DEMS
Department of Economics, Management and Statistics

# DEMS WORKING PAPER SERIES

## Estimating high dimensional multivariate stochastic volatility models

Matteo Pelagatti and Giacomo Sbrana

No. 428 – January 2020

Department of Economics, Management and Statistics
University of Milano – Bicocca
Piazza Ateneo Nuovo 1 – 2016 Milan, Italy
http://dems.unimib.it/

# Estimating high dimensional multivariate stochastic volatility models*

Matteo Pelagatti†
Department of Economics, Management and Statistics
University of Milano - Bicocca

Giacomo Sbrana
Department of Information Systems, Supply Chain & Decision Making
NEOMA Business School

**Abstract**

This paper proposes tree main results that enable the estimation of high dimensional multivariate stochastic volatility models. The first result is the closed-form steady-state Kalman filter for the multivariate AR(1) plus noise model. The second result is an accelerated EM algorithm for parameters estimation. The third result is an estimator of the correlation of two elliptical random variables with time-varying variances that is consistent and asymptotically normal regardless of the variances evolution. Speed and precision of our methodology are evaluated in a simulation experiment. Finally, we implement our method and compare its performance with other approaches in a minimum variance portfolio composed by the constituents of the CAC40 and S&P100 indexes.

*Keywords:* Riccati equation, EM algorithm, Kalman filter, Correlation estimation, Large covariance matrix, Multivariate stochastic volatility

*We greatly acknowledge the DEMS Data Science Lab for supporting this work by providing computational resources.

†Corresponding author. Email: matteo.pelagatti@unimib.it, tel: +39 02 6448 5834, address: Via Bicocca degli Arcimboldi 8, Milano, I-20126, Italy.

# 1 Introduction

Models for time series with time-varying variances and covariances have become very popular in finance because they capture features that are typical of many asset returns, such as volatility clustering and heavy tails. Two main econometric approaches emerged in the literature: ARCH-type models (Engle, 1982; Bollerslev, 1986; Nelson, 1991) and stochastic volatility (SV) models (Hull and White, 1987; Heston, 1993; Harvey et al., 1994; Alizadeh et al., 2002). ARCH-type models have the advantage of being simple to estimate, while SV models are harder to implement but are backed by financial theory.

For a long period of time, the multivariate versions of both models have been only moderately successful since their estimation on realistically large portfolios of assets was not feasible. Indeed, the number of parameters tends to grow more quickly than the number of observations (curse of dimensionality) and the numerical maximisation of the likelihood function with respects to hundreds or thousands of parameters using standard algorithms is deemed to fail.

For ARCH-type models a convincing solution to the curse of dimensionality was proposed by Engle (2002) with his Dynamic Conditional Correlation (DCC) model. The multi-step estimation approach proposed in that article allows the DCC model to be applicable to portfolios of hundreds of assets in only few minutes of computation time on a PC. No equally successful model has been introduced in the SV world. To the best of our knowledge, the only work that tries to solve the curse of dimensionality in multivariate SV models is that of Chib et al. (2006). Indeed, the authors of that article reduce the dimensionality using a factor model and carry out Bayesian estimation of the unknown quantities using a combination of Markov Chain Monte Carlo and particle filters. However, implementing their algorithm is rather complex and, by extrapolating the information on computation times reported in the article, one can expect that on a modern workstation the application of their method to 100 assets can take few hours.

In this work we derive three general results that can be jointly used to estimate high dimensional multivariate SV models cast in linear state-space form such as the one in Harvey et al. (1994, Sec. 3), from now on HRS, and a multivariate extension of the one in Alizadeh et al. (2002), from now on ABD.

Let us call $d$ the number of time series and $n$ their (common) length. In a multivariate time series set-up, the problem with the approaches of HRS and ABD is that quasi maximum-likelihood (QML) estimation faces numerical issues and computation time exploding with the dimension. Indeed, each pass of the Kalman filter implies, for every time point $t \in \{1, 2, \dots, n\}$, sums, multiplications and an inversion of $d \times d$ matrices. Thus, for large $d$ the computational burden becomes too expensive. Furthermore, the typical quasi-Newton optimisers used to maximise the log-likelihood function become very unstable when the number of parameters is very large (for a portfolio with $d = 100$ assets the parameters to estimate are more than 10,000!).

Our solution to the aforementioned issues in estimating large HRS and ABD models is based on three results. Firstly, we substitute the Kalman filter recursions with the steady-state Kalman filter, which we obtain in closed-form for the multivariate AR(1) plus noise model (i.e., we derive the analytical solution to the implied Riccati equation). This approximation does not harm the asymptotic properties of the estimates of the parameters and reduces by a factor of $n$ the number of operations on the $d \times d$ matrices of the regular Kalman filter. Secondly, we design an expectation maximisation (EM) algorithm based on the steady-state filter and smoother to be used in substitution of quasi-Newton optimisers. Our algorithm is numerically very stable and, as any EM algorithm, it moves very quickly towards a neighbourhood of the solution. Finally, we propose a simple estimator of the correlation between returns that is consistent and asymptotically normal regardless of the evolution of the variances, provided that the returns are drawn from elliptical distributions with time invariant correlations. This last result allows the quick estimation of one

of the two large covariance matrices of the multivariate SV models, which, as it will be clear later, is only a deterministic transformation of the correlation matrix of the returns.

The paper is organized as follows: Section 2 contains the main results; Section 3 introduces the two types of multivariate SV models (i.e., HRS and a multivariate extension of ABD) and illustrates the use of our results to build the estimation procedure; Section 4 applies our procedure to the returns of the CAC40 and S&P100 constituents and carries out a comparison with the DCC model, the RiskMetrics methodology and with the historical covariance matrix approach in the minimum variance portfolio framework; Section 5 concludes. The Appendix contains the proofs of the theorems and other side-results.

# 2 Main results

## 2.1 Steady state Kalman filter and smoother

Consider the multivariate AR(1) plus noise process,

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \text{WN}(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon), \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\kappa} + \phi \boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{WN}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{1}$$

where $\text{WN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a sequence of serially uncorrelated vectors with time-invariant mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\kappa}$ is a vector of constants and $\phi$ is a scalar parameter taking values in the closed interval $[-1, 1]$. In what follows it is assumed that $\mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\eta}_s^\top) = \boldsymbol{0}$ for all $s, t \in \{1, \ldots, n\}$. All vectors and matrices in (1) have dimension $d$ and $d \times d$, respectively.

Let us name $\boldsymbol{a}_t$ the projection of the state vector $\boldsymbol{\alpha}_t$ onto the linear span of the observations $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}\}$,

$$\boldsymbol{a}_t = \mathbb{P}[\boldsymbol{\alpha}_t | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}]$$

and $\boldsymbol{P}_t$ its mean squared error (MSE),

$$\boldsymbol{P}_t = \mathbb{E}[\boldsymbol{\alpha}_t - \boldsymbol{a}_t][\boldsymbol{\alpha}_t - \boldsymbol{a}_t]^\top,$$

and assume that $\boldsymbol{\alpha}_1$ is a random vector with mean $\boldsymbol{a}_1$ and covariance matrix $\boldsymbol{P}_1$. Then, the Kalman filter recursions for this model can be written as follows: for $t = 1, 2, \ldots, n$,

**Innovation** : $\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{a}_t$,

**Innovation variance** : $\boldsymbol{F}_t = \boldsymbol{P}_t + \boldsymbol{\Sigma}_\epsilon$ ,

**Kalman gain** : $\boldsymbol{K}_t = \phi \boldsymbol{P}_t \boldsymbol{F}_t^{-1}$,

**Prediction** : $\boldsymbol{a}_{t+1} = \phi \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{v}_t$,

**Prediction error** : $\boldsymbol{P}_{t+1} = \phi^2 \boldsymbol{P}_t - \phi^2 \boldsymbol{P}_t \boldsymbol{F}_t^{-1} \boldsymbol{P}_t + \boldsymbol{\Sigma}_\eta$.

As $t$ diverges, the sequence of covariance matrices $\boldsymbol{P}_t$ converges to the steady-state matrix $\boldsymbol{P}$ that solves the *algebraic Riccati equation*,

$$\boldsymbol{P} = \phi^2 \boldsymbol{P} - \phi^2 \boldsymbol{P} (\boldsymbol{P} + \boldsymbol{\Sigma}_\epsilon)^{-1} \boldsymbol{P} + \boldsymbol{\Sigma}_\eta, \tag{2}$$

where $\boldsymbol{P}$ is symmetric positive definite[1]. The following result provides the closed-form solution of equation (2) for the state-space form (1).

[1]When $|\phi| < 1$, then the system is stable and the exponentially fast convergence to the steady-state holds because of Result 3.3.1 in Harvey (1989). When $\phi = 1$, then Result 3.3.2 in Harvey (1989) applies since the system is *controllable* (eq. 3.3.4 in Prof. Harvey's book) and *observable* (eq. 3.3.5 in the same volume) and these two properties imply the properties of *stabilisability* and *detectability*, which, in turn, guarantee the exponentially fast convergence to the steady-state.

