Estimation of a Wishart Multivariate Stochastic Volatility Model via Efficient Importance Sampling

Aishameriane Schmidt*

Preliminary Version: Do not cite nor quote

December, 2017

Abstract

In the past decades, multivariate models for volatilities were developed in order to improve estimates and forecasting, specially in the financial applications. As an alternative to multivariate GARCH models, the multivariate stochastic volatility models provide a more flexible structure to accomodate time-variations observed in the correlation structure of the series. In this paper we studied the Wishart MSV model and estimated the model using an algorithm based on the Efficient Importance Sampling (EIS) procedure proposed by Richard & Zhang (2007). As an application, we simulated the selection of the minimum variance portfolio comparing the EIS Wishart MSV model with the multivariate DCC GARCH.

Key-words: Multivariate Stochastic Model. Efficient Importance Sampling. Wishart Process. Portfolio optimization.

Introduction

The problem of modeling conditional volatility has been subject in several studies, specially those in financial areas, like portfolio optimization, asset and derivatives pricing, and risk management. Asset returns have the property of not displaying mean tendencies or shifts (the series in most cases fluctuates around zero) while carrying information in their variances, which makes this series ideal candidates for volatility models. Generalized autoregressive conditional heteroskedasticity (GARCH) and Stochastic Volatility (SV) models were developed in the past years in order to better accommodate the time-varying structure of volatilities in time-series by improving covariance forecasts.

The need to model multiple series simultaneous volatility behavior led to the development of multivariate models, such as the GARCH-in-mean, first proposed by Engle & Kraft (1983) and extended by Bollerslev, Engle & Wooldridge (1988). Modeling the series conjointly has empirical applications: consider, for example, the problem concerning portfolio allocation where one needs to choose how to allocate their investment, i.e.,

*Contact: aishameriane.schmidt@posgrad.ufsc.br

Article for the Econometrics II course - PPGECO UFSC. Professor: André Alves Portela Santos.

one must decide what fraction of their invested money will be allocated in each asset from the portfolio. In order to do that, it is necessary to understand the correlation structure between assets within the portfolio. Moreover, there are evidences of forecasting improvement due to inclusion of time-varying covariances when comparing prediction performance of multivariate volatility structure with simpler models (CHIB; NARDARI; SHEPHARD, 2006).

Although present in many studies, multivariate GARCH models have the drawback of assuming that the contemporary returns variance is a deterministic function of past returns (ASAI; MCALEER; YU, 2006). Multivariate Stochastic Volatility (MSV) models, on other hand, assume a random process to the volatility, allowing that variances are driven by a latent flow of price-relevant information (LIESENFELD; RICHARD, 2003). Philipov & Glickman (2006b) say this framework allows MSV models to better describe stylized facts about returns and variances, such as heavy tails and long memory in comparison to multivariate GARCH models.

The first MSV models used log-volatilities driven by gaussian processes for the univariate series. But this specification does not have a directly interpretation in the multivariate case, as described by Philipov & Glickman (2006b, 2006a), Chib, Omori & Asai (2009), Liesenfeld & Richard (2003), among others. Immune to this problem, the Wishart MSV model allows a direct conection between univariate and multivariate setup (see Uhlig 1997, Philipov & Glickman 2006b, 2006a), since the Wishart distribution is already in matrix form and corresponds to a generalization of the Gamma distribution. In particular, Philipov & Glickman (2006b) have developed a general model of multivariate stochastic volatility driven by a Wishart processes that aims to provide a direct connection between univariate and multivariate models, increasing the flexibility in describing the behavior of stochastic covariances and improving volatility estimates and forecasts.

One problem concerning all MSV models, in particular the Wishart MSV specification, is estimation: since the variance is a non-linear latent component in the model, the model likelihood function depends on high-dimensional integrals that could be challenging to compute (LIESENFELD; RICHARD, 2003). Note that due to the (likely) non-Gaussian nature of the covariance matrices, the Kalman filter cannot be employed. Philipov & Glickman (2006b, 2006a) and Asai & McAleer (2009) employed bayesian methods using Gibbs sampling techniques to estimate the model, but reported problems in estimating jointly the parameters (which they solved by fixing one of them) and with inefficiency factors associated to persistance parameters, respectively.

In this paper, we propose the maximum likelihood estimate of the Wishart MSV model as proposed by Philipov & Glickman (2006a) using the efficient importance sampling (EIS) method from Richard & Zhang (2007). The EIS algorithm is a generic Monte Carlo (MC) techique that is suitable to evaluate high dimension integrals like the ones in the Wishart MSV model. A simulation exercise were performed to see if the algorithm can produce a suitable approximation for the covariance matrix of a synthetic Wishart GDP as well estimate the model's true parameters.

The remainder of the paper is organized as follows. Section 2 presents the MSV model proposed by Philipov & Glickman (2006b) while section 3 describes the EIS procedure both in generic form and applied to the Wishart MSV model. Then, section 4 presents the simulation results. Section 5 concludes.

1 Wishart MSV model

Consider the Wishart MSV model as described in Philipov & Glickman (2006b, 2006a) and Asai & McAleer (2009):

$$y_t = \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}_k(0_k, \Omega_t^{-1}), \tag{1}$$

$$\Omega_t | \nu, \Omega_{t-1} \sim \mathcal{W}_k(\nu, S_{t-1}), \tag{2}$$

$$S_t = \frac{1}{\nu}\ A^{1/2}\ \Omega_t^d\ A^{1/2'}, \tag{3}$$

where:

- y_t is a $k \times 1$ vector containing gaussian random variables observed at time t;
- Ω_t^{-1} is the covariance matrix of y_t at time t;
- ν and S_{t-1} are the degrees of freedom and the scale parameter of the Wishart distribution of dimension k, respectively;
- A is a positive definite symmetric parameter matrix that is decomposed through a Cholesky decomposition as $A = A^{1/2}A^{1/2'}$ and d is a scalar parameter.

