authors upon request.

also increase and on the estimates of Γ that decrease. The estimates of δ^2 seem to be unaffected by the change of n. However, increasing n increases the precision of all estimates, which is due to an increase in the information content regarding the common dynamics of the series.

Comparing the estimation results between the two tables, we observe that increasing the complexity of the model by allowing to disentangle the factors structure between variances and covariances leads to some changes in the parameter estimates and their precision: the factor describing the common dynamics of the realized variances and covariances becomes more persistent, while the factor typical only to the variances inherits the persistence of the one of FAW(1). While the estimates of Γ increase, the ones of b decrease, but there are no significant changes in the estimates of the degrees of freedom m. The precision of the estimates worsens in general, as FAW(1,1) has more parameters to estimate than FAW(1).

Comparing the values of the log-likelihoods and BIC, we see that for each n, the goodness-of-fit is improved by adding a second factor in the expression of Σ_t. Moreover the values of the Ljung-Box statistics computed up to lag 20 and averaged over all $Y_{ij,t}$ series, although still very large, drastically decrease by adding further factors leading to a drop in their values by six compared to the values corresponding to the original $Y_{ij,t}$ series.

In order to evaluate the forecasting performance of the two models, we apply them to the $n = 5$ and $n = 10$ stocks mentioned above and divide the sample of $T = 3931$ series in an in-sample series of length $T_1 = 3431$ and an out of sample series of $T_2 = 500$. The forecasts are obtained by expanding window: i.e., use the first 3431 observations to estimate the model and make one step ahead forecast $\hat{Y}_{3432|3431}$. Then, add to the previous window one observation, i.e., the in-sample window includes the observations $1, \ldots, 3432$, to estimate the model and make a second one-step ahead forecast $\hat{Y}_{3433|3432}$ and go on until the one before the last observation in the series.

To compare the forecast quality of our new approaches, we also consider a simple alternative, which is standard in most of the comparisons undergone in the literature (Golosnoy et al. (2012), among others), namely the Exponentially Weighted Moving Average (EWMA) model that is the model used by most of the practitioners through the RiskMetrics package of J.P. Morgan (1996) and based on which the forecasts are obtained recursively through the formula:

$$\hat{Y}_{t+1|t} = (1 - \lambda)Y_t + \lambda \hat{Y}_{t-1|t-2}, \tag{15}$$

where $\lambda = 0.94$.

Besides EWMA, we also compare our forecasts against the ones stemming from the CAW model of Golosnoy et al. (2012), which might be seen to be the closest to our approach. In particular, for our purposes we implement the following two CAW specifications:

$$Y_{t+1|t} \sim \mathcal{W}_n \left(m, \frac{\Sigma_{t+1}}{m} \right) \quad , \quad m \geq n \tag{16}$$

where

$$\Sigma_{t+1} = CC' + B\Sigma_t B' + AY_t A' \tag{17}$$

and the model is denoted CAW and another specification composing the HAR model of Corsi (2009), where:

$$\Sigma_{t+1} = CC' + A_1 Y_t A_1' + A_2 Y_t^{5d} A_2' + A_3 Y_t^{20d} A_3', \tag{18}$$

with Y_t^{5d} and Y_t^{20d} being the averages of Y_t over the last 5 and 20, respectively, days and the model is denoted CAW-HAR. The CAW-HAR model aims at capturing the long memory in the series. Given that the number of parameters of these approaches is of order $O(n^2)$, for $n = 10$, we choose to implement only their diagonal version (denoted here by dCAW), i.e. the matrices A, B, A_1, A_2 and A_3 are diagonal. Thus, the total number of parameters of the CAW specifications is equal for $n = 5$ to 66 for CAW and 91 for CAW-HAR, 26 for dCAW and 31 for dCAW-HAR. for $n = 10$, dCAW has 76 and dCAW-HAR has 86 parameters.

To compare the forecasting performance of the models we follow Laurent et al. (2013) and Golosnoy et al. (2012) and use the Root Mean Squared Error (RMSE) computed based on the Frobenius norm[3] of the matrix of forecast errors defined by the difference between the forecast matrix $\hat{Y}_{t+1|t}$ and the true value of the covariance matrix at time $t+1$. Given that this true covariance matrix is not observable, we follow Patton (2011) and use instead the realized value Y_{t+1}. The RMSE loss function is consistent with respect to the choice of the covariance matrix proxy (Laurent et al. (2013)) and is given by:

$$RMSE(\hat{Y}_{t+1|t}) = \frac{1}{T_2} \sum_{t=1}^{T_2} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} (\hat{Y}_{ij,t+1|t} - Y_{i,j,t+1})^2 \right]^{1/2}. \tag{19}$$

To compare the RMSE results, we implement the Diebold-Mariano (DM) test by choosing the FAW model with the smallest RMSE as the benchmark.

Table B.5 presents the forecast results for the two choices of n. As one may observe from the table, FAW(1,1) provides the smallest RMSE for both choices of n. This is not the case of FAW(1), which seems to provide relatively poor forecasts compared to the alternatives and FAW(1,1). Moreover, we reject at 5% significance level that the alternative models provide a RMSE equal to the one of FAW(1,1). The superior performance of FAW(1,1) compared to the alternatives becomes more obvious when increasing n, for which some of the alternatives are not even implementable due to their extremely large parametrization.

5 Conclusions

This paper introduces a very parsimonious framework to model and forecast daily RCov series, denoted the Factor Autoregressive model. It automatically assures the positive definiteness and symmetry of the forecasts and accounts for the common dynamics and long memory of the series without any parametric constraints. It consists in a additive stochastic volatility latent factor structure with Wishart underlying distribution. The model is

[3]The Frobenius norm of a matrix $M = \{m_{kl}\}$ of dimension $s \times q$ is equal to $||M|| = \sqrt{\sum_{k=1}^{s} \sum_{l=1}^{q} m_{kl}^2}$

a non-linear non-Gaussian state-space model that is estimated here by exact Maximum Likelihood with the Kitagawa filtering approach. In the empirical exercise, we show that the model outperforms the existing ones in terms of forecasting in sample and out of sample at very low parametric costs.