**Theorem 1** (Riccati equation solution)**.** *Consider the system as in (1) with* $\boldsymbol{\Sigma}_\eta$ *positive semi-definite and* $\boldsymbol{\Sigma}_\epsilon$ *strictly positive definite. Apply the Cholesky decomposition to* $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{M}\boldsymbol{M}^\top$ *and the eigen-decomposition to* $\boldsymbol{Q} = \boldsymbol{M}^{-1}\boldsymbol{\Sigma}_\eta(\boldsymbol{M}^{-1})^\top = \boldsymbol{\Psi}\boldsymbol{\Delta}\boldsymbol{\Psi}^\top$*, with* $\boldsymbol{\Psi}$ *matrix of eigenvectors and* $\boldsymbol{\Delta} = \text{diag}(\delta_1, \delta_2, \cdots, \delta_d)$ *diagonal matrix of eigenvalues.*

*Then the unique positive definite solution for* $\boldsymbol{P}$ *is*

$$\boldsymbol{P} = \frac{1}{2}\boldsymbol{M}\boldsymbol{\Psi}\left\{\boldsymbol{\Delta} - \boldsymbol{I} + \phi^2\boldsymbol{I} + \left[(\boldsymbol{\Delta} - \boldsymbol{I} + \phi^2\boldsymbol{I})^2 + 4\boldsymbol{\Delta}\right]^{\frac{1}{2}}\right\}\boldsymbol{\Psi}^\top\boldsymbol{M}^\top. \tag{3}$$

The knowledge of the steady-state solution leads to considerable computational savings since some of the Kalman filter recursions become redundant (see discussions in Harvey, 1989; Durbin and Koopman, 2001). The analytical expressions in (3) provides also the steady-state solutions for $\boldsymbol{F}$ and $\boldsymbol{K}$. Indeed, given (3), it follows immediately that the steady-state innovations variance $\boldsymbol{F}$ is:

$$\boldsymbol{F} = \boldsymbol{P} + \boldsymbol{\Sigma}_\epsilon = \frac{1}{2}\boldsymbol{M}\boldsymbol{\Psi}\left\{\boldsymbol{\Delta} + \boldsymbol{I} + \phi^2\boldsymbol{I} + [(\boldsymbol{\Delta} - \boldsymbol{I} + \phi^2\boldsymbol{I})^2 + 4\boldsymbol{\Delta}]^{\frac{1}{2}}\right\}\boldsymbol{\Psi}^\top\boldsymbol{M}^\top. \tag{4}$$

Moreover the Kalman gain in the steady-state is

$$\boldsymbol{K} = \phi\boldsymbol{P}\boldsymbol{F}^{-1} = \frac{1}{2\phi}\boldsymbol{M}\boldsymbol{\Psi}\left\{\phi^2\boldsymbol{I} - \boldsymbol{\Delta} - \boldsymbol{I} + [(\boldsymbol{\Delta} - \boldsymbol{I} + \phi^2\boldsymbol{I})^2 + 4\boldsymbol{\Delta}]^{\frac{1}{2}}\right\}(\boldsymbol{\Psi}^{-1})^\top(\boldsymbol{M}^{-1})^\top. \tag{5}$$

Note that, while (3) and (4) are symmetric, the Kalman gain matrix (5) is non-symmetric in general.

A nice consequence is that, at each step in time, one does not need to store the expressions for $\boldsymbol{P}_t$, $\boldsymbol{F}_t$ and $\boldsymbol{K}_t$, and the only Kalman recursions that need to be run are the updating equation for the state vector and the innovation equation. This greatly simplifies the filtering process as well as the smoothing process as shown in next section.

## 2.2 Approximate maximum likelihood estimation by EM

We used the steady-state matrices derived in the previous section to design an EM algorithm that approximates the maximum likelihood estimator (MLE). The basic idea is using the steady-state Kalman filter straight form the beginning of the time series instead of waiting the steady-state to be reached by the regular Kalman filter after few step. For sufficiently long time series like the ones generally used in finance this approximation should be harmless. Indeed, the asymptotic behaviour of our approximate maximum likelihood estimation is the same as the one of exact MLE when the convergence to the steady-state is exponentially fast (cf. footnote 1). As already remarked in the previous section, using the steady-state Kalman filter greatly reduces the computational burden as the matrices $P$, $K$, $F$ and $F^{-1}$ become time-invariant and they are computed only once for each iteration of the optimisation procedure (be it EM or Quasi-Newton).

This approach was advocated by Harvey (1989, p. 434), who named it *Approximate Maximum Likelihood* (see also Harvey and Peters, 1990, p. 93). For example, for the local level model, Harvey shows that the likelihood can be maximised numerically by choosing the signal-noise parameter that minimises the sum of squared innovations (see Harvey, 1989, Section 8.3.3). This approach, requiring the knowledge of the steady-state quantities, can also be implemented using the so-called Expectation-Maximization (EM) algorithm. This is especially recommended for large-scale systems, avoiding the issue of numerically maximizing the likelihood by Quasi-Newton methods. The EM algorithm for the estimation of state space models is fully discussed by Watson and Engle (1983); Koopman (1993); Durbin and Koopman (2001); Shumway and Stoffer (2017). This iterative algorithm requires the implementation of the smoothing process

(for the Expectation-step) and the updating of the parameters (Maximization step) involving the smoothing quantities derived from the Expectation-step.

In high dimensions, the estimation of models in state space form using the EM algorithm can be rather slow because of the multiple passes of the smoother, which involves many matrix products, additions, inversions and the storage of a great number of large matrices. However, the use of the closed-form steady-state Kalman filter greatly reduces the computational burden also of the smoother, making the Expectation step much faster. Indeed, using the steady-state matrices, the smoothing algorithm proposed by de Jong (1988, 1989) (see also Ansley and Kohn, 1985; Koopman, 1997) becomes: $\boldsymbol{r}_n = \mathbf{0}$, $\boldsymbol{N}_n = \mathbf{0}$,

$$\boldsymbol{r}_{t-1} = \boldsymbol{F}^{-1}\boldsymbol{v}_t + \boldsymbol{L}^\top \boldsymbol{r}_t \tag{6}$$

$$\boldsymbol{N}_{t-1} = \boldsymbol{F}^{-1} + \boldsymbol{L}^\top \boldsymbol{N}_t \boldsymbol{L} \tag{7}$$

where $\boldsymbol{L} = \phi \boldsymbol{I} - \boldsymbol{K}$, which can also be computed as

$$\boldsymbol{L} = \frac{1}{2\phi} M\Psi \left\{ \phi^2 \boldsymbol{I} + \boldsymbol{\Delta} + \boldsymbol{I} - \left[ (\boldsymbol{\Delta} - \boldsymbol{I} + \phi^2 \boldsymbol{I})^2 + 4\boldsymbol{\Delta} \right]^{\frac{1}{2}} \right\} (\Psi^{-1})^\top (M^{-1})^\top.$$

Therefore, given a set of structural parameters $\phi$, $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$, the smoothing recursions quickly deliver the desired results (refer to Chapter 4 of Durbin and Koopman, 2001, for an extensive treatment of the filtering and smoothing approach for linear Gaussian state-space models).

Using the quantities provided by the Expectation-step, the Maximization step for the matrices $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_\eta$ can be carried out using the simple updating expressions (3.5), (3.6) and (3.7) of Section 3 in Koopman (1993). More specifically, for model (1) these expressions can be restated as follows:

$$\boldsymbol{\Sigma}_\epsilon(\iota + 1) = \boldsymbol{\Sigma}_\epsilon(\iota) + \boldsymbol{\Sigma}_\epsilon(\iota) \boldsymbol{\Theta}_e \boldsymbol{\Sigma}_\epsilon(\iota) \tag{8}$$

$$\boldsymbol{\Sigma}_\eta(\iota + 1) = \boldsymbol{\Sigma}_\eta(\iota) + \boldsymbol{\Sigma}_\eta(\iota) \boldsymbol{\Theta}_r \boldsymbol{\Sigma}_\eta(\iota) \tag{9}$$

where $\iota = 0, 1, \ldots$ is the iteration index. Here, $\boldsymbol{\Sigma}_\epsilon(0)$ and $\boldsymbol{\Sigma}_\eta(0)$ are arbitrary covariance matrices to be used as starting values,

$$\boldsymbol{\Theta}_r = \frac{1}{n} \sum_{t=1}^{n} (\boldsymbol{r}_t \boldsymbol{r}_t' - \boldsymbol{N}_t), \tag{10}$$

where $\boldsymbol{r}_t$ and $\boldsymbol{N}_t$ are constructed as in (6) and (7), and

$$\boldsymbol{\Theta}_e = \frac{1}{n} \sum_{t=1}^{n} (\boldsymbol{e}_t \boldsymbol{e}_t' - \boldsymbol{D}_t), \tag{11}$$

with

$$\boldsymbol{e}_t = \boldsymbol{F}^{-1}\boldsymbol{v}_t - \boldsymbol{K}' \boldsymbol{r}_t \tag{12}$$

and

$$\boldsymbol{D}_t = \boldsymbol{F}^{-1} + \boldsymbol{K}' \boldsymbol{N}_t \boldsymbol{K}. \tag{13}$$

Finally, the Maximization step for the vector of constants $\boldsymbol{\kappa}$ and the autoregressive parameter $\phi$ is given by

$$\boldsymbol{\kappa}(\iota) = \frac{1}{n-1} \sum_{t=2}^{n} (\boldsymbol{a}_{t|n} - \phi \boldsymbol{a}_{t-1|n}) \tag{14}$$

$$\phi(\iota) = \frac{\mathrm{Tr}\left( \sum_{t=2}^{n} (\boldsymbol{I} - \boldsymbol{P}\boldsymbol{N}_{t-1}) \boldsymbol{L}\boldsymbol{P} + \boldsymbol{a}_{t|n} \boldsymbol{a}_{t-1|n}' - \boldsymbol{\kappa}(\iota) \boldsymbol{a}_{t-1|n}' \right)}{\mathrm{Tr}\left( \sum_{t=2}^{n} \boldsymbol{P} - \boldsymbol{P}\boldsymbol{N}_{t-1}\boldsymbol{P} + \boldsymbol{a}_{t|n} \boldsymbol{a}_{t|n}' \right)}, \tag{15}$$

where $\boldsymbol{a}_{t|n} = \boldsymbol{a}_t + \boldsymbol{P}\boldsymbol{r}_{t-1}$. Note that (14) and (15) are, respectively, the partial derivative of the likelihood function with respect to $\boldsymbol{\kappa}$ and to $\phi$ as shown in Appendix C.

Given that both the Expectation and Maximization steps make use of the steady-state quantities, this algorithm guarantees important computational gains (see Table 1).

In order to assess the speed and precision of our approach and to compare it to standard MLE, we implemented few Monte Carlo experiments. We simulated sample paths of length $n = 1000$ from model (1) using Gaussian disturbances and random structural parameters generated as follows:

- the common AR(1) coefficient $\phi$ was generated from a uniform distribution with support $[.85, .95]$;
- all constants in the vector $\boldsymbol{\kappa}$ were set to zero;
- both matrices $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\epsilon$ were correlation matrices generated by the following steps:
  1. we generated a matrix $\boldsymbol{A}$ of uniform numbers on $[0, 1]$;
  2. we computed the symmetric matrix $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^\top$;
  3. we extracted the eigenvalues and eigenvectors of $\boldsymbol{B}$ and built the matrix $\tilde{\boldsymbol{B}}$ with the same eigenvectors of $\boldsymbol{B}$ and with eigenvalues transformed as
  $$\tilde{\lambda_j} = 1 + 29\frac{\lambda_j - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}},$$
  where $\lambda_{\max}, \ldots, \lambda_{\min}$ are the ordered eigenvalues of $\boldsymbol{B}$;
  4. finally, we set $\boldsymbol{\Sigma}$ equal to the correlation matrix obtained from the covariance matrix $\tilde{\boldsymbol{B}}$ .

We used the eigenvalue transformation at step 3 to keep the condition number of the covariance matrices approximatively equal to 30 for any considered dimension and, thus, avoiding numerically non-invertible matrices.