The scale parameter of the Wishart distribution, S_{t-1} is the link that allows the covariance matrix Ω_t evolve trough time depending on past covariances. Note that the quadratic form in (3) ensures that the covariances Ω_t^{-1} will be symmetric positive for all t. The parameters A and d are vital to understand the temporal dynamic behavior of the covariance structure (PHILIPOV; GLICKMAN, 2006a).

According to Philipov & Glickman (2006a), the parameter A represents a measure of temporal sensitivity because its components have the information on how the elements of Ω_t depends from the ones in Ω_{t-1}. The magnitude of the elements in the diagonal of A indicates the influence that past variances have on the values from the present, while the values outside the main diagonal represents the influence of the covariances, i.e., reflects the contagion between series. Using the Wishart properties, it is possible to obtain the conditional expectation of Ω_t given Ω_{t-1} to see the effect of A over the expected value of the covariance matrix:

$$\mathbb{E}[\Omega_t | A, \Omega_{t-1}] = \nu\ S_{t-1} = A^{1/2}\,(\Omega_{t-1})^d\,A^{1/2'} \tag{4}$$

It follows that the conditional distributions of the covariance matrices themselves follow the inverse-Wishart distribution and have the conditional expectation given by:

$$\mathbb{E}[\Omega_t^{-1} | A, \Omega_{t-1}^{-1}] = (\nu - k - 1)^{-1}\ S_{t-1}^{-1} = \frac{\nu}{\nu - k - 1} A^{1/2} \left(\Omega_{t-1}^{-1}\right)^d A^{1/2'} \tag{5}$$

where the parameters ν, d and A are as defined earlier.

The scalar d measures as the overall intensity of the covariances relationships across series. It ranges from 0 to 1 and higher values are associated with longer memory of the processes. In financial time series, the presence of long memory implies that today's returns

have a high impact on the returns from many periods ahead (PHILIPOV; GLICKMAN, 2006a). By equation (3) this effect is clear: since the past returns covariance Ω_{t-1} is on the d-th power, a d value near 0 implies that the term Ω_{t-1} will be close to 1 and S_{t-1} enters in (2) depending almost exclusively on A (which itself is time invariant). In the limit case when $d = 0$ we have the constant correlation case, i.e., the process is time-invariant with respect to the correlation. This can be seen using equation (5):

$$\begin{aligned}\mathbb{E}[\Omega_t^{-1}|A, \Omega_{t-1}^{-1}] &= (\nu - k - 1)^{-1}\ S_{t-1}^{-1} \\ &= \frac{\nu}{\nu - k - 1} A^{1/2} \left(\Omega_{t-1}^{-1}\right)^0 A^{1/2'} \\ &= \frac{\nu}{\nu - k - 1} A\end{aligned}$$

On other hand, when $d \approx 1$, the effect of past returns covariance matrix Ω_{t-1} will tend to have full strength. Using $d = 1$ and $A = \mathbb{I}_k$ (with $\mathbb{I}_k$ being the $k \times k$ identity matrix) in (4) we see the special case of a matricial random walk:

$$\mathbb{E}[\Omega_t|A, \Omega_{t-1}] = \nu\ S_{t-1} = \nu\ \frac{1}{\nu}\ \mathbb{I}_k^{1/2} \left(\Omega_{t-1}\right)^1 \mathbb{I}_k^{1/2'} = \Omega_{t-1}$$

Philipov & Glickman (2006a) say that the autoregressive stochastic matrix process for the covariances will be well-defined if the following conditions are met: i) A must be symmetric positive definite, ii) d must lie between 0 and 1, and iii) the degrees of freedom ν must be greater than the number of variables in the model.

As pointed out by Liesenfeld & Richard (2003), one major critique of the MSV models in general is the lack of correspondence between the univariate and multivariate parts. For example, consider the following log-volatilities in a MSV structure, based on Primiceri (2005):

$$y_t = \epsilon_t, \tag{6}$$

$$\epsilon_t = \Omega_t^{-1/2} \xi_t, \qquad \xi_t \sim \mathcal{N}_k(\mathbf{0}_k, 1), \tag{7}$$

$$\Omega_t^{-1} = B_t^{-1} H_t B_t^{-1'}, \tag{8}$$

where H_t is a diagonal matrix whose elements follows a geometric driftless random walk given by

$$ln(h_{it}) = ln(h_{it-1}) + \sigma_i \eta_{it}, \qquad \eta_{it} \sim \mathcal{N}(0, 1), \tag{9}$$

and B_t represents a lower triangular matrix given by:

$$B_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,t} & \beta_{k2,t} & \cdots & 1 \end{bmatrix} \qquad \text{with} \qquad \beta_t = \beta_{t-1} + \upsilon_t, \quad \upsilon_t \sim \mathcal{N}(0, Q). \tag{10}$$

Note that the state transition equation for h_{it} in (9) imposes a gaussian distribution for the log volatilities. To see that this characteristic does not hold for the multivariate setup, consider the special case with 2 series. Then, define $\tilde{B}_t$ and $\tilde{H}_t$ as follows:

$$\tilde{H}_t = \begin{bmatrix} h_{1,t} & 0 \\ 0 & h_{2,t} \end{bmatrix} \quad \tilde{B}_t = \begin{bmatrix} 1 & 0 \\ \beta_{21,t} & 1 \end{bmatrix}$$

with motion laws given by (9) e (10). Now, we can compute $\tilde{\Omega}_t$ using the relation $\tilde{\Omega}_t^{-1} = \tilde{B}_t^{-1}\tilde{H}_t\tilde{B}_t^{-1'}$:

$$\begin{aligned} \tilde{\Omega}_t &= \begin{bmatrix} 1 & 0 \\ -\beta_{21,t} & 1 \end{bmatrix} \cdot \begin{bmatrix} h_{1,t} & 0 \\ 0 & h_{2,t} \end{bmatrix} \cdot \begin{bmatrix} 1 & -\beta_{21,t} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} h_{1,t} & 0 \\ -\beta_{21,t}\ h_{1,t} & h_{2,t} \end{bmatrix} \cdot \begin{bmatrix} 1 & -\beta_{21,t} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} h_{1,t} & -\beta_{21,t}\ h_{1,t} \\ -\beta_{21,t}\ h_{1,t} & \beta_{21,t}^2\ h_{1,t} + h_{2,t} \end{bmatrix} \end{aligned} \tag{11}$$