The FAW model can be easily extend to account for specific empirical features of the underlying series, such as GARCH effects, especially in the realized variance dynamics, fat-tailedness by assuming a Student's t or a stable distribution within the SV structure of the latent factors or a fat-tailed matrix distribution , such as the matrix F-distribution for the idiosyncratic noises or to add observable factors in the spirit of Fama-French approach. Although empirically very appealing, these extensions are left for further research.

Acknowledgements

For helpful comments we would like to thank Luigi Brugnano, Bastian Gribisch, Ingmar Nolte, James Taylor, Winfried Pohlmeier and participants at the 11th International Conference on Computational and Financial Econometrics (CFE 2017) in December 2017 in London, UK. All remaining errors are ours. Giorgio Calzolari gratefully acknowledges financial support from University of Konstanz through the Graduate School of Decision Sciences. Roxana Halbleib gratefully acknowledges financial support from the European Social Fund, Ministry of Science, Research and Arts of Baden–Württemberg, Germany, from the Deutsche Forschungsgemeinschaft through the PO375/11-1 titled "Robust Risk Measures in Real Time Settings and from the Zukunftskolleg at University of Konstanz. This work was primarily conducted using the computational resource bwUniCluster funded by the Ministry of Science, Research and Arts of Baden-Württemberg and the universities of the State of Baden-Württemberg, Germany, within the bwHPC framework program.

References

Asai, M. and McAleer, M. (2006), 'Asymmetric multivariate stochastic volatility', *Econometric Reviews* **25**, 453–473.

Barndorff-Nielsen, O. E., Hansen, P., 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. (2004), 'Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics', *Econometrica* **72**, 885–925.

Bauwens, L., Braioni, M. and Storti, G. (2017), 'A dynamic component model for forecasting high-dimensional realized covariance matrices', *Econometrics and Statistics* **1**, 40–61.

Bauwens, L., Laurent, S. and Rombouts, J. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Bilio, M. and Sartore, D. (2003), Stochastic volatility models: A survey with applications to option pricing and value at risk, *in* L. Dunis, J. Laws and P. Naim, eds, 'Applied Quantitative Methods for Trading and Investment', New York: John Wiley, pp. 239–291.

Calzolari, G., Halbleib, R. and Zagidullina, A. (2017), A latent factor model for forecasting realized volatilities. GSDS Working Paper No. 2017-14.

Chiriac, R. and Voev, V. (2011), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* **26**, 922–947.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012), 'The conditional autoregressive wishart model for multivariate stock market volatility', *Journal of Econometrics* **167**, 211–223.

Gouriéroux, C., Jasiak, J. and Sufana, R. (2009), 'The wishart autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**, 167–181.

Granger, C. W. J. (1980), 'Long-memory relationships and the aggregation of dynamic models', *Journal of Econometrics* **14**, 227–238.

Jin, X. and Maheu, J. M. (2013), 'Modelling realized covariances and returns', *Journal of Financial Econometrics* **11**, 335–369.

J.P. Morgan (1996), *RiskMetrics - Technical Document*, Morgan Guaranty Trust Company.

Kitagawa, G. (1987), 'Non gaussian state spacemodeling of nonstationary time series', *Journal of the American Statistical Association* **82**, 1032–1041.

Laurent, S., Rombouts, J. and Violante, F. (2013), 'On loss functions and ranking forecasting performances of multivariate volatility models', *Journal of Econometrics* **173**, 1–10.

Lunde, A., Shephard, N. and Sheppard, K. (2016), 'Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice'.

Muirhead, R. (1982), *Aspects of Multivariate Statistical Theory*, Wiley.

Patton, A. (2011), 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics* **160/1**, 246–256.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high frequency data', *Journal of the American Statistical Association* **100**, 1394–1411.

Appendix A: Kitagawa Filter

Here we present the Kitagawa filtering procedure for FAW(1) and FAW(1,1). For simplicity, in what follows we drop $\boldsymbol{\theta}$ from the pdf specifications (Bilio and Sartore (2003)). In what follows, $\overline{Y_t}$ indicates the whole sample of observations till time t: $\overline{Y_t} = Y_1, Y_2, \ldots, Y_t$. Thus $\overline{Y_t} = \{Y_t, \overline{Y_{t-1}}\}$

A.1: FAW(1)

$\mathbf{h_t} = h_t$. Let R be the number of nodes in the grid of values of h_t for numerical integration (e.g., $R = 50$, $R = 100$ or $R = 200$) in such a way that the grid practically covers the entire domain of h_t for any t. As the domain of Gaussian random variables is from $-\infty$ to $+\infty$, the grid must be wide enough to exclude tails with a probability numerically negligible.

Let $p = 1, 2, ..., R$ indicate the nodes in the grid of h_t: $h_t^1, h_t^2, ..., h_t^R$ and $q = 1, 2, ..., R$, indicate the nodes in the grid of h_{t+1}: $h_{t+1}^1, h_{t+1}^2, ..., h_{t+1)}^R$.

$f(Y_t \mid h_t)$ and $f(h_t \mid h_{t-1})$ are known in closed form from the observation equation (1) and the state equation (7), respectively. More precisely, while $f(Y_t \mid h_t)$ is given in Equation (14), $f(h_t \mid h_{t-1})$ is given by:

$$f(h_t \mid h_{t-1}) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{h_t - \rho h_{t-1}}{2\delta^2}}. \tag{A.1}$$

The Kitagawa filtering procedure consists in several recursive steps that repeat for each $t = 1, \ldots, T$. For this, we assume that $f(h_1|Y_0)$ is equal to the unconditional pdf of h_1, i.e. $f(h_1|Y_0) = f(h_1) = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi\delta^2}} e^{-\frac{(1-\rho^2)h_1}{2\delta^2}}$ and that the numerical values of all conditional densities of h_t given $\overline{Y_{t-1}}$ are known at any node of the grid h_t^p. $f(h_t^1 \mid \overline{Y_{t-1}}), \ldots f(h_t^R \mid \overline{Y_{t-1}})$. Thus the steps involved by the Kitagawa filtering are the following:

1. Compute $f(Y_t \mid h_t, \overline{Y_{t-1}})$.

   As Y_t is a function of the past only through h_t, at any node of the grid one can write $f(Y_t \mid h_t^p, \overline{Y_{t-1}}) = f(Y_t \mid h_t^p)$ that can be computed from the observation equation (1) in closed form.