Table 1 reports the timing (in seconds) and precision (in term of mean absolute error[2]) of our EM-based approximate MLE for a range of dimensions $d$. For $d \leq 10$ we report also the same quantities for the regular MLE based on the full Kalman filter. The precision of the two methods is comparable, but the computation time makes the classical MLE infeasible in high dimensions.[3]

## 2.3 Consistent estimation of correlations between processes with unknown time-varying variances

In this section we show how to consistently estimate the correlation matrix of a vector process such as those used in multivariate stochastic volatility models, regardless of the evolution law of the volatilities.

**Assumption 1.** *Consider the vector process $\boldsymbol{x}_t$, whose i-th element is $X_{it} = Z_{it}\sigma_{it}$, for $i = \{1, \ldots, d\}$, $t = \{1, \ldots, n\}$, $0 < \sigma_{it} < \infty$, and let $Z_{it}$ be i.i.d. following a continuous d-dimensional elliptical*

[2]Since the estimated covariances are correlations in the interval $[0, 1]$, we based our comparisons on the simple-to-interpret mean absolute error (MAE) rather than on more typical matrix distances such as Frobenius.

[3]All simulations ran on a virtual machine with Intel Xeon CPU E5-2667 v3 @ 3.20GHz 8-core processor, 56.0 GB RAM, running Microsoft R Open 3.5.1 (with Intel MKL) on Windows 10 Pro. The timing has been measured using the *microbenchmark* package for R (Mersmann, 2018).

Table 1: Comparisons of Kalman filter based MLE with the estimates obtained by our EM algorithm on a multivariate local level model of dimension $d$ with length $n = 1,000$.

| | Execution times (sec.) | | | MAE $\Sigma_\epsilon$ | | MAE $\Sigma_\eta$ | | MAE $\phi$ | | MAE $\kappa$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $d$ | MLE | EM | MLE/EM | MLE | EM | MLE | EM | MLE | EM | MLE | EM |
| 3 | 1.87 | 0.05 | 40 | .060 | .056 | .072 | .064 | .008 | .009 | .029 | .027 |
| 5 | 7.77 | 0.21 | 38 | .061 | .058 | .069 | .067 | .008 | .010 | .029 | .027 |
| 10 | 115.33 | 0.27 | 425 | .060 | .058 | .069 | .066 | .006 | .009 | .029 | .027 |
| 25 | | 1.52 | | | .060 | | .065 | | .005 | | .025 |
| 50 | | 10.89 | | | .056 | | .062 | | .004 | | .026 |
| 100 | | 273.24 | | | .058 | | .061 | | .004 | | .025 |
| 200 | | 1250.25 | | | .049 | | .052 | | .012 | | .023 |

*distribution with zero median (and, thus, also zero mean when it exists) and correlation matrix* $\boldsymbol{R} = \{\rho_{ij}\}$.[4]

**Theorem 2.** *Under Assumption 1,*

$$\nu_{ij} := \mathbb{E}\,\mathrm{sign}\left(X_{it}X_{jt}\right) = \frac{2}{\pi}\arcsin\rho_{ij}. \tag{16}$$

**Remark 1.** Notice that the evolution of the scale parameters $\sigma_{it}$ can be of any kind: deterministic or stochastic. The simple trick we exploit to get rid of the arbitrary scale parameters and estimate the correlations is

$$\mathrm{sign}(X_{it}X_{jt}) = \frac{X_{it}X_{jt}}{|X_{it}||X_{jt}|} = \frac{(Z_{it}\sigma_{it})(Z_{jt}\sigma_{jt})}{|Z_{it}\sigma_{it}||Z_{jt}\sigma_{jt}|} = \frac{Z_{it}Z_{jt}}{|Z_{it}||Z_{jt}|} = \mathrm{sign}(Z_{it}Z_{jt}).$$

**Remark 2.** The reader acquainted with copula theory has certainly noticed that the map between $\rho_{ij}$ and $\nu_{ij}$ is the same as the map between $\rho_{ij}$ and Kendall's rank correlation coefficients, $\tau_{ij}$, when the random variables are elliptically distributed (Lindskog et al., 2003). Indeed, it can be proven that, under Assumption 1, the population $\tau_{ij}$ and $\nu_{ij}$ of the pair $(Z_{it}, Z_{jt})$ coincide. This result is relevant in large samples, because the computation of the sample Kendall's $\tau$ can be quite demanding: the most efficient way of computing Kendall's $\tau$ is by using the algorithm of Knight (1966), which is $O(n\log n)$, while the computational complexity of the sample version of $\nu_{ij}$ (see equation (17) below) is only $O(n)$.

**Theorem 3.** *Let us define the sample mean sign*

$$\hat{\nu}_{ijn} := \frac{1}{n}\sum_{t=1}^{n}\mathrm{sign}(X_{it}X_{jt}), \tag{17}$$

*and its sine transform*

$$\hat{\rho}_{ijn} = \sin\left(\frac{\pi}{2}\hat{\nu}_{ijn}\right). \tag{18}$$

*Under Assumption 1,*

- $\hat{\nu}_{ijn} \xrightarrow{a.s.} \nu_{ij}$;
- $\sqrt{n}\left(\hat{\nu}_{ijn} - \nu_{ij}\right) \xrightarrow{d} \mathcal{N}\left(0, 1-\nu^2\right)$;
- $\hat{\rho}_{ijn} \xrightarrow{a.s.} \rho_{ij}$;

[4]The i.i.d. hypothesis can be relaxed provided that the marginal distribution of the processes $Z_{it}$ is continuous and elliptical with correlation matrix $\boldsymbol{R}$ and that a law of large number and a central limit theorem apply.

- $\sqrt{n}\left(\hat{\rho}_{ijn} - \rho_{ij}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$ *with*

$$\sigma^2 = \left[\frac{\pi}{2}\cos\left(\frac{\pi}{2}\nu_{ij}\right)\right]^2 (1-\nu^2) = (1-\rho_{ij}^2)\left(\frac{\pi^2}{4} - \arcsin^2\rho_{ij}\right).$$

Thus, if two random variables are elliptically distributed, their correlation can be consistently estimated, regardless of any form of heteroskedasticity, using a transformation of the average sign of their product. This result is very useful in financial applications because when the variances evolve as non-stationary processes (as numerous applications of GARCH and SV models to real data seem to suggest) the sample correlation is an inconsistent estimator.

# 3 Multivariate stochastic volatility models

## 3.1 The Harvey-Ruiz-Shephard (HRS) model

The basic univariate SV model in discrete time is defined by

$$\begin{aligned} y_t &= \sigma_t \zeta_t, & \zeta_t &\sim \text{NID}(0,1), \\ \log\sigma_t^2 &= \kappa + \phi\log\sigma_{t-1}^2 + \eta_t, & \eta_t &\sim \text{NID}(0,\tau^2), \end{aligned} \tag{19}$$

where $y_t$ is the time series of returns, NID denotes a normally independently distributed sequence of random variables and $\kappa$, $\phi$, $\tau^2$ are parameters.

Harvey et al. (1994) propose to linearise the above SV model by taking a *log-of-square* transform of the first equation of (19) and approximate it with the Gaussian state-space form

$$\begin{aligned} \log y_t^2 &= \log\sigma_t^2 + \log\zeta_t^2, & \log\zeta_t^2 &\approx \text{NID}(1.27, \pi^2/2), \\ \log\sigma_t^2 &= \kappa + \phi\log\sigma_{t-1}^2 + \eta_t, & \eta_t &\sim \text{NID}(0,\tau^2). \end{aligned} \tag{20}$$

and carry out numerical QML estimation using the Kalman filter.

Harvey et al. (1994) propose also the multivariate generalisation of the above model, which can be written as

$$\begin{aligned} \boldsymbol{v}_t &= \boldsymbol{h}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\approx \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \\ \boldsymbol{h}_t &= \boldsymbol{\kappa} + \boldsymbol{\Phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{21}$$

where $\boldsymbol{v}_t$ is a vector whose generic element is $v_{it} = \log y_{it}^2 - 1.27$, $\boldsymbol{\varepsilon}_t$ is a vector whose generic element is $\varepsilon_{it} = \log\zeta_{it}^2 - 1.27$, $\boldsymbol{\Phi}$ is a diagonal matrix with $\phi_1, \phi_2, \ldots, \phi_d$ on the main diagonal, $\boldsymbol{\Sigma}_\eta$ is a free covariance matrix and $\boldsymbol{\Sigma}_\varepsilon$ is a covariance matrix whose generic $ij$-th element is given by

$$\begin{cases} \pi^2/2 & \text{if } i = j, \\ \sum_{k=1}^{\infty} \frac{(k-1)!}{(1/2)_k}\rho_{ij}^{2k} & \text{if } i \neq j, \end{cases} \tag{22}$$

with $\rho_{ij}$ denoting the correlation between $\zeta_{it}$ and $\zeta_{jt}$, and $(x)_k := x(x+1)\ldots(x+k-1)$.

## 3.2 The Alizadeh-Brandt-Diebold (ABD) model

Various authors (Jacquier et al., 1994; Andersen and Sørensen, 1997; Kim et al., 1998) argued that, since the Gaussian approximation used in models (20)–(21) is poor, then QML estimation is expected to be highly inefficient and severely biased in finite samples. Alizadeh et al. (2002) proposed an alternative SV model that should overcome these issues. The measure of daily volatility used in their model is the log-range of daily prices defined by

$$w_{it} = \log\left(P_{it}^{\max} - P_{it}^{\min}\right),$$

where $P_{it}^{\max}$ and $P_{it}^{\min}$ denote the maximum and minimum price of asset $i$ in the time span $\delta_t$ (which throughout this paper is one trading day). If the evolution of the price within the time span $\delta_t$ is well approximated by a Brownian motion, then the log-range has a distribution characterised by the following first four moments[5]: mean$= 0.43 + \frac{1}{2}\log\delta_t + \log\sigma_t$, sd$= 0.29$, skewness$= 0.17$ and kurtosis$= 2.80$ (cf. Alizadeh et al., 2002, Table I). Since skewness and kurtosis are close to those of a Gaussian distribution, Alizadeh et al. (2002) expect their model to perform better than HRS'.