While the first element of the diagnonal of $\tilde{\Omega}_t$, $h_{1,t}$, follows the same gaussian distribution, this is not true for $\beta_{21,t}^2\ h_{1,t} + h_{2,t}$ (and could be verified for the more general case as well). Besides, in the specific case of the TVP-VAR proposed by Primiceri (2005), changing the order of the variables in the measure equation implies in a diferent distribution for the elements of the new covariance matrix compared to the original setup, implying that changes in the variables lead to different estimates.

The model proposed by Philipov & Glickman (2006b), on other hand, has the advantage of the natural extension of scalar variances into matrices, since the Wishart distribution is a generalization of the Gamma distribution. For a single serie, in the Wishart MSV model, the measure equation (1) became $y_t|\sigma_t^{-2} \sim \mathcal{N}(0, \sigma_t^{-2})$ with $\sigma_t^2|\alpha, \beta_{t-1} \sim \mathcal{G}(\alpha, \beta_{t-1})$. Furthermore, the invariance problem pointed earlier in Primiceri (2005) framework disappears. Without loss of generality, consider again the bivariate case with:

$$A = \begin{bmatrix} a_{11} & a \\ a & a_{22} \end{bmatrix} \qquad \text{and} \quad \Omega_t = \begin{bmatrix} \sigma_{11,t}^2 & 0 \\ 0 & \sigma_{22,t}^2 \end{bmatrix}.$$

This implies that S_t will be given by:

$$\begin{aligned} S_t &= \begin{bmatrix} \sqrt{a_{11}} & \sqrt{a} \\ \sqrt{a} & \sqrt{a_{22}} \end{bmatrix} \cdot \begin{bmatrix} \sigma_{11,t}^{2d} & 0 \\ 0 & \sigma_{22,t}^{2d} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{a_{11}} & \sqrt{a} \\ \sqrt{a} & \sqrt{a_{22}} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{a_{11}}\sigma_{11,t}^{2d} & \sqrt{a}\sigma_{22,t}^{2d} \\ \sqrt{a}\sigma_{11,t}^{2d} & \sqrt{a_{22}}\sigma_{22,t}^{2d} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{a_{11}} & \sqrt{a} \\ \sqrt{a} & \sqrt{a_{22}} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}\sigma_{11,t}^{2d} + a\sigma_{22,t}^{2d} & \sqrt{a_{11}a}\sigma_{11,t}^{2d} + \sqrt{a_{22}a}\sigma_{22,t}^{2d} \\ \sqrt{a_{11}a}\sigma_{11,t}^{2d} + \sqrt{a_{22}a}\sigma_{22,t}^{2d} & a\sigma_{11,t}^{2d} + a_{22}\sigma_{22,t}^{2d} \end{bmatrix} \end{aligned} \tag{12}$$

Now the elements in the diagonal of S_t in (13) (corresponding to the variances) have the same structure, therefore, the same distribution, meaning that if one changes the order of the series, this will permute elements without loss of the adjacent probability distribution. Philipov & Glickman (2006a) summarize the advantages of their model in comparison to the existing MSV-models at that time as bellow:

1. Their framework extends scalar variances into covariance matrices (instead vectors of log-variances from other models), meaning their model is a multivariate extension to the scalar case. They show that some univariate specifications of their general model have a close connection to existing univariate models, like the conditional mean structure presents in the traditional SV models.

2. The time variation of covariance matrices is not solely restricted by changes in variances (this assumption is used, for example, in Cogley & Sargent 2005 among others). The Wishart distribution scale parameter, which underlies the conditional expectation of the asset covariance matrix, allows for both variances and correlations to evolve stochastically over time.

3. Their model allows for the conditional volatility of an asset to depend not only on its past volatility but also on past covariances with other assets. This model formulation incorporates the observed contagion among asset returns into the covariance structure.

4. Most of the (then) existing MSV models are particular cases of Philipov & Glickman's model.

Estimation of the model described in system (1)-(3) involves finding the vector θ containing ν (degrees of freedom of the Wishart distribution), the elements from the matrix A (if A is $k \times k$, then we have ${}^{(k^2+k)}/_2$ elements to estimate, since A is symmetric by construction) plus the time-persistent parameter d. Thus, $\theta = (\nu, d, vec(A))$ with lenght equal to ${}^{(k^2+k)}/_2 + 2$ parameters.

The likelihood associated with the system (1)-(3) is obtained by integrating the joint density of the observable variables from measurement equation (y_t) and the unobserved precision matrices Ω_t with respect to all past precision matrices. Defining $f(\underline{y}, \underline{\Omega}, \theta)$ as the joint density of $\underline{y} = \{y_t\}_{t=1}^T$ and $\underline{\Omega} = \{\Omega_t\}_{t=1}^T$, the likelihood has the form:

$$\begin{aligned} L(\theta|\underline{y}) &= \int f(\underline{y}.\underline{\Omega}, \theta) d\underline{\Omega} \\ &= \int \cdots \int \prod_{t=1}^{T} f(y_t, \Omega_t|\underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) d\Omega_1 \ldots d\Omega_T \end{aligned} \tag{13}$$

with $\underline{y}_t$ and $\underline{\Omega}_t$ given by $\{y_j\}_{j=1}^t$ e $\{\Omega_j\}_{j=1}^t$, respectively.