2. Compute $f(Y_t, h_t \mid \overline{y_{t-1}})$

   $f(Y_t, h_t \mid \overline{Y_{t-1}}) = f(Y_t \mid h_t, \overline{Y_{t-1}})\ f(h_t \mid \overline{Y_{t-1}}) = f(Y_t \mid h_t)\ f(h_t \mid \overline{Y_{t-1}})$ whose numerical values at all nodes of the grid $h_t^1, h_t^2, ..., h_t^R$ can be computed as $f(Y_t, h_t^p \mid \overline{Y_{t-1}}) = f(Y_t \mid h_t^p)\ f(h_t^p \mid \overline{Y_{t-1}})$

3. Compute $f(Y_t \mid \overline{Y_{t-1}})$

   Numerical integration of the previous step's function over the R nodes of the grid provides the numerical value of $f(Y_t \mid \overline{Y_{t-1}})$, which can be used to update the likelihood till time t. This step requires only one numerical integration, on a grid of R nodes.

4. Compute $f(h_t \mid \overline{Y_t})$

   From $f(Y_t, h_t \mid \overline{Y_{t-1}}) = f(h_t \mid Y_t, \overline{Y_{t-1}})\ f(Y_t \mid \overline{Y_{t-1}}) = f(h_t \mid \overline{Y_t})\ f(Y_t \mid \overline{Y_{t-1}})$ one gets $f(h_t \mid \overline{Y_t}) = \frac{f(Y_t, h_t \mid \overline{Y_{t-1}})}{f(Y_t \mid \overline{Y_{t-1}})} = \frac{f(Y_t \mid h_t)\ f(h_t \mid \overline{Y_{t-1}})}{f(Y_t \mid \overline{Y_{t-1}})}$ whose numerical values $f(h_t^p \mid \overline{Y_t})$ can therefore be computed at all nodes of the grid $h_t^1, h_t^2, ..., h_t^R$.

5. Compute $f(h_{t+1} \mid h_t, \overline{Y_t})$

   First of all it must be noticed that h_{t+1} depends on $\overline{Y}_t$ *only* through h_t. Thus $f(h_{t+1} \mid h_t, \overline{Y_t}) = f(h_{t+1} \mid h_t)$. For each pair h_{t+1} and h_t the density $f(h_{t+1} \mid h_t)$ is given by Equation (A.1) in closed form. Thus, for each of the $R \times R$ pairs of values (h_{t+1}^q, h_t^p) (from the grids $h_{t+1}^1, h_{t+1}^2, ..., h_{t+1}^R$ and $h_t^1, h_t^2, ..., h_t^R$) the Equation (A.1) provides the numerical value of $f(h_{t+1}^q \mid h_t^p)$. This step requires therefore R^2 computations of the conditional density function provided by the Equation (A.1).

6. Compute $f(h_{t+1}, h_t \mid \overline{Y_t})$

   $f(h_{t+1}, h_t \mid \overline{Y_t}) = f(h_{t+1} \mid h_t, \overline{Y_t})\ f(h_t \mid \overline{Y_t}) = f(h_{t+1} \mid h_t)\ f(h_t \mid \overline{Y_t})$. Thus, for each of the $R \times R$ pairs of values (h_{t+1}^q, h_t^p) (from the grids $h_{t+1}^1, h_{t+1}^2, ..., h_{t+1}^R$ and $h_t^1, h_t^2, ..., h_t^R$), the above equation provides the numerical value of $f(h_{t+1}^q, h_t^p \mid \overline{Y_t})$.

7. Compute $f(h_{t+1} \mid \overline{Y_t})$

   For each value (node) h_{t+1}^q of the grid, numerical integrations of the previous step's function over the R nodes of the grid $h_t^1, h_t^2, ..., h_t^R$ provides the numerical values of $f(h_{t+1}^q \mid \overline{Y_t})$, which is therefore available for each node of the grid $h_{t+1}^1, h_{t+1}^2, ..., h_{t+1}^R$. This step requires, therefore, R numerical integration, each of which on a grid of R nodes.

8. Restart from the first step

   Time, data, likelihood and the grid are "shifted" from $(t-1, t)$ to $(t, t+1)$. Recursions continue till $t = T$. At $t = T$, the complete likelihood is available.

A.2: FAW(1,1)

$\mathbf{h_t} = (h_{1,t}, h_{2,t})'$. Let R be the number of nodes in the grid of values of $h_{1,t}$, as well as in the grid of values of $h_{2,t}$, for numerical integration (for instance, the largest value could be $R = 100$, thus $R \times R = 10000$) such that the grids cover the entire domains of $h_{1,t}$ and $h_{2,t}$ for any t(in practice, as before, the grids should be wide enough to exclude tails with a probability numerically negligible);

Let $p = 1, 2, ..., R$ indicate the nodes in the grid of $h_{1,t}$, $r = 1, 2, ..., R$ indicate the nodes in the grid of $h_{2,t}$; $q = 1, 2, ..., R$ indicate the nodes in the grid of $h_{1,t+1}$ and $s = 1, 2, ..., R$ indicate the nodes in the grid of $h_{2,t+1}$.

$f(Y_t \mid h_{1,t}, h_{2,t})$ is given in Equation (14), $f(h_{1,t} \mid h_{1,t-1})$ and $f(h_{2,t} \mid h_{2,t-1})$ are obtained from equations (9) and (10), respectively:

$$f(h_{1,t} \mid h_{1,t-1}) = \frac{1}{\sqrt{2\pi\delta_1^2}} e^{-\frac{h_{1,t}-\rho_1 h_{1,t-1}}{2\delta_1^2}} \tag{A.2}$$

$$f(h_{2,t} \mid h_{2,t-1}) = \frac{1}{\sqrt{2\pi\delta_2^2}} e^{-\frac{h_{2,t}-\rho_2 h_{2,t-1}}{2\delta_2^2}} \tag{A.3}$$

Given that $h_{1,t}$ and $h_{2,t}$ are independent, one can derive $f(h_{1,t}, h_{2,t} \mid h_{1,t-1}, h_{2,t-1}) = f(h_{1,t} \mid h_{1,t-1}) f(h_{2,t} \mid h_{2,t-1})$.