The model of Alizadeh et al. (2002) for the generic asset $i$ can be easily cast in linear state space form as

$$\begin{aligned} w_{it} - 0.43 - \frac{1}{2}\delta_t &= \log\sigma_{it} + \varepsilon_{it}, & \varepsilon_{it} &\approx \text{NID}(0, 0.29^2) \\ \log\sigma_{it} &= \kappa_i + \phi_i \log\sigma_{i,t-1} + \eta_{it}, & \eta_{it} &\sim \text{NID}(0, \tau^2). \end{aligned}$$

It is simple to write the multivariate version of the model, however the exact computation of how the correlation between two (Brownian) prices translates into correlation between two log-ranges is rather involved. We solved this problem by simulation and found that a degree 10 polynomial only with even exponents provides an excellent approximation (see Appendix 2.3). The multivariate version of the ABD model is

$$\begin{aligned} \boldsymbol{w}_t &= \boldsymbol{h}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\approx \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \\ \boldsymbol{h}_t &= \boldsymbol{\kappa} + \boldsymbol{\Phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{23}$$

where $\boldsymbol{\kappa}$ and $\boldsymbol{\Phi}$ are as above, the generic element of $\boldsymbol{w}_t$ is $w_{it} - 0.43 - \delta_t/2$, the $i$-element of $\boldsymbol{h}_t$ is $\log\sigma_{it}$, $\boldsymbol{\Sigma}_\eta$ is a covariance matrix of unknown quantities and the $ij$-th element of matrix $\boldsymbol{\Sigma}_\varepsilon$ is given (approximately) by

$$0.29^2\left(0.7447\rho_{Wij}^2 + 0.5738\rho_{Wij}^4 - 1.1100\rho_{Wij}^6 + 1.0524\rho_{Wij}^8 - 0.2609\rho_{Wij}^{10}\right), \tag{24}$$

with $\rho_{Wij}$ correlation between the two Brownian motions driving the prices of assets $i$ and $j$.

Empirical applications of GARCH and SV models tend to find a high persistence of the conditional volatility. For SV models this translates into estimates of $\phi_i$ in a left neighbourhood of 1 (cf. Harvey et al., 1994, Table 2). For this reason and, more importantly, to reduce the number of parameters to estimate, in our applications we consider only SV models in which $\boldsymbol{\kappa} = \mathbf{0}$ and $\boldsymbol{\Phi} = \boldsymbol{I}_d$ (the same constraints are used in the multivariate model of Harvey et al., 1994, Sections 3–6).[6]

## 3.3 Estimating large SV models

If the volatility of returns is not mean-reverting, as in the case $\boldsymbol{\Phi} = \boldsymbol{I}_d$, then the sample correlation of returns is not consistent for the population correlation, but our estimator based on signs is still consistent. Thus, our procedure for estimating the two SV models when log-volatilities are random walks is the following:

1. estimate the returns' correlation matrix using (17)–(18);
2. transform the estimated correlations using (22) for the HRS model or (24) for the ABD model;

[5] In the following formulae we assume that the volatility $\sigma_t$ is annualised and that $\delta_t$ is measured in years (generally, if $\delta_t$ is one day its value is set to $1/252$ since there are approximately 252 trading days per year).

[6] We estimated the above models also without constraints on $\boldsymbol{\kappa}$ and $\phi$ but we always obtained estimates of $\phi$ extremely close to 1 and of $\boldsymbol{\kappa}$ virtually equal to zero for all its elements. These results are available upon request.

3. run the EM algorithm using only updating equation (9) until convergence;

4. run the classical or the steady-state smoother using the estimated parameters to obtain inference for the in-sample log-volatilities $\boldsymbol{h}_t$ and their out-of-sample predictions.

Alternatively, if a model with mean-reverting volatilities is preferred, then our results can be applied only under the constraint that all log-volatilities share the same persistence: $\phi_i = \phi$ for $i = 1, 2, \ldots, d$. In this case the procedure becomes:

1. estimate the returns' correlation matrix using either sample correlations or (17)–(18);

2. transform the estimated correlations using (22) for the HRS model or (24) for the ABD model;

3. run the complete EM algorithm until convergence;

4. run the classical or the steady-state smoother using the estimated parameters to obtain inference for the in-sample log-volatilities $\boldsymbol{h}_t$ and their out-of-sample predictions.

# 4 Applications

## 4.1 Minimum variance portfolios

We applied our estimation procedure to the HRS and ABD models using two different baskets fo stocks. The first one is composed by 38 stocks belonging to CAC40 index (we had two exclude a couple of stocks whose time series were too short for the analysis); the second includes 88 stocks in the S&P100 index (again, we excluded 12 stocks because their time series were too short or for the particularly odd behaviours of their prices).[7]

For CAC40 stocks, we used the time span [2009-03-27]–[2018-03-27] as estimation sample and [2018-03-28]–[2019-03-28] as evaluation window, while for S&P100 stocks we used the interval [2008-03-14]–[2017-03-13] for estimation and [2017-03-14]–[2018-03-14] for evaluation. We decided to evaluate the SV models on two different time-spans and stock indexes as a robustness check.

The performance of the two SV models is assessed using a minimum variance framework. In particular, we use the models fitted on the estimation window to produce one-step-ahead predictions of the covariance metrics of the stock returns over the evaluation window. These matrices are used to build a minimum variance portfolio updated on a daily basis. The realized variance of the portfolios over the one-year evaluation window is the quantity we rely on for the assessment (the smaller, the better). The same criteria are applied to three alternative methods to obtain minimum variance portfolios:

**Fixed:** a static portfolio based on historical covariances computed using the estimation window;

**RM:** a daily updated portfolio based on the RiskMetrics approach, which predicts one-step-ahead covariances using the recursion

$$\sigma_{ij,t+1} = \lambda \sigma_{ij,t} + (1-\lambda) y_{it} y_{jt},$$

with $y_{it}$ and $y_{jt}$ representing the returns and $\lambda = 0.94$;

[7] Closing prices, daily minima and maxima were downloaded from FactSet.

**DCC:** a daily updated portfolio based on the covariance matrices produced by a DCC model fitted to the estimation window (we used the R package *rmgarch* by Ghalanos (2019)).

The computation times of our procedure and of the DCC are comparable: less than 20 seconds for the CAC40 and slightly more than a minute for the SP100[8]. Figure 1 depicts the recursive volatilities of the portfolios with CAC40 stocks as a function of the time $t$. The recursive volatility is computed as

$$vol_t = \left( \frac{1}{t} \sum_{s=1}^{t} r_s^2 \right)^{1/2},$$

where $r_s$ is the portfolio return at time $s$. The top panel of Figure 1 refers to unconstrained minimum variance portfolios, while for the portfolios in the bottom panel short-selling was not allowed.

In both cases (unconstrained and constrained) the HRS model performs extremely well with a recursive volatility, which is very close to that of the DCC-based portfolio. The ABD model, on the contrary, seems to produce the worst performing portfolios. The behaviour of the RiskMetrics-based portfolio is rather odd, being among the worst when unconstrained and among the best under no short-selling constraints.

Figure 2 reports the results of the same portfolio strategies applied to the S&P100 stocks. In this case the DCC-based portfolio is the clear winner, but the HRS is performing almost as well. The ABD is the worst performing model and, again, the RiskMetrics-based portfolio performs very poorly in the unconstrained case and extremely well if no short-selling constraints are imposed.

The partial conclusions we can draw from these applications are: in a minimum variance portfolio framework,

- the HRS and the DCC seem to perform similarly even though the former is a constant conditional correlation model and the latter a dynamic conditional correlation model;
- the multivariate ABD model performs poorly on both datasets and portfolio types;
- the RiskMetrics-based strategy performance is bad when the minimum variance portfolio is unconstrained and good under no short-selling constraints;
- the fixed portfolio based on the historical covariance lays persistently in the middle of the ranking.

## 4.2 Common factors in the evolution of the volatilities

One advantage of multivariate SV models over GARCH-DCC is that using the former one obtains an estimate of the covariance structure of the (log-)variance processes. This information can be used in several ways, but probably the most relevant is to reveal the presence of common factors in the evolution of volatilities. Furthermore, we can also exploit this information to understand why HRS and ABD perform so differently.

Figure 3 represents the correlation matrices of the log-variance processes for the two models. All correlations are non-negative, but the correlations of ABD are much weaker than those of HRS. Thus, according to the ABD model, volatilities tend to evolve in rather idiosyncratic ways, while the HRS model suggests strong co-movements among stock volatilities. The eigenvalues of these two matrices are represented in Figure 4: for the HRS model the first 4 principal components

[8]All computations ran on a virtual machine with Intel Xeon CPU E5-2667 v3 @ 3.20GHz 8-core processor, 56.0 GB RAM, running Microsoft R Open 3.5.1 (with Intel MKL) on Windows 10 Pro. The timing refers to the *elapsed time* returned by the R function system.time().