Using the definitions above, (13) can be written in a more compact notation:

$$L(\theta|\underline{y}) = \int \prod_{t=1}^{T} f(y_t, \Omega_t|\underline{y}_{t-1}, \underline{\Omega}_{t-1}, \theta) d\underline{\Omega}_T \tag{14}$$

Assuming that y_t is independent from Ω_{t-1}, it is possible to factor $f(y_t, \Omega_t | \underline{y}_{t-1}, \underline{\Omega}_{t-1}, \theta)$ as follows:

$$f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) = g(y_t | \Omega_t, \underline{y}_{t-1}; \theta) \cdot p(\Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) \tag{15}$$

The formulation in (15) is a state-space representation of the dynamic variable model. The density $g(\cdot)$ represents the conditional density of all observable variables on time t given their past values and latent variables, while $p(\cdot)$ denotes the transition density of the non-observable states. Using the Wishart MSV model (1)-(3), we have:

$$g(y_t | \Omega_t, \underline{y}_{t-1}; \theta) \propto |\Omega_t|^{1/2} \exp\left\{-\frac{1}{2} y_t' \Omega_t y_t\right\} \tag{16}$$

$$p(\Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) \propto |\Omega_t|^{\frac{\nu-k-1}{2}} \exp\left\{-\frac{1}{2} tr\left[S_{t-1}^{-1} \cdot \Omega_t\right]\right\} \tag{17}$$

In order to evaluate the likelihood in (15) associated with the observable variables y_t, one needs to integrate out the latent states Ω_t from the joint density. But since Ω_t depends from its previous states and enters the model in a non-linear form, the integration problem cannot be solved by standard filtering methods (like de Kalman Filter) (LIESENFELD; RICHARD, 2003). Our proposal is to use the EIS procedure as described in Richard & Zhang (2007) to approximate (15). Next session explains the importante sampling method before introducing the EIS technique in a general way and then in the specific case of the Wishart MSV-Model.

1.1 Importance Sampling

Monte Carlo (MC) methods are procedures of stochastic integration built to approximate solutions for integrals whose analytical solution does not exist and that are too complex to solve with traditional numerical methods. More specifically, MC methods makes use of the central limit theorem and the law of large number to approximate an expectancy by a sample average.

For example, consider the following integral:

$$I = \int_{\chi} g(x) f(x) dx. \tag{18}$$

If x is a random vector with probability density function (p.d.f.) given by f, then equation (18) could be seen as the expectancy of $g(\cdot)$ applied to x:

$$I = \mathbb{E}_f[g(x)]. \tag{19}$$

If it is possible to draw samples from f, one could approximate (19) (thus, (18)) by a sample mean:

$$I \approx \frac{1}{N} \sum_{i=1}^{N} g(x_i). \tag{20}$$

Under regularity conditions, the strong law of large numbers assures that (20) converges almost surely to (19) (GEWEKE, 1989). Implementation of (20) requires a pseudo-random number generator that produces samples from an Uniform density and then (using the inversion theorem) sample from f. In many applications this is not possible, either because the functional form of f is unknown or because the commom support between f and g is small, raising the necessity of other methods to evaluate (18).

The importance sampling (IS) procedure is a MC methods that aims to improve the sampling method in order to reduce the high variance resultant from the lack of commom support between $f(x)$ and $g(x)$. In order to achieve this, one needs to choose an auxiliary density $m(x)$, called *importance sampler*, that allows concentrate the draws in a region containing more values of $g(x)$. Note that equation (18) can be rewritten as:

$$I = \int_{\chi} g(x)\frac{f(x)}{m(x)}m(x)dx = \mathbb{E}_m\left[\frac{g(x)f(x)}{m(x)}\right]. \tag{21}$$

And in the same way as before, expression (21) can be approximated by a sample mean:

$$I \approx I_N^{IS} = \frac{1}{N}\sum_{i=1}^{N}\omega(x_i)g(x_i), \tag{22}$$

where $\omega_i = \frac{f(x_i)}{m(x_i)}$ are the so called *importance weights* and x_i are now draw from $m(\cdot)$. The IS advantage over the traditional MC is that IS, by using the importance weights, corrects the bias that occur when sampling values from f that are not on the same domain as g. Like (20), equation (22) also converges to (18) (assuming the regularity conditions). Since I_N^{IS} is an unbiased estimator for I, its variance will be given by:

$$\sigma_{IS}^2 = \int g^2(x)\frac{f^2(x)}{m(x)}dx - I^2. \tag{23}$$

It is important to note that, in order to obtain convergence of the importance sampler estimator, it is necessary that $\int g^2(x)\frac{f^2(x)}{m(x)}dx < \infty$. This does not occur for cases where the tails from $m(x)$ are lighter than the ones from $f(x)$, which reinforces the relevance of choosing and estimator that assures that (23) is well defined. Further details about the IS procedure can be found in Moura (2010), Robert & Casella (2010) and Durbin & Koopman (2012).

One challenge concerning IS procedures it finding a suitable sampler that reduces the variance in (23). This leads to two different problems: 1) finding a class M of parametric estimators and 2) the selection of the parameters that define a estimator $m \in M$ which minimizes the variance of the estimates (MOURA, 2010). EIS is a generic method to address the second problem, i.e., it searches the best global approximation for the product $g(x)f(x)$ inside a class of parametric densities.