When starting the Kitagawa filter at time t, one should know the numerical values of all joint conditional densities of $(h_{1,t}, h_{2,t})$ given $\overline{Y_{t-1}}$ at any of the $R \times R$ nodes of the bi-dimensional grid $\{h_{1,t}^p, h_{2,t}^r\}$: $f(h_{1,t}^p, h_{2,t}^r \mid \overline{Y_{t-1}})$ at $(h_{1,t}^1, h_{2,t}^1)$, $(h_{1,t}^1, h_{2,t}^2)$, ..., $(h_{1,t}^1, h_{2,t}^R)$, ..., $(h_{1,t}^2, h_{2,t}^1)$, ..., $(h_{1,t}^R, h_{2,t}^R)$.

The Kitagawa filtering procedure consists in the following steps:

1. Compute $f(Y_t \mid h_{1,t}, h_{2,t}, \overline{Y_{t-1}})$

   As Y_t is a function of the past only through $h_{1,t}$ and $h_{2,t}$, at any node of the bi-dimensional grid it is $f(Y_t \mid h_{1,t}^p, h_{2,t}^r, \overline{Y_{t-1}}) = f(Y_t \mid h_{1,t}^p, h_{2,t}^r)$, and it can therefore be computed from the observation equation (1).

2. Compute $f(Y_t, h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}})$

   $f(Y_t, h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}}) = f(Y_t \mid h_{1,t}, h_{2,t}, \overline{Y_{t-1}})\ f(h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}})$, whose numerical values can be computed in any of the $R \times R$ nodes of the bi-dimensional grid $\{h_{1,t}^p, h_{2,t}^r\}$ as $f(Y_t, h_{1,t}^p, h_{2,t}^r \mid \overline{Y_{t-1}}) = f(Y_t \mid h_{1,t}^p, h_{2,t}^r) f(h_{1,t}^p, h_{2,t}^r \mid \overline{Y_{t-1}})$

3. Compute $f(Y_t \mid \overline{Y_{t-1}})$

   Numerical integration of $f(Y_t, h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}})$ over the $R \times R$ nodes of the bi-dimensional grid $\{h_{1,t}^p, h_{2,t}^r\}$ provides the numerical value of $f(Y_t \mid \overline{Y_{t-1}})$. It can be used to update the likelihood till time t.

4. Compute $f(h_{1,t}, h_{2,t} \mid \overline{Y_t})$

   From $f(Y_t, h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}}) = f(h_{1,t}, h_{2,t} \mid Y_t, \overline{Y_{t-1}})\ f(Y_t \mid \overline{Y_{t-1}}) = f(h_{1,t}, h_{2,t} \mid \overline{Y_t})\ f(Y_t \mid \overline{Y_{t-1}})$

   one gets

   $f(h_{1,t}, h_{2,t} \mid \overline{Y_t}) = \frac{f(Y_t, h_{1,t}, h_{2,t} \mid \overline{Y_{t-1}})}{f(Y_t \mid \overline{Y_{t-1}})}$

   whose numerical values $f(h_{1,t}^p,\ h_{2,t}^r \mid \overline{Y_t})$ can therefore be computed at all nodes of the bi-dimensional grid $\{h_{1,t}^p, h_{2,t}^r\}$. It must be remarked that, even if the two factors are independent, *conditionally* on $\overline{Y_t}$ they are usually *not* independent.

5. Compute $f(h_{1,t+1}, h_{2,t+1} \mid h_{1,t}, h_{2,t})$

   For each pair $(h_{1,t+1}, h_{2,t+1})$ and each pair $(h_{1,t}, h_{2,t})$ the joint conditional density $f(h_{1,t+1}, h_{2,t+1} \mid h_{1,t}, h_{2,t}) = f(h_{1,t+1} \mid h_{1,t}) f(h_{2,t+1} \mid h_{2,t}))$ where $f(h_{1,t+1} \mid h_{1,t})$ and $f(h_{2,t+1} \mid h_{2,t}))$ are given in equations (A.2) and (A.3). Thus, R^4 values of this function should be computed.

6. Compute $f(h_{1,t+1}, h_{2,t+1}, h_{1,t}, h_{2,t} \mid \overline{Y_t})$

   $f(h_{1,t+1}, h_{2,t+1}, h_{1,t}, h_{2,t} \mid \overline{Y_t}) = f(h_{1,t+1}, h_{2,t+1} \mid h_{1,t}, h_{2,t}, \overline{Y_t})\ f(h_{1,t}, h_{2,t} \mid \overline{Y_t}) = f(h_{1,t+1}, h_{2,t+1} \mid h_{1,t}, h_{2,t})\ f(h_{1,t}, h_{2,t} \mid \overline{Y_t})$.

   Thus, for each of the R^4 quadruples of values $(h^q_{1,t+1}, h^s_{2,t+1}, h^p_{1,t}, h^r_{2,t})$ the above equation provides the numerical value of $f(h^q_{1,t+1}, h^s_{2,t+1}, h^p_{1,t}, h^r_{2,t} \mid \overline{Y_t})$.

7. Compute $f(h_{1,t+1}, h_{2,t+1} \mid \overline{Y_t})$

   Numerical integration of the previous step's function over the $R \times R$ nodes of the bi-dimensional grid $\{h^p_{1,t}, h^r_{2,t}\}$ provides the numerical values of $f(h^q_{1,t+1}, h^s_{2,t+1} \mid \overline{Y_t})$ that will, therefore, be available for each node of the bi-dimensional grid $\{h^q_{1,t+1}, h^s_{2,t+1}\}$.

   This step would involve a quite large number of function evaluation (the order would be R^4). Given that the two latent factors $h_{1,t}$ and $h_{2,t}$ are independent, the computational procedure can be done more efficiently, with a number of function evaluations of order R^3 as it follows:

   7.1 Compute $f(h_{2,t+1} \mid h_{1,t}, h_{2,t}, \overline{Y_t})$

   When the two factors are independent, the above conditional density is simply given by the state equation of the second factor: $f(h_{2,t+1} \mid h_{1,t}, h_{2,t}, \overline{Y_t}) = f(h_{2,t+1} \mid h_{1,t}, h_{2,t}) = f(h_{2,t+1} \mid h_{2,t})$.