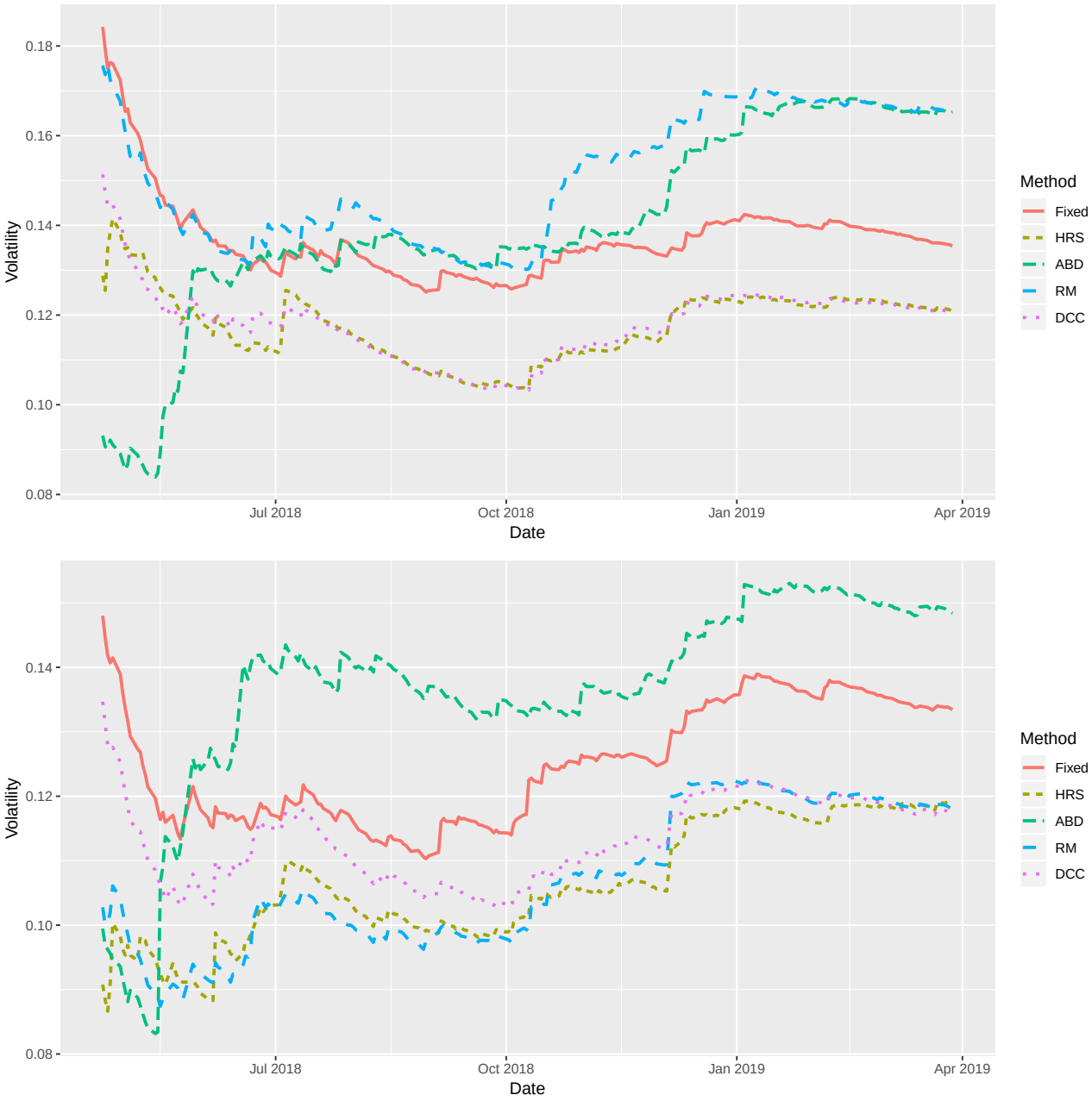


Figure 1: CAC40: recursive volatilities of the portfolios over the evaluation window. The plot on the top refers to unconstrained minimum variance portfolios, the bottom plot to minimum variance portfolios with non-negative weights (no short-selling).

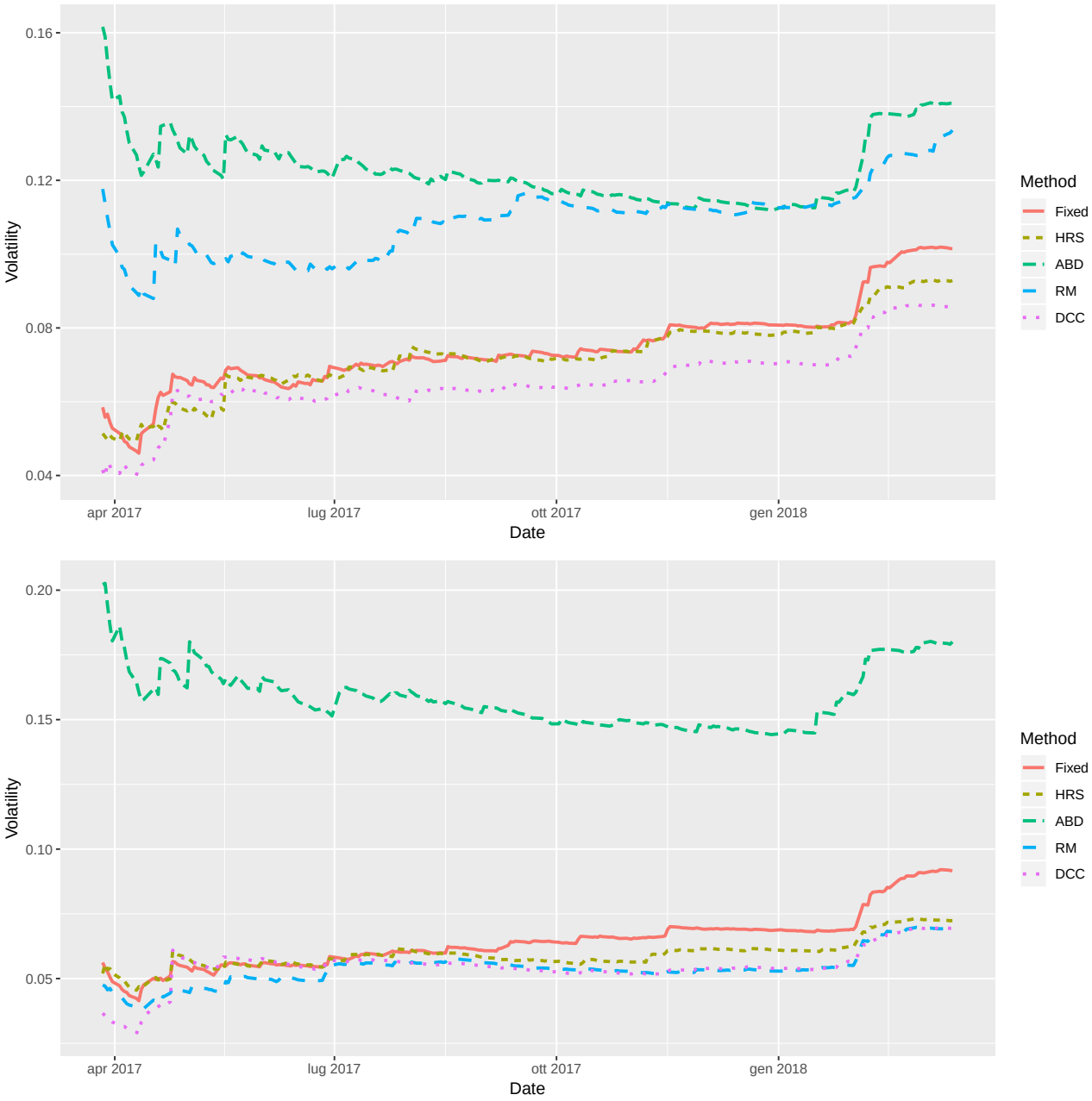


Figure 2: S&P100: recursive volatilities of the portfolios over the evaluation window. The plot on the top refers to unconstrained minimum variance portfolios, the bottom plot to minimum variance portfolios with non-negative weights (no short-selling).

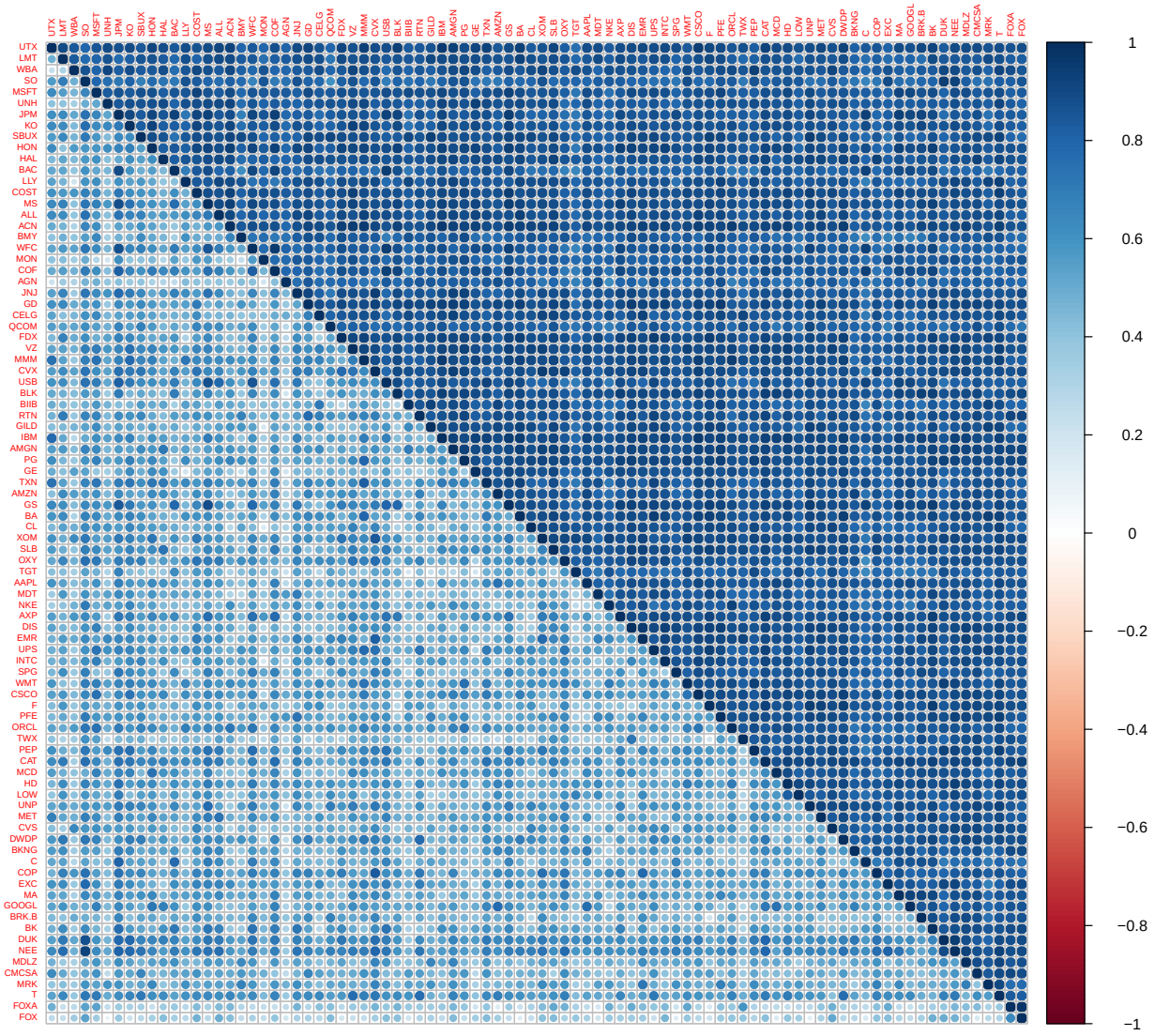


Figure 3: Graphical representation of the correlations in the estimated matrices $\boldsymbol{\Sigma}_\eta$: the correlation of the HRS model are above the main diagonal, while those of the ABD model are below the same diagonal.

explain 90% of the total variance, while for the ABD model 21 components are needed to achieve the same result.