Assuming that one has already choose a parametric class of estimators $M : \{m(x;a); a \in \mathcal{A}\}$ where $\mathcal{A}$ denotes the parameter space, the goal of the EIS algorithm is to search for the optimum value a^* in which (23) is minimum as possible. For example, for a low dimension problem to be solved via EIS, the product $f(x)g(x)$ is approximated by the kernel $k(x,a)$, $a \in \mathcal{A}$ whose relation to the importance sampler is given by:

$$m(x;a) = \frac{k(x;a)}{\chi(a)}, \tag{24}$$

$$\text{where} \quad \chi(a) = \int_X k(x;a)dx \quad \text{is the integration constant.}$$

It is possible to show that a almost optimum value of a can be found solving a least squares problem defined as:

$$(\hat{a}, \hat{\gamma}) = \underset{a \in \mathcal{A}, \gamma \in \mathbb{R}}{\arg\min}\ q(x,a) \tag{25}$$

$$q(x,a) = \int_\chi d^2(x;a) \cdot \omega(x,a) \cdot m(x,a)dx, \tag{26}$$

with

$$d(x,a) = \ln\ f(x)g(x) - \gamma - \ln\ k(x,a), \tag{27}$$

$$w(x,a) = \frac{f(x)g(x)}{m(x,a)}. \tag{28}$$

where the constant γ will be treated as the intercept of the optimization problem.

The EIS algorithm consists in calculating the MC approximation for $q(x,a)$, but, sinnce the importance sampler depends on a, this solution is not a direct one. Richard & Zhang (2007) propose a recursive method where a sequence $\{m(x,\hat{a}_j)\}_{j=0}^{*}$ of estimators that converges to $m(x,a^*)$ is build. For a given $\hat{a}_j \in \mathcal{A}$, a MC estimate for (26) is given by:

$$\bar{q}_N(a|\hat{a}_j) = \frac{1}{N}\sum_{i=1}^{N} \underbrace{\left[\ln f(\tilde{x}_i^j)g(\tilde{x}_i^j) - \gamma - \ln k(\tilde{x}_i^j, a_j)\right]^2}_{d(\tilde{x}_i^j, \hat{a}_j)^2} \cdot w(\tilde{x}_i^j, \hat{a}_j). \tag{29}$$

The values $\{\tilde{x}_i^j\}_{i=1}^N$ are obtained sampling from $m(x,\hat{a}_j)$, which yelds a recursion problem of the form:

$$(\hat{a}_{j+1}, \hat{\gamma}_{j+1}) = \underset{a \in \mathcal{A}, \gamma \in \mathbb{R}}{\arg\min}\ \bar{q}_N(a|\hat{a}_j). \tag{30}$$

with $\bar{q}_N(a|\hat{a}_j)$ as defined in (29).

In cases where the integral being evaluate have higher dimensions, the EIS is generalized to the so called *sequential efficient importante sampling* methods. In cases of big dimensionality, (25) must be replaced by a sequence of lower dimension optimization problems, which is achieved by factoring the target density in a sequence of conditional densities.

Assuming that is possible to form a partition of x in lower dimensional components, i.e., $x = (x_1, x_2, \ldots, x_L)$, then f and g can be written as:

$$g(x) = \prod_{\ell=1}^{L} g_\ell(X_\ell) \qquad \text{e} \qquad f(x) = \prod_{\ell=1}^{L} f_\ell(x_\ell|X_{\ell-1})$$

where $X_\ell = (x_1, \ldots, x_\ell)$.

Also, equation (18) can be written as:

$$I = \int_\chi g(x) \cdot f(x)dx = \int_\chi \prod_{\ell=1}^{L} g_\ell(X_\ell) \cdot f_\ell(x_\ell|X_{\ell-1})dx \tag{31}$$

In this case, the importance sampler is factorized as follows:

$$m(x;a) = \prod_{\ell=1}^{L} m_\ell(x_\ell|X_{\ell-1};a), \quad \text{with} \quad a = (a_1, \ldots, a_\ell) \in A = \prod_{\ell=1}^{L} A_\ell \tag{32}$$

and expression (24), which links $m(x;a)$ with the kernel $k(x;a)$ and the constant $\chi(a)$ becomes:

$$m_\ell(x_\ell|X_{\ell-1};a) = \frac{k_\ell(X_\ell;a_\ell)}{\chi_\ell(X_{\ell-1};a_\ell)}, \quad \text{onde} \quad \chi_\ell(X_{\ell-1};a_\ell) = \int_X k_\ell(X_\ell;a_\ell)dx_\ell. \tag{33}$$

The integral of $g_\ell(X_\ell) \cdot f_\ell(x_\ell|X_{\ell-1})$ with respect to x_ℓ is an unknown function of $X_{\ell-1}$ and cannot be approximated by $m_\ell(x_\ell|X_{\ell-1};a)$, whose integral w.r.t. $X_{\ell-1}$ is equal to one (because is a p.d.f.). In order to solve this, the ℓ-th integrand $g_\ell(X_\ell) \cdot f_\ell(x_\ell|X_{\ell-1})$ is approximated by a kernel $k_\ell(X_\ell;a_\ell)$, which itself depends only from variables of current period. This implies that the integration constant $\chi_\ell(X_{\ell-1};a_\ell)$ isn't affected by the EIS procedure, even thought it could contains information about $X_{\ell-1}$ (MOURA, 2010). As a result, it is possible to "transfer" $\chi_\ell(X_{\ell-1};a_\ell)$ to the least squares optimization in $a_{\ell-1}$, in such a way that the sampler $m_\ell(x_\ell|X_{\ell-1;\hat{a}_\ell})$ will have all the information contained in the sample about the importance region of x_ℓ. Using this, the integral given in (31) can be written as:

$$I = \chi_1(a_1) \int \prod_{\ell=1}^{L} \frac{g_\ell(X_\ell) f_\ell(x_\ell|X_{\ell-1})\chi_{\ell+1}(X_\ell;a_{\ell+1})}{k_\ell(x_\ell;a_\ell)} \cdot \prod_{\ell=1}^{L} m_\ell(x_\ell|X_{\ell-1};a)dx \tag{34}$$