   7.2 Compute $f(h_{2,t+1}, h_{1,t}, h_{2,t} \mid \overline{Y_t})$

$$\begin{aligned} f(h_{2,t+1}, h_{1,t}, h_{2,t} \mid \overline{Y_t}) &= f(h_{2,t+1} \mid h_{1,t}, h_{2,t}, \overline{Y_t})\ f(h_{1,t}, h_{2,t} \mid \overline{Y_t}) \\ &= f(h_{2,t+1} \mid h_{2,t})\ f(h_{1,t}, h_{2,t} \mid \overline{Y_t}). \end{aligned}$$

   Thus, for each of the R^3 triples of values $(h^s_{2,t+1}, h^p_{1,t}, h^r_{2,t})$ the above equation provides the numerical value of $f(h^s_{2,t+1}, h^p_{1,t}, h^r_{2,t} \mid \overline{Y_t})$.

   7.3 Compute $f(h_{2,t+1}, h_{1,t} \mid \overline{Y_t})$

   Numerical integration of the previous step's function over the one-dimensional grid $\{h^r_{2,t}\}$ provides the numerical values of $f(h^s_{2,t+1}, h^p_{1,t} \mid \overline{Y_t})$, that will therefore be available for each node of the bi-dimensional grid $\{h^s_{2,t+1}, h^p_{1,t}\}$.

   7.4 Compute $f(h_{1,t+1}, h_{2,t+1}, h_{1,t} \mid \overline{Y_t})$

   $f(h_{1,t+1}, h_{2,t+1}, h_{1,t} \mid \overline{Y_t}) = f(h_{1,t+1} \mid h_{2,t+1}, h_{1,t}, \overline{Y_t})\ f(h_{2,t+1}, h_{1,t} \mid \overline{Y_t})$. Given that the two factors are independent, and $h_{1,t+1}$ depends on $\overline{Y_t}$ only through $h_{1,t}$, $f(h_{1,t+1}, h_{2,t+1}, h_{1,t} \mid \overline{Y_t}) = f(h_{1,t+1} | h_{1,t})\ f(h_{2,t+1}, h_{1,t} \mid \overline{Y_t})$.

   7.5 Compute $f(h_{1,t+1}, h_{2,t+1} \mid \overline{Y_t})$

   Numerical integration of the previous step's function over the one-dimensional grid $\{h^p_{1,t}\}$ provides the numerical values of $f(h^q_{1,t+1}, h^s_{2,t+1} \mid \overline{Y_t})$, that will, therefore, be available for each node of the bi-dimensional grid $\{h^q_{1,t+1}, h^s_{2,t+1}\}$.

8. Restart from the first step.

   Time, data, likelihood and the grids are "shifted" from $(t-1, t)$ to $(t, t+1)$. Recursions continue till $t = T$. At $t = T$, the complete likelihood is available.

A.3: Computational Issues

Numerical integration is necessary in many steps of the filtering procedure presented above, as a closed form integration is usually unfeasible. As in our approach we deal with "Gaussian type" probability distributions, the Gauss-Hermite integration is expected to provide very accurate numerical values for all the integrals. Its application, however, is not straightforward, given the particular form of the functions to be integrated, whose "location and width" change "unpredictably" over time. Parallel to the Gauss-Hermite, we, therefore, implement numerical integration procedures of a standard "trapezoid rule". They are remarkably simpler from the computational point of view, so we use them in the algorithm. Of course, trapezoid integration leads to higher costs in terms of computation time, due to more "grid points" necessary to ensure the same numerical accuracy as Gauss-Hermite. In all cases, where both procedures are applied, the differences in the results are only after four decimal digits after the comma of those reported in our tables.

Fixing the grid for the trapezoid integration rule requires some care in order to assure that the grid reasonably covers the entire domain of the integration variable. Dealing with Gaussian-type distributions, the domain of integration is conceptually infinite. The integration grid should be wide enough to exclude tails with a negligible probability. For this, we adopt the rule of excluding tails, where the probability density is less than 10^{-15} with respect to the maximum value of the density in the integration interval. In most cases, an integration interval between minus and plus 8 times the unconditional standard deviation of each factor is more than sufficient.

Appendix B: Tables

Table B.1: **Monte Carlo results from estimating FAW(1)**: $n = 5, 8, 10$ series, $T = 4000$ observations and $R = 1000$ replications.

Parameters	$n = 5$			$n = 8$			$n = 10$		
	True	Mean	Std. dev	True	Mean	Std. dev	True	Mean	Std. dev
ρ	0.9600	0.9595	0.0053	0.9400	0.9394	0.0057	0.9400	0.9395	0.0050
δ^2	0.1000	0.1001	0.0032	0.1000	0.0999	0.0028	0.1000	0.1000	0.0023
m	15.0000	15.0032	0.0781	20.0000	20.0018	0.0617	25.0000	25.0056	0.0585
Γ	2.0000	2.0020	0.0382	2.0000	2.0011	0.0290	2.0000	1.9993	0.0229
	2.0000	2.0016	0.0338	2.0000	2.0006	0.0275	2.0000	1.9996	0.0213
	2.0000	2.0013	0.0345	2.0000	2.0000	0.0267	2.0000	1.9993	0.0215
	2.0000	2.0015	0.0337	2.0000	2.0006	0.0273	2.0000	1.9999	0.0215
	2.0000	2.0021	0.0340	2.0000	2.0001	0.0273	2.0000	1.9993	0.0218
				2.0000	2.0002	0.0277	2.0000	1.9994	0.0219
				2.0000	2.0003	0.0265	2.0000	1.9995	0.0215
				2.0000	2.0003	0.0272	2.0000	1.9993	0.0219
							2.0000	1.9998	0.0213
							2.0000	1.9997	0.0217
b	1.0000	1.0004	0.0106	1.0000	1.0000	0.0083	1.0000	0.9998	0.0065
	1.0000	1.0006	0.0103	1.0000	0.9999	0.0081	1.0000	0.9998	0.0068
	1.0000	1.0002	0.0107	1.0000	1.0002	0.0082	1.0000	0.9999	0.0062
	1.0000	1.0004	0.0109	1.0000	1.0002	0.0081	1.0000	1.0000	0.0066
				1.0000	1.0001	0.0084	1.0000	0.9999	0.0067
				1.0000	1.0003	0.0084	1.0000	0.9999	0.0065
				1.0000	1.0000	0.0082	1.0000	0.9999	0.0064
							1.0000	0.9999	0.0065
							1.0000	0.9999	0.0064

Table B.2: **Monte Carlo results from estimating FAW(1,1)**: $n = 5, 8, 10$ series, $T = 4000$ observations and $R = 1000$ replications.