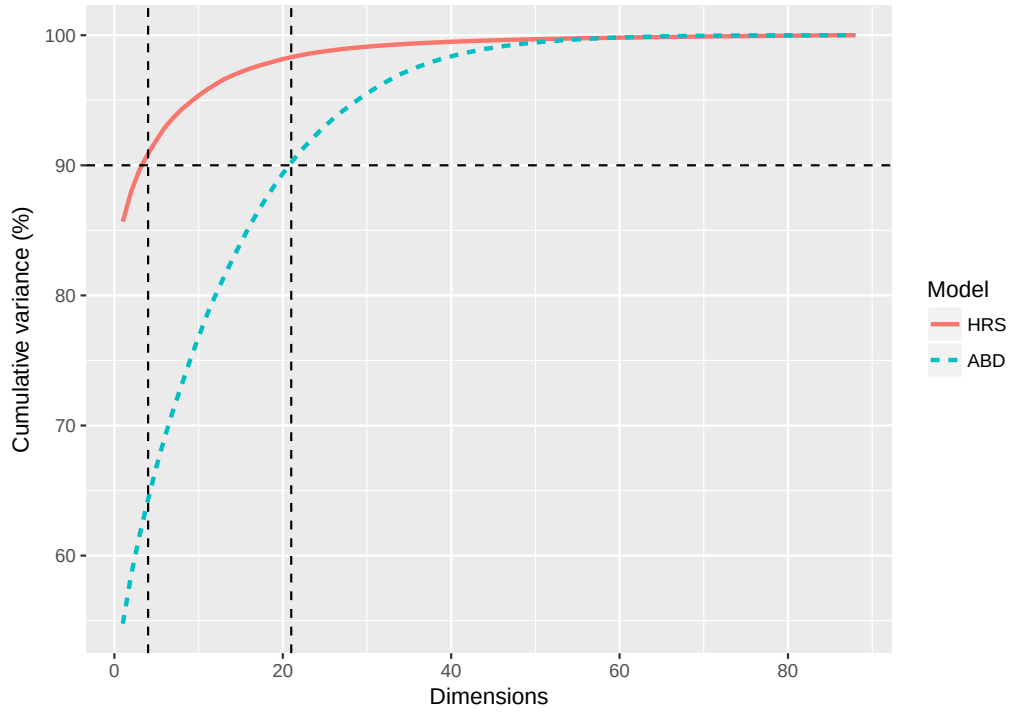


Figure 4: Percentage of variance explained by the first principal components.

Finally, for each SV model Figure 5 depicts the volatility index produced by using the first eigenvector of the estimated $\mathbf{\Sigma}_\eta$ matrix as a portfolio of the 88 individual smoothed and predicted (one-step-ahead) volatilities. Both indexes are compared with the well-known VIX index as published on the Cboe website. The index based on the HRS model (top of Figure 5) is very close to the VIX but it enhances the peaks as, for example, in the last part of year 2008 and in the second half of year 2011. The index built on the ABD model has a different scale and shows a growing volatility over the years 2012–2018, which is not witnessed by the VIX and the HRS-based index.

We tried to understand why the ABD model seems to perform so poorly. Our conjecture is that the approximation of intra-day price movements with a Brownian motion is not very sharp. This conjecture is supported by the plots in Figure 4.2, where we try to represent the relationship between log of volatility and log-ranges for the Amazon stock. The left panel of that figure depicts the log of absolute returns and log-ranges, while the right panel shows the daily return standard deviation and the mean daily log ranges computed over periods of 30 days. The relations in both plots are quite far from their theoretical version: $\mathbb{E}\, w_t = 0.43 + \log \sigma_t$, where $w_{it}$ is the daily log-range and $\sigma_t$ is the daily standard deviation. A deeper investigations on the quality of approximation of the ABD model to the empirical behaviour of stock log-ranges goes beyond the scope of this paper; however the issue should be addressed in future works because few adjustments based on the actual empirical behaviour of log-ranges could make their approach improve significantly also in the multivariate context.

# 5 Conclusion

We proposed a computationally efficient way to estimate large multivariate AR(1)-plus-noise models and, in particular, large multivariate local level models. This result was achieved by showing

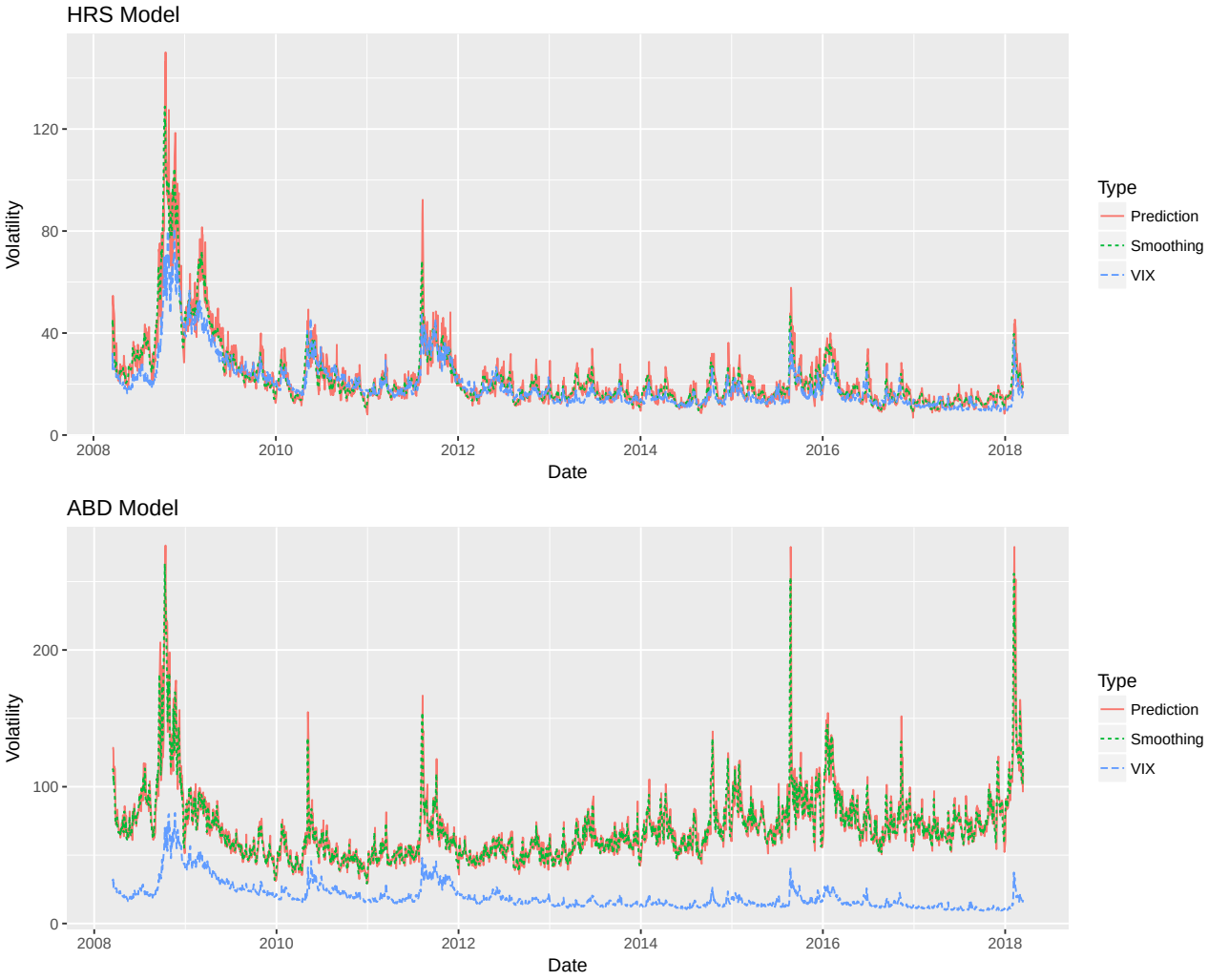


Figure 5: VIX index and volatility indexes built using the first principal component of the smoothed and one-step-ahead predicted volatilities in models HRS (top) and ABD (bottom).

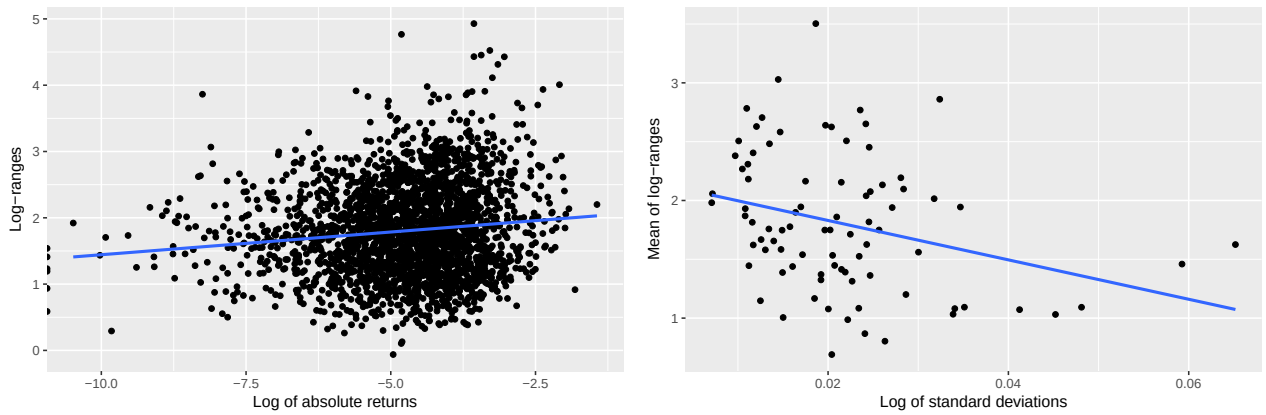


Figure 6: Empirical relationship between the log of standard deviation and the expectations of log-returns of the Amazon stock. The left panel depicts log-ranges as a function of the square roots of absolute returns. The right panel depicts mean daily log-ranges computed over 30-day periods as a function of the log of daily returns standard deviations computed over 30-day periods.

how to solve the Riccati equation implied by the steady-state Kalman filter and by proposing an EM algorithm that exploits the steady-state smoother to estimate the numerous unknown parameters of the model. Furthermore, we proposed a consistent and asymptotically normal estimator of the correlation between two random sequences, whose variances are free to evolve in arbitrary ways. We exploit our results to estimate high dimensional stochastic volatility models of two types: the model of Harvey et al. (1994) (HRS) and the model by Alizadeh et al. (2002) (ABD). We applied the two models to minimum-variance portfolios composed by the constituents of the CAC40 and of the S&P100 indexes, and compared their performance to those of portfolios built on historical covariances, RiskMetrics predictions and DCC-GARCH predictions (Engle, 2002). Despite being a constant conditional correlation model, the HRS model shows a performance that is very close to that of the DCC model, while the ABD model tends to perform poorly. Furthermore, the covariance matrix of the log-volatility processes produced by our estimation approach allows the investigation of the factor structure of the time-varying volatilities, or stated differently, the identification of common features in the volatility of volatility.