with $\chi_{L+1} \equiv 1$. In order to "get back" all the integration constants in the least squares optimization of a_ℓ, a sequence of recursive procedures going from $\ell = L$ to $\ell = 1$ is implemented as:

$$(\hat{a}_\ell^{k+1}, \hat{\gamma}_\ell^{j+1}) = \underset{a_\ell \in \mathcal{A}_\ell, \gamma_\ell \in \mathbb{R}}{\arg\min} \; \bar{q}_N(X_\ell;a_\ell;\hat{a}_{\ell+1}) \tag{35}$$

with

$$\bar{q}_N(X_\ell;a_\ell;\hat{a}_{\ell+1}) = \frac{1}{N}\sum_{i=1}^{N} d_\ell^2(X_{i,\ell};a_\ell;\hat{a}_{\ell+1}\gamma_\ell) \cdot w_\ell(X_{i,\ell}|X_{i,\ell-1};a_\ell) \tag{36}$$

and

$$d_\ell(X_{i,\ell};a_\ell;\hat{a}_{\ell+1}\gamma_\ell) = \ln[g_\ell(X_{i,\ell})f_\ell(x_{i,\ell}|X_{i,\ell-1})\chi_{\ell+1}(X_{i,\ell};\hat{a}_{\ell+1})] - \gamma - \ln\ k_\ell(x_{i,\ell};a_\ell) \tag{37}$$

Note that equation (37) considers the integration constant $\chi_{\ell+1}(X_{i,\ell};\hat{a}_{\ell+1})$ that refers to the last period, since the procedure is recursive). This term possibly contains all relevant information about the importance region of x_ℓ.

After the sequence of EIS regressions has converged to a fixed point $\hat{a}^*$, a MC estimate for (21) is obtained by the following equation:

$$I \approx \bar{I}_N^{EIS} = \frac{1}{N}\sum_{i=1}^{N}\left[\prod_{\ell=1}^{L}\frac{g_\ell(X_{i;\ell})f_\ell(x_{i;\ell}|X_{i,\ell-1})}{m_\ell(x_{i,\ell}|X_{i,\ell-1};\hat{a}_\ell)}\right], \tag{38}$$

where the sequence $\{X_{i,\ell}\}_{i=1}^N$ is obtained from independent draws from the EIS sampler $m(x;\hat{a}^*)$.

Although EIS is usefull finding the values which minimizes the MC variance, the problem of finding the parametric class $\mathcal{M}$ of estimator remains open. This problem colapses in the choice of the parametric class of EIS density kernels k - once we have a kernel, to find m one just needs to integrate the kernel and set it equal 1 in order to find χ. Back to the Wishart MSV model, the kernel k must provide a close functional approximation to the product in (15), i.e.:

$$f(y_t,\Omega_t|\underline{\Omega}_{t-1},\underline{y}_{t-1};\theta)\chi(\underline{\Omega}_t;\Gamma_{t+1}) = g(y_t|\Omega_t,\underline{y}_{t-1};\theta)\cdot p(\Omega_t|\underline{\Omega}_{t-1},\underline{y}_{t-1};\theta)\chi(\underline{\Omega}_t;\Gamma_{t+1}) \tag{39}$$

where Γ_t is the matrix containing all auxiliary parameters (it is the generalization of a introduced before). The expression in (39) can be used in system (35)-(37) and it is operational and computationally feasible for dimensions T (the sample size) and k (number of variables). Moreover, it is possible to take advantage of the conjugacy between the normal and Wishart densities defined in (16) and (17): their product defines a Wishart density kernel for Ω_t given (Ω_{t-1}, y_t), simplifying even more the problem. Since the Wishart distribution is closed under multiplication (see DeGroot 2005), a natural choice of the EIS kernel k is a Wishart kernel in $\underline{\Omega}_t$, which contains $g(\cdot)p(\cdot)$ and an additional Wishart kernel approximating the remaining terms from (39). Thus, the EIS density kernel will be given by:

$$k(\underline{\Omega},\Gamma_t) = g(y_t|\Omega_t,\underline{y}_{t-1};\theta)\cdot p(\Omega_t|\underline{\Omega}_{t-1},\underline{y}_{t-1};\theta)\zeta(\Omega_t;\Gamma_t) \tag{40}$$

where ζ is a Wishart kernel used to approximate $\chi(\underline{\Omega}_t;\Gamma_{t+1})$ and is parametrized as:

$$\zeta(\Omega_t;\Gamma_t) = |\Omega_t|^{c_{1,t}/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[C_{2,t}\Omega_t\right]\right\}, \qquad \Gamma_t = (c_{1,t},\mathrm{vech}(C_{2,t})) \tag{41}$$

This choice of sampler implies a great simplification in the EIS optimization problem posed in(35)-(37), since the products $g(\cdot)p(\cdot)$ present in the original equation and in the EIS kernel will cancel out.

In order to perform out-of-sample forecasts for a given function of interest $h(\Omega_{T+1},\underline{\Omega}_T)$, we can use the following conditional formula:

$$\begin{aligned}&\mathbb{E}[h(\Omega_{T+1},\underline{\Omega}_T)|\underline{y}_T] = \\ &= \frac{\int\int h(\Omega_{T+1},\underline{\Omega}_T)p(\Omega_{T+1|\Omega_T})\prod_{t=1}^{T}f(y_t,\Omega_t|\underline{\Omega}_{t-1},\underline{y}_{t-1})d\Omega_{T+1}d\underline{\Omega}_T}{\int\prod_{t=1}^{T}f(y_t,\Omega_t|\underline{\Omega}_{t-1},\underline{y}_{t-1})d\underline{\Omega}_T}.\end{aligned} \tag{42}$$