Parameters	$n = 5$			$n = 8$			$n = 10$		
	True	Mean	Std. dev	True	Mean	Std. dev	True	Mean	Std. dev
ρ_1	0.9800	0.9793	0.0043	0.9700	0.9695	0.0041	0.9700	0.9694	0.0043
ρ_2	0.9500	0.9489	0.0053	0.9400	0.9388	0.0057	0.9400	0.9387	0.0058
δ_1^2	0.1000	0.0996	0.0095	0.1000	0.1002	0.0052	0.1000	0.0998	0.0055
δ_2^2	0.1000	0.1000	0.0033	0.1000	0.1000	0.0027	0.1000	0.1000	0.0027
m	15.0000	14.9947	0.1093	20.0000	19.9986	0.0666	20.0000	19.9998	0.0491
Γ	1.0000	1.0016	0.0979	1.0000	1.0058	0.0853	1.0000	1.0016	0.0837
	1.0000	1.0013	0.0981	1.0000	1.0061	0.0857	1.0000	1.0016	0.0834
	1.0000	1.0015	0.0980	1.0000	1.0061	0.0858	1.0000	1.0017	0.0834
	1.0000	1.0017	0.0980	1.0000	1.0063	0.0857	1.0000	1.0017	0.0836
	1.0000	1.0019	0.0977	1.0000	1.0062	0.0855	1.0000	1.0021	0.0836
				1.0000	1.0059	0.0856	1.0000	1.0023	0.0835
				1.0000	1.0059	0.0859	1.0000	1.0021	0.0837
				1.0000	1.0060	0.0854	1.0000	1.0020	0.0835
							1.0000	1.0016	0.0833
							1.0000	1.0020	0.0837
b	1.0000	1.0000	0.0027	1.0000	1.0001	0.0028	1.0000	1.0000	0.0028
	1.0000	1.0001	0.0028	1.0000	1.0000	0.0028	1.0000	1.0001	0.0027
	1.0000	1.0000	0.0026	1.0000	0.9999	0.0028	1.0000	1.0002	0.0028
	1.0000	1.0000	0.0027	1.0000	1.0000	0.0028	1.0000	1.0002	0.0028
				1.0000	1.0000	0.0028	1.0000	1.0002	0.0027
				1.0000	1.0000	0.0029	1.0000	1.0001	0.0027
				1.0000	1.0000	0.0028	1.0000	1.0002	0.0028
							1.0000	1.0002	0.0027
							1.0000	1.0001	0.0028

Table B.3: **Results from estimating FAW(1)** from realized covariance matrices computed from the first $n = 5, 8, 10$ series over the window 01.11.2001 until 08.08.2017 ($T = 3931$).

Parameters	$n = 5$		$n = 8$		$n = 10$	
	Estimate	Std. dev	Estimate	Std. dev	Estimate	Std. dev
ρ	0.9620	0.0045	0.9386	0.0056	0.9371	0.0057
δ^2	0.0803	0.0028	0.0804	0.0024	0.0809	0.0022
m	15.5701	0.0816	19.9202	0.0636	23.4778	0.0597
Γ	4.5225	0.1013	2.7317	0.0435	2.7629	0.0395
	1.4873	0.0326	1.0260	0.0155	1.0220	0.0140
	2.1567	0.0467	1.3341	0.0202	1.2772	0.0175
	1.3759	0.0302	1.2729	0.0191	1.3584	0.0186
	1.5165	0.0332	1.4031	0.0211	1.4790	0.0203
			1.3234	0.0203	1.2989	0.0180
			1.1163	0.0170	1.1122	0.0153
			1.0400	0.0160	1.0178	0.0141
					1.1877	0.0163
					0.8665	0.0119
b	1.0006	0.0114	0.8690	0.0071	0.8706	0.0064
	0.7347	0.0098	0.7061	0.0065	0.7255	0.0059
	1.3270	0.0147	1.0222	0.0084	1.0074	0.0075
	1.3928	0.0154	1.0693	0.0088	1.0499	0.0078
			0.9033	0.0074	0.9171	0.0068
			0.6477	0.0059	0.6576	0.0054
			0.7873	0.0065	0.7989	0.0059
					0.6959	0.0057
					0.7816	0.0058
Nr. of parameters	12		18		22	
Log-likelihood	-9.4850		-11.4239		-7.4428	
BIC	18.9954		22.8856		14.9318	
Average LB-statistics(20) $Y_{ij,t}$ series	20016		20397		19567	
Average LB-statistics(20) residuals	8379		10016		9434	

Table B.4: **Results from estimating FAW(1,1)** from realized covariance matrices computed from the first $n = 5, 8, 10$ series over the window 01.11.2001 until 08.08.2017 ($T = 3931$).