Our proposal represents a good alternative to the popular DCC model. The advantages of using large multivariate SV models instead of DCC-GARCH type models are manifolds: the properties of SV models are more intuitive and easier to derive, discrete time SV models are natural counterparts to those continuous time SV models used in the financial literature, the direct estimation of the covariance matrix of the volatility components opens the field to many interesting applications that are not so immediate in the DCC-GARCH world.

# References

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance 57*(3), 1047–1091.

Andersen, T. G. and B. E. Sørensen (1997). GMM and QML asymptotic standard deviations in stochastic volatility models: Comments on Ruiz (1994). *Journal of Econometrics 76*(1), 397 – 403.

Ansley, C. and R. Kohn (1985). Estimation, filtering and smoothing in state space models with incomplitely specified initial conditions. *The Annals of Statistics 13*, 1286–1316.

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

Cambanis, S., S. Huang, and G. Simons (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis 11*, 368–385.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*(2), 341 – 371.

de Jong, P. (1988). A cross-validation filter for time series models. *Biometrika 76*, 594–600.

de Jong, P. (1989). Smoothing and interpolation with state-space models. *Journal of the American Statistical Association 75*, 594–600.

Durbin, J. and S. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

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

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Fang, K.-T., S. Kotz, and K.-W. Hg (1987). *Symmetric Multivariate and Related Distributions*. Chapman & Hall.

Ghalanos, A. (2019). *rmgarch: Multivariate GARCH models.* R package version 1.3-6.

Harvey, A. (1989). *Forecasting Structural Time Series and the Kalman Filter*. Cambridge University Press.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*(2), 247–264.

Harvey, A. C. and S. Peters (1990). Estimation procedures for structural time series models. *Journal of Forecasting 9*, 89–108.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies 6*(2), 327–343.

Hull, J. and A. White (1987). Hedging the risks from writing foreign currency options. *Journal of International Money and Finance 6*(2), 131–152.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*(4), 371–389.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies 65*(3), 361–393.

Knight, W. R. (1966). A computer method for calculating kendall's tau with ungrouped data. *Journal of the American Statistical Association 61*(314), 436–439.

Koopman, S. (1993). Disturbance smoother for state space model. *Biometrika 80*(1), 117–126.

Koopman, S. (1997). Exact initial kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association 92*(400), 1630–1638.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*(4), 823–826.

Lindskog, F., A. McNeil, and U. Schmock (2003). Kendall's tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk: Measurement, Evaluation and Management*, pp. 149–156. Heidelberg: Physica-Verlag.

Mersmann, O. (2018). *microbenchmark: accurate timing functions*. R package version 1.4-4.

Nadarajah, S. and T. K. Pogány (2016). On the distribution of the product of correlated normal random variables. *Comptes Rendus Mathematique 354*(2), 201–204.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Shumway, R. H. and D. S. Stoffer (2017). *Time Series Analysis and Its Applications. With R Examples*. Springer.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC, and varying coefficient regression models. *Journal of Econometrics 23*, 385–400.

# Appendix

## A Computing the correlations between log-ranges

Alizadeh et al. (2002, Table I) provide the first two moments of the univariate log-range, however they do not consider multivariate models and, thus, do not compute the correlation of log-range pairs relative to two correlated Browninan motions. Instead of trying to get an exact formula that, as in equation (15) in Alizadeh et al. (2002), can be represented only as an infinite expansion, we approximated the function mapping the Brownian motion correlations in the log-range correlations by simulation and regression. As it will appear our approximation is very accurate for working with real data.

Since the behaviour of two log-ranges generated by two correlated Brownian motions is symmetric with respect to the sign of the correlation (i.e., the behaviour of the two log-ranges will be stochastically the same for $\rho_W$ and $-\rho_W$, where $\rho_W$ is a positive correlation between the two Brownian motions), we based the simulations on a grid of only positive correlations (i.e., $\rho_W = 0.10, 0.20, \ldots, 0.90, 0.91, \ldots, 0.99$). For the simulations, we assumed that the volatility is constant over one day of trading and sampled the continuous paths of the Brownian motions on a grid of 1,000 points. This was repeated 1,000,000 times and the correlation between the two log-ranges was estimated using the sample correlation.

Table 2: Estimated correlation of two log-ranges, $\rho_\varepsilon$, as a function of the correlation of two Brownian motions, $\rho_W$.

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho_W$ | 0.00 | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.91 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 0.99 | 1.00 |
| $\rho_\varepsilon$ | 0.00 | 0.01 | 0.03 | 0.07 | 0.13 | 0.21 | 0.31 | 0.43 | 0.57 | 0.75 | 0.77 | 0.80 | 0.82 | 0.84 | 0.86 | 0.89 | 0.91 | 0.94 | 0.97 | 1.00 |

We regressed the estimated log-range correlations on the even powers of the Brownian motion correlation from 2 to 10. Furthermore, we imposed to the regression function to pass through the origin by eliminating the intercept and we imposed to the function to pass through the points $(-1,1)$ and $(1,1)$ by constraining the regression coefficients to sum to one. The constrained least squares solution for the regression coefficients is

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} - \frac{\boldsymbol{\iota}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} - 1}{\boldsymbol{\iota}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{\iota}} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{\iota},$$

where $\boldsymbol{\iota}$ is a conformable vector of ones.

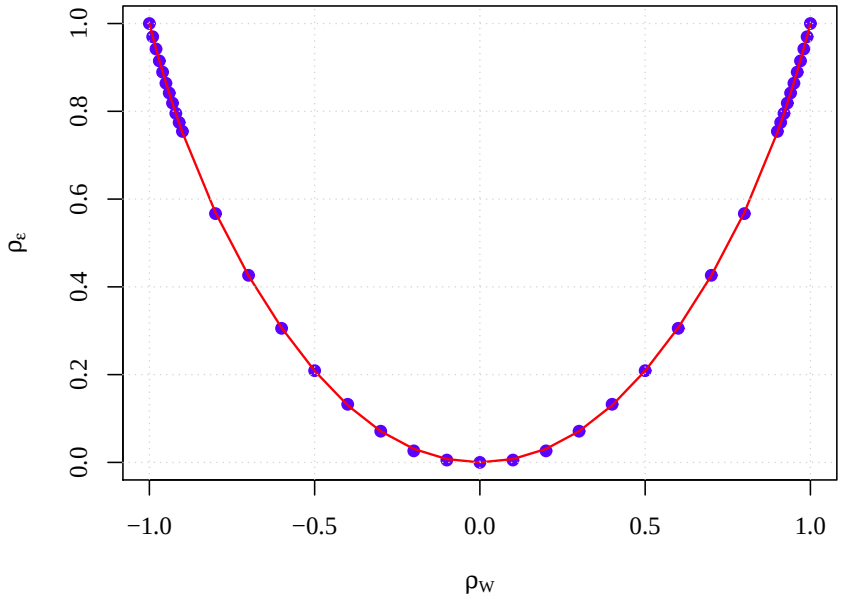


Figure 7: Approximating function that maps the correlation of two Brownian motions into the correlation of the two resulting log-ranges.

Figure 7 shows the almost perfect fit (the $R^2$ is numerically 1) of the estimated function to the log-range correlations obtained by simulation. Thus, the correlations between daily returns can be accurately mapped into correlations between daily log-ranges using the following formula

$$\rho_\varepsilon = 0.7447\rho_W^2 + 0.5738\rho_W^4 - 1.1100\rho_W^6 + 1.0524\rho_W^8 - 0.2609\rho_W^{10}. \tag{25}$$

# B Proof of Theorem 1

The transformation $\tilde{\boldsymbol{y}}_t = \boldsymbol{M}^{-1}\boldsymbol{y}_t$ allows the re-parametrization of the system (1) as

$$\begin{aligned} \tilde{\boldsymbol{y}}_t &= \tilde{\boldsymbol{\alpha}}_t + \tilde{\boldsymbol{\epsilon}}_t, \qquad \tilde{\boldsymbol{\epsilon}}_t \sim \mathrm{WN}(\boldsymbol{0}, \boldsymbol{I}) \\ \tilde{\boldsymbol{\alpha}}_{t+1} &= \phi\tilde{\boldsymbol{\alpha}}_t + \tilde{\boldsymbol{\eta}}_t, \quad \tilde{\boldsymbol{\eta}}_t \sim \mathrm{WN}(\boldsymbol{0}, \boldsymbol{Q}) \end{aligned} \tag{26}$$

where $\boldsymbol{Q} = \boldsymbol{M}^{-1}\boldsymbol{\Sigma}_\eta(\boldsymbol{M}^{-1})^\top$ is a symmetric positive definite matrix. The Riccati equation for the Kalman prediction error variance of the transformed system (26) in steady-state is

$$\tilde{\boldsymbol{P}} = \phi^2\tilde{\boldsymbol{P}} - \phi^2\tilde{\boldsymbol{P}}(\tilde{\boldsymbol{P}} + \boldsymbol{I})^{-1}\tilde{\boldsymbol{P}} + \boldsymbol{Q}, \tag{27}$$

that can be rewritten as:

$$\phi^2\tilde{\boldsymbol{P}}((\tilde{\boldsymbol{P}}+\boldsymbol{I})^{-1}\tilde{\boldsymbol{P}}-I)+\tilde{\boldsymbol{P}}=\boldsymbol{Q}, \tag{28}$$

Since $[(\tilde{\boldsymbol{P}}+\boldsymbol{I})^{-1}\tilde{\boldsymbol{P}}-\boldsymbol{I}]=-(\tilde{\boldsymbol{P}}+\boldsymbol{I})^{-1}$ then (28) simplifies to

$$-\phi^2\tilde{\boldsymbol{P}}(\tilde{\boldsymbol{P}}+\boldsymbol{I})^{-1}+\tilde{\boldsymbol{P}}=\boldsymbol{Q} \tag{29}$$

Now consider the decomposition $\tilde{\boldsymbol{P}}=\boldsymbol{\Psi}\boldsymbol{G}\boldsymbol{\Psi}'$, where $\boldsymbol{G}=\text{diag}(g_1,g_2,\cdots,g_d)$. Note that, $(\tilde{\boldsymbol{P}}+\boldsymbol{I})=\boldsymbol{\Psi}(\boldsymbol{I}+\boldsymbol{G})\boldsymbol{\Psi}'$. Hence, it holds

$$-\phi^2\tilde{\boldsymbol{P}}(\tilde{\boldsymbol{P}}+\boldsymbol{I})^{-1}+\tilde{\boldsymbol{P}}=\boldsymbol{\Psi}\Big[-\phi^2\boldsymbol{G}(\boldsymbol{I}+\boldsymbol{G})^{-1}+\boldsymbol{G}\Big]\boldsymbol{\Psi}^\top.$$