Using (15), (16), (17), we can find the one-step predictive mean $\mathbb{E}[\Omega_{T+1}|\underline{y}_T]$ by defining $h(\Omega_{T+1}, \underline{\Omega}_T) \equiv \Omega_{T+1}$ and substituting all in (42) as follows:

$$
\begin{aligned}
&\mathbb{E}[\Omega_{T+1}|\underline{y}_T] = \\
&= \frac{\int\int \Omega_{T+1}|\Omega_{T+1}|^{(\nu-k-1)/2} \exp\left\{-\frac{1}{2}tr[S_{t-1}\Omega_{T+1}]\right\} \prod_{t=1}^{T} |\Omega_t|^{(\nu-k)/2} \exp\left\{-\frac{1}{2}(y_t'\Omega_t y_t + tr[S_{t-1}^{-1}\Omega_t])\right\} d\Omega_{T+1} d\underline{\Omega}_T}{\int \prod_{t=1}^{T} |\Omega_t|^{(\nu-k)/2} \exp\left\{-\frac{1}{2}(y_t'\Omega_t y_t + tr[S_{t-1}^{-1}\Omega_t])\right\} d\underline{\Omega}_T} \\
&= \int \Omega_{T+1}|\Omega_{T+1}|^{(\nu-k-1)/2} \exp\left\{-\frac{1}{2}tr[S_{t-1}\Omega_{T+1}]\right\} d\Omega_{T+1} \\
&= \mathbb{E}[\Omega_{T+1}|\Omega_T] \\
&= A^{1/2} (\Omega_{t-1})^d A^{1/2'}
\end{aligned} \tag{43}
$$

where the last equality came from (5).

Next section, we will describe the procedure to obtain the likelihood for the Wishart MSV model via EIS algorithm and compare the one-step prediction with a prediction obtained using a multivariate GARCH model.

2 Implementing the sequential EIS to a Wishart MSV Model

For this paper, the EIS algorithm was implemented in R language in order to obtain the likelihood and covariance estimates for 3 series of simulated returns[1] comprising 240 periods.

In order to generate data, the results obtained by Philipov & Glickman (2006a) (page 321) were used as true parameters, i.e, we used $d = 0.5$, $\nu = 14$ and A, where

$$A = \begin{bmatrix} 0.0238 & 0.0057 & 0.0145 \\ 0.0057 & 0.0239 & 0.0056 \\ 0.0145 & 0.0056 & 0.0330 \end{bmatrix}$$

to simulate T precision matrices and the 3 observable variables.

The EIS procedure comprises two big steps: first, we obtain the trajectories in a forward sequence, without using the information of the whole sample and second, we obtain the smoothed estimates using the backward recursion described in the previous section, to transfer information between Ω_t and y_{t+1}. Since this is a method relying on convergence, an stop rule based on the difference between estimated values and maximum number of iterations were used.

The process described above[2] allows to obtain the MC estimate for the likelihood function as defined in (13) as well the estimates for the covariance matrix Ω_t. The next step would be parameter estimation and requires an optmizer that finds θ who minimizes the negative of the computed likelihood. Due to the high computational time demanded in this step, this part was not explored for the current work.

In order to see the performance of the covariance estimates, we used the simulated data y_t in a portfolio optimization problem and compared the returns of each method.

[1] Ideally this would be an empirical application, but the likelihood produced strange values when using real asset returns and due to time restrictions I had to use simulated data.

[2] The code documentation explains better each step

Specifically, after obtaining the covariances via EIS and GARCH modelling, we used them in an minimum variance optimization routine to obtain the portfolio weights for each time t, $t \in \{1, \cdots, 240\}$.

3 Results from the portfolio optimization exercise

The descriptives for the three simulated series are in table (1) and their respective plots are in figure (1). Since this is an exercise to verify the best portfolio, the series from now on will be refered as assets with associated returns. The assets came from a Wishart MSV model and they appear to mimic well the behavior of traditional assets. Since the parameter d was equal to 0.5 in generating the series, it is expected to have a time persistant behavior across the observations, as well a correlation structure imposed by the matrix A (defined in the previous section).

Figure 1 – Plot of the simulated returns

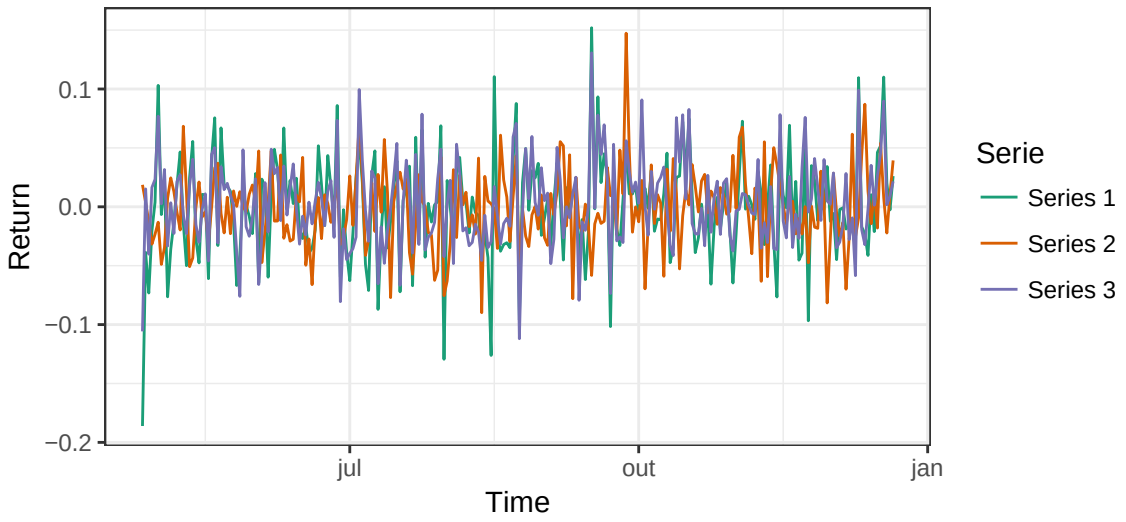


Table 1 – Descriptives of the simulated series

	Series 1	Series 2	Series 3
N. of Observations	240	240	240
Mean return	-0.001	-0.001	0.003
Min. Return	-0.186	-0.090	-0.112
Max. Return	0.152	0.147	0.131
Return std. dev.	0.044	0.033	0.038
Return Kurtosis	1.747	1.316	0.553
Return Skewness	-0.163	0.220	0.268

As benchmark model, we used a Multivariate DCC GARCH model specified as below:

$$r_t = \mu_t + \alpha_t \tag{44}$$

$$\alpha_t = H_t^{1/2} z_t \tag{45}$$

$$H_t = D_t R_t D_t \tag{46}$$

where r_t is a 3×1 vector of returns from the 3 assets at time t; α_t is a 3×1 vector of mean-corrected[3] returns of the three assets at time t; μ_t is the 3×1 vector of expected value of r_t; $H_t^{1/2}$ is a 3×3 matrix at time t such that H_t is the conditional variance matrix of α_t; D_t is a 3×3 diagonal matrix with the standard deviations of α_t at time t; R_t is the 3×3 conditional correlation matrix of α_t at time t and z_t is a vector of white noise. μ_t may be modelled using a time-series model, but since table (1) and figure (1) show, it seems that the returns are centered around zero, so we are not modelling this part. The standard deviations in the diagonal matrix D_t were modelled following univariate GARCH(1,1) models with standard student-t for the errors distribution (the choice was due to the distribution fat tails). For the multivariate distribution, the multivariate t was chosen. In order to obtain the predictions one-step-ahead for the series, we used the *dccroll* function from the rmgarch R package.

One application of multivariate models (GARCH or MSV) is to obtain estimates for the covariances of asset returns in order to select the best asset composition, where best could stands for minimum risk, maximum returns, etc. In this paper we are considering the problem of finding the minimum variance portfolio. Using the same notation as Tsay (2014), let r_t denote the k returns with covariance matrix given by V_t. We are interested in finding the weights of the portfolio, denoted by $w_t = (\omega_{1t}, \ldots, \omega_{kt})$. The return of the porfolio is given by $w_t' r_t$ with variance equals to $w_t' \, V_t \, w_t$. So, the minimization problem to solve in order to find the optimal weights will be given by:

$$\min_w w_t' \, V_t \, w_t \qquad \text{such that} \quad \sum_{i=1}^{k} \omega_i = 1. \tag{47}$$

Using the EIS estimates for the 240 covariance matrices as well the covariance estimates obtained via the DCC model, the weights of the minimum variance portolio were calculated as well the descriptives for the returns over each weighted portfolio. Figures (2) and (3) contains the 240 period portfolio composition for the DCC and MSV-Wishart models. While the covariances estimated via DCC model lead to a composition almost constant over time with the majority part of the investment allocated in the second asset, the minimum variance portfolio using the Wishart MSV model presents more composition fluctuations over time and attributed less investment in the second asset.

The results from table (2) shows that the returns obtained in 240 periods using the Wishart MSV using EIS are negative, on average and lower than the expected returns using the covariances estimated by the DCC model. Although the maximum return from the EIS Wishart MSV model are approximated 11% higher than the maximum return from the DCC model, this may not be an advantage, since the variance of the Wishart MSV portfolio is higher as well.

Using one-step-ahead prediction to built an efficient frontier for each model shows that the DCC portfolio is indeed better when compared to the MSV Wishart portfolio:

[3] This means $\mathbb{E}[\alpha_t] = 0$ and $Cov[\alpha_t] = H_t$

Figure 2 – Weights for the minimum variance portfolio using a multivariate GARCH

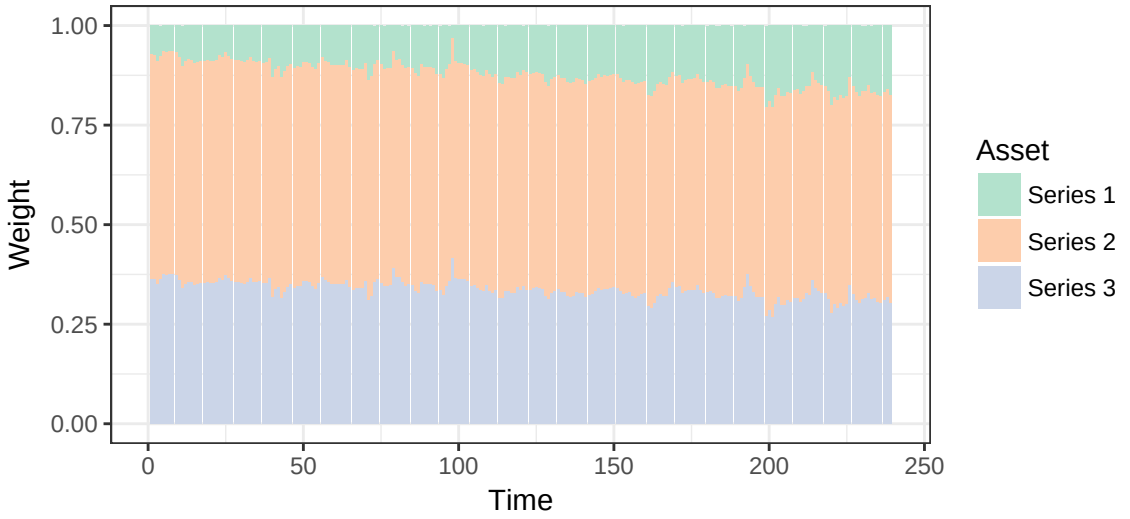