Parameters	$n = 5$		$n = 8$		$n = 10$	
	Estimate	Std. dev	Estimate	Std. dev	Estimate	Std. dev
ρ_1	0.9809	0.0033	0.9700	0.0042	0.9622	0.0047
ρ_2	0.9505	0.0052	0.9358	0.0058	0.9347	0.0059
δ_1^2	0.0777	0.0050	0.0791	0.0047	0.0990	0.0052
δ_2^2	0.0741	0.0028	0.0770	0.0024	0.0778	0.0022
m	16.3750	0.0878	20.4789	0.0663	24.0805	0.0620
Γ	2.7991	0.2451	2.4203	0.1668	2.4449	0.1650
	0.9444	0.0827	0.9067	0.0625	0.9061	0.0612
	1.3338	0.1168	1.1847	0.0816	1.1327	0.0764
	0.8098	0.0712	1.1145	0.0769	1.2066	0.0815
	0.9381	0.0824	1.2357	0.0852	1.3199	0.0891
			1.1839	0.0816	1.1567	0.0781
			0.9876	0.0680	0.9825	0.0663
			0.9294	0.0641	0.9048	0.0611
					1.0504	0.0709
					0.7693	0.0519
b	1.0206	0.0114	0.8847	0.0071	0.8799	0.0063
	0.7173	0.0095	0.7045	0.0064	0.7251	0.0058
	1.4357	0.0156	1.0566	0.0085	1.0254	0.0074
	1.4706	0.0159	1.0990	0.0089	1.0633	0.0077
			0.9016	0.0073	0.9165	0.0066
			0.6500	0.0058	0.6608	0.0053
			0.7862	0.0064	0.7994	0.0058
					0.6992	0.0056
					0.7889	0.0057
Nr. of parameters	14		20		24	
Log-likelihood	-9.1984		-11.0022		-6.8088	
BIC	18.4220		22.0424		13.6639	
Average LB-statistics(20) $Y_{ij,t}$ series	20016		20397		19567	
Average LB-statistics(20) residuals	6302		4442		3808	

Table B.5: **Forecasting results:** RMSE and the p-values of DM test with FAW(1,1) as the benchmark: out of sample window from 07.08.2015 until 08.08.2017 ($T_2 = 500$)

Model	$n = 5$		$n = 10$	
	RMSE	DM p-value	RMSE	DM p-value
FAW(1)	3.9024	0.0000	5.4783	0.0000
FAW(1,1)	3.2191	-	4.3817	-
EWMA	3.5556	0.0000	4.9043	0.0000
CAW	3.3814	0.0041	n.a.	n.a.
CAW-HAR	3.3155	0.0210	n.a.	n.a.
dCAW	3.3752	0.0032	5.5056	0.0000
dCAW-HAR	3.3234	0.0164	4.5667	0.0166

Appendix C: Figures

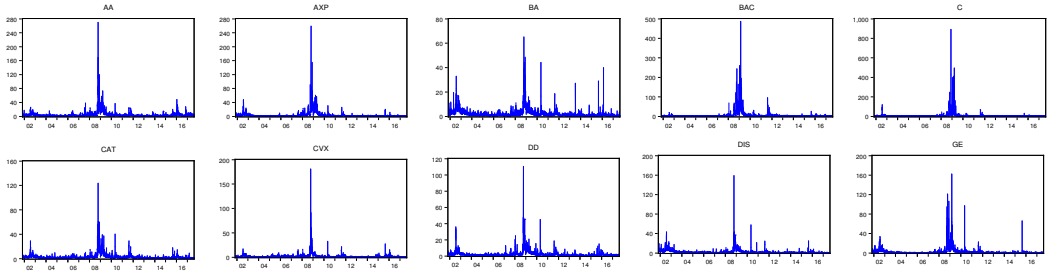

Figure C.1: **Line graph of daily realized variances** over the period 01.11.2001 until 08.08.2017 ($T = 3931$ trading days). On X-axis we plot the years.

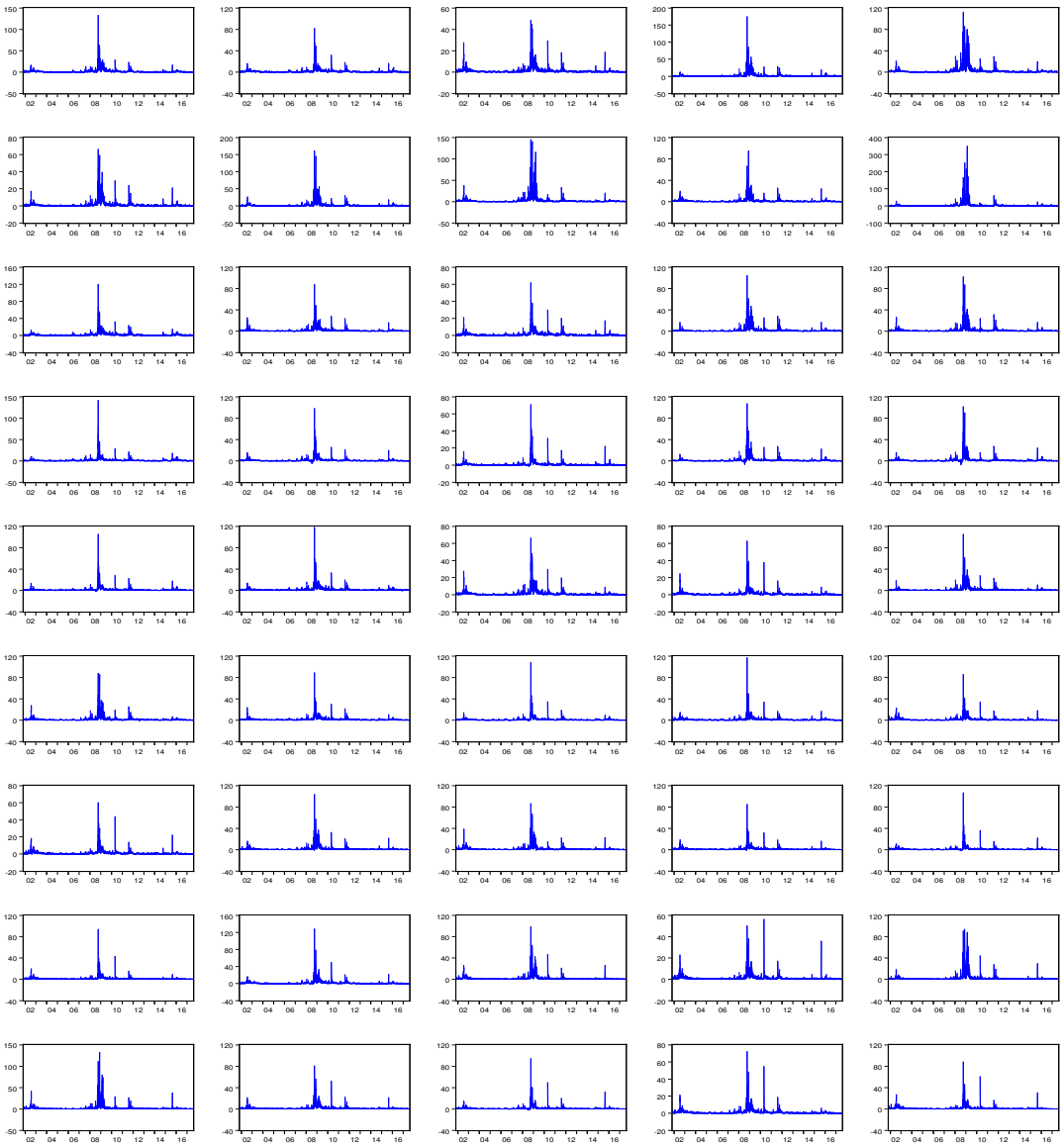

Figure C.2: **Line graph of daily realized covariances** over the period 01.11.2001 until 08.08.2017 ($T = 3931$ trading days). On X-axis we plot the years.