It follows from (29) that $\tilde{\boldsymbol{P}}$ shares the same eigenvectors of $\boldsymbol{Q}$. Moreover, the eigenvalues of $\tilde{\boldsymbol{P}}$ can also be expressed as combination of the eigenvalues of $\boldsymbol{Q}$. Indeed, the relation between the generic $i$-th eigenvalue of $\tilde{\boldsymbol{P}}$ and the $i$-th eigenvalue of $\boldsymbol{Q}$ can be expressed through the following quadratic equation:

$$g_i-\frac{\phi^2 g_i}{1+g_i}=\delta_i.$$

This equation has two different solutions for $g_i$. However, the only strictly positive solution (for $\delta_i>0$) is:

$$g_i=\frac{\delta_i+\phi^2-1+\sqrt{(\delta_i+\phi^2-1)^2+4\delta_i}}{2}$$

As a consequence we can write the algebraic Riccati solution as:

$$\tilde{\boldsymbol{P}}=\frac{1}{2}\boldsymbol{\Psi}\Big[\boldsymbol{\Delta}-I+\phi^2 I+((\boldsymbol{\Delta}-I+\phi^2 I)^2+4\boldsymbol{\Delta})^{\frac{1}{2}}\Big]\boldsymbol{\Psi}^\top \tag{30}$$

Note that, since $g_i$ are strictly positive for any $\delta_i>0$, then the algebraic Riccati solution is also strictly positive definite. Finally, one obtains the result as in (3) through the following transformation $\boldsymbol{P}=\boldsymbol{M}\tilde{\boldsymbol{P}}\boldsymbol{M}^\top$, which is also positive definite.

# C Proof of the maximization steps

Koopman and Shephard (1992) derives the score vector of Gaussian state-space models. Here, bearing in mind that $\Theta=\{\Sigma_\epsilon;\Sigma_\eta;\boldsymbol{\kappa};\phi\}$, the log-likelihood function for model (1) can written as in their eq. 2.1 by considering this modification:

$$\begin{aligned}
&Q(\Theta,\Theta*)=-\tfrac{1}{2}\textstyle\sum_{t=1}^n\{\log|\boldsymbol{\Sigma}_\epsilon(\Theta)|+\log|\boldsymbol{\Sigma}_\eta(\Theta)|\}-\\
&-\tfrac{1}{2}\textstyle\sum_{t=1}^n\text{Tr}\Big[\boldsymbol{\Sigma}_\epsilon(\Theta)^{-1}\{\boldsymbol{y}_t\boldsymbol{y}_t'-\boldsymbol{y}_t\boldsymbol{a}_{t|n}'-\boldsymbol{a}_{t|n}\boldsymbol{y}_t'+\boldsymbol{a}_{t|n}\boldsymbol{a}_{t|n}'+\boldsymbol{P}_{t|n}\}\Big]-\\
&-\tfrac{1}{2}\textstyle\sum_{t=1}^n\text{Tr}[\boldsymbol{\Sigma}_\eta(\Theta)^{-1}\{\boldsymbol{a}_{t|n}\boldsymbol{a}_{t|n}'-\boldsymbol{a}_{t|n}\boldsymbol{\kappa}'-\phi\,\boldsymbol{a}_{t|n}\boldsymbol{a}_{t-1|n}'-\\
&-\boldsymbol{\kappa}\boldsymbol{a}_{t|n}'+\boldsymbol{\kappa}\boldsymbol{\kappa}'+\phi\,\boldsymbol{\kappa}\boldsymbol{a}_{t-1|n}'-\phi\,\boldsymbol{a}_{t-1|n}\boldsymbol{a}_{t|n}'+\phi\,\boldsymbol{a}_{t-1|n}\boldsymbol{\kappa}'+\phi^2\boldsymbol{a}_{t-1|n}\boldsymbol{a}_{t-1|n}'+\\
&\boldsymbol{P}_{t|n}-\phi\,\boldsymbol{P}_{t-1,t|n}-\phi\,\boldsymbol{P}_{t,t-1|n}+\phi^2\boldsymbol{P}_{t-1|n}\}]
\end{aligned} \tag{31}$$

In the steady-state the quantities in (31) converge as follows: $\boldsymbol{a}_{t|n}=\boldsymbol{a}_t+\boldsymbol{P}\boldsymbol{r}_{t-1}$; $\boldsymbol{P}_{t|n}=\boldsymbol{P}-\boldsymbol{P}\boldsymbol{N}_{t-1}\boldsymbol{P}$ ; $\boldsymbol{P}_{t,t-1|n}=\boldsymbol{P}\boldsymbol{L}'(\boldsymbol{I}-\boldsymbol{N}_{t-1}\boldsymbol{P})$. Now, differentiating (31) with respect to, respectively, $\boldsymbol{\kappa}$ and $\phi$, one has:

$$\sum_{t=1}^{n}\{\boldsymbol{a}'_{t|n}-\boldsymbol{\kappa}'-\phi\,\boldsymbol{a}'_{t-1|n}\}\boldsymbol{\Sigma}_{\eta}(\Theta)^{-1} \tag{32}$$

and

$$\begin{array}{c}\sum_{t=1}^{n}\mathrm{Tr}[\{\boldsymbol{a}_{t|n}\boldsymbol{a}'_{t-1|n}-\boldsymbol{\kappa}\boldsymbol{a}'_{t-1|n}-\phi\,\boldsymbol{a}_{t-1|n}\boldsymbol{a}'_{t-1|n}+\\ +\boldsymbol{P}_{t,t-1|n}-\phi\,\boldsymbol{P}_{t-1|n}\}\boldsymbol{\Sigma}_{\eta}(\Theta)^{-1}]\end{array} \tag{33}$$

Imposing these to zero, and considering the quantities in steady-state, one obtains (14) and (15).

# D Proof of Theorem 2

For the proof of Theorem 2 we need the following results. The symbol $E_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi)$ denotes a $p$-dimensional elliptical distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and characteristic function $\phi$.

**Theorem 4** (Representation of an elliptically distributed random vector)**.** *Let $\boldsymbol{x}$ be a $p$-dimensional random vector; then $\boldsymbol{x}$ is distributed as $E_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi)$ with* $\mathrm{rank}(\boldsymbol{\Sigma})=k\leq p$ *if and only if there is a random variable $R\geq 0$, independent of the $k$-dimensional random vector $\boldsymbol{u}$ uniformely distributed on the unit hypersphere $\{\boldsymbol{z}\in\mathbb{R}^k|\boldsymbol{z}^\top\boldsymbol{z}=1\}$ and a $p\times k$ matrix $\boldsymbol{A}$ with $\boldsymbol{A}\boldsymbol{A}^\top=\boldsymbol{\Sigma}$, such that*

$$\boldsymbol{x}\overset{d}{=}\boldsymbol{\mu}+R\boldsymbol{A}\boldsymbol{u}.$$

*Proof.* For the proof see Cambanis et al. (1981) or the monografy by Fang et al. (1987). □

**Theorem 5** (Distribution of the product of two standard normal variables)**.** *Let $(X,Y)$ be a bivariate normal random vector with zero means, unit variances and correlation coefficient $\rho$. Then, the probability density function of $Z=XY$ is*

$$f_Z(z)=\frac{1}{\pi\sqrt{1-\rho^2}}\exp\left[\frac{\rho z}{1-\rho^2}\right]K_0\left(\frac{|z|}{1-\rho^2}\right),$$

*for $-\infty<z<\infty$, where $K_0(\cdot)$ denotes the modified Bessel function of the second kind of order zero:*

$$K_0(x)=\int_0^\infty\cos(x\sinh t)dt=\int_0^\infty\frac{\cos(xt)}{\sqrt{t^2+1}}dt.$$

*Proof.* See Nadarajah and Pogány (2016). □

**Corollary 6.** *If $(X,Y)$ are jointly normally distributed with zero means, unit variances and correlation $\rho$, then the probability* $\Pr(Z>0)=\Pr(X>0\wedge Y>0)+\Pr(X<0\wedge y<0)$ *is given by*

$$\int_0^\infty f_Z(z)dz=1-\frac{\arccos(\rho)}{\pi}=\frac{1}{2}+\frac{\arcsin(\rho)}{\pi}.$$

We are now in the condition to prove Theorem 2.

*Theorem 2.* Let $(X,Y)$ be jointly elliptically distributed with zero location and scale matrix given by

$$\Sigma=\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}.$$

If $(X, Y)$ are jointly normal, then the result is just an application of the Corollary 6:

$$\mathbb{E}[\text{sign}(XY)] = \Pr(XY > 0) - \Pr(XY < 0) = 2\Pr(XY > 0) - 1 = \frac{2}{\pi}\arcsin(\rho).$$

Theorem 4 generalizes this result to all continuous elliptical random variables. Indeed, in our case (bivariate with zero location and unit scale), the representation of that theorem specializes to

$$\begin{bmatrix} X \\ Y \end{bmatrix} = R\boldsymbol{A}\boldsymbol{u} = R\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}\begin{bmatrix} \cos(U) \\ \sin(U) \end{bmatrix},$$

with $U$ uniformly distributed on $[0, 2\pi)$. Since the random variable $R$ is positive (almost surely), the signs of $X$ and $Y$ do not depend on $R$:

$$\Pr(X > 0 \wedge Y > 0) = \Pr(R\boldsymbol{A}\boldsymbol{u} > \boldsymbol{0}) = \Pr(\boldsymbol{A}\boldsymbol{u} > \boldsymbol{0}) = \Pr\left(\tilde{R}\boldsymbol{A}\boldsymbol{u} > \boldsymbol{0}\right) = \frac{1}{2} + \frac{\arcsin(\rho)}{\pi},$$

where $\tilde{R}$ is the random variable that makes the elliptical distribution Gaussian ($\tilde{R}$ is half-normally distributed). □ □

# E Proof of Theorem 3

The first point of Theorem 3 is just an application of the strong of large numbers, since $\nu_{ij} = \mathbb{E}\,\text{sign}(X_{it}X_{jt})$ is finite.

The second point is an application of the classical central limit theorem, as

$$\text{Var}[\text{sign}(X_{it}X_{jt})] = \mathbb{E}\,\text{sign}(X_{it}X_{jt})^2 + [\mathbb{E}\,\text{sign}(X_{it}X_{jt})]^2 = 1 - \nu_{ij}^2.$$

The third point is an application of the continuous mapping theorem, being the sin a continuous function.

The fourth point obtains using the delta method.