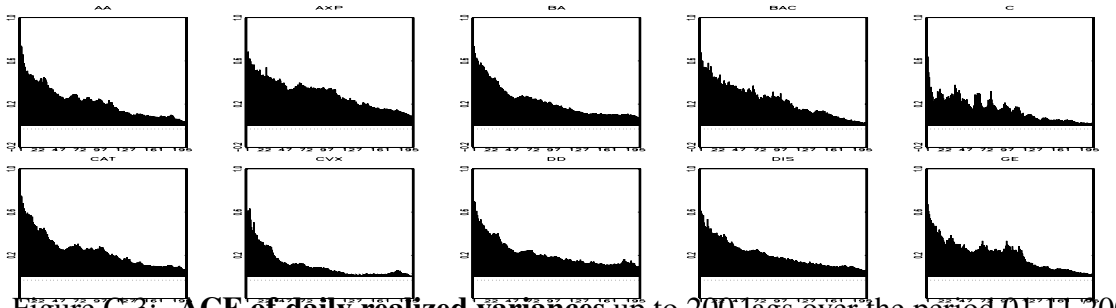

Figure C.3: **ACF of daily realized variances** up to 200 lags over the period 01.11.2001 until 08.08.2017 ($T = 3931$ trading days). On X-axis we plot the lags. The dotted-line is the 95% confidence interval.

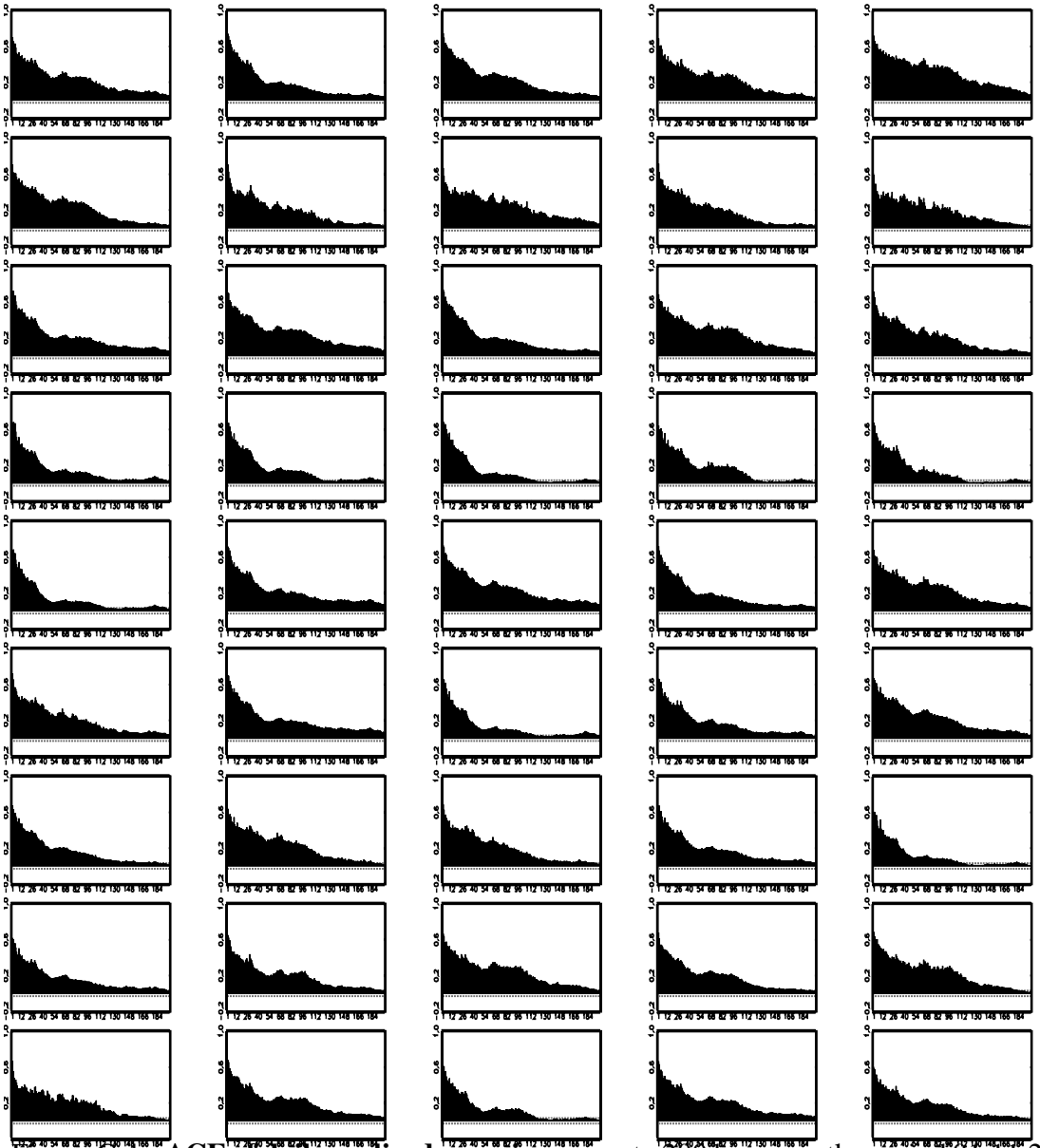

Figure C.4: **ACF of daily realized covariances** up to 200 lags over the period 01.11.2001 until 08.08.2017 ($T = 3931$ trading days). On X-axis we plot the lags. The dotted-line is the 95% confidence interval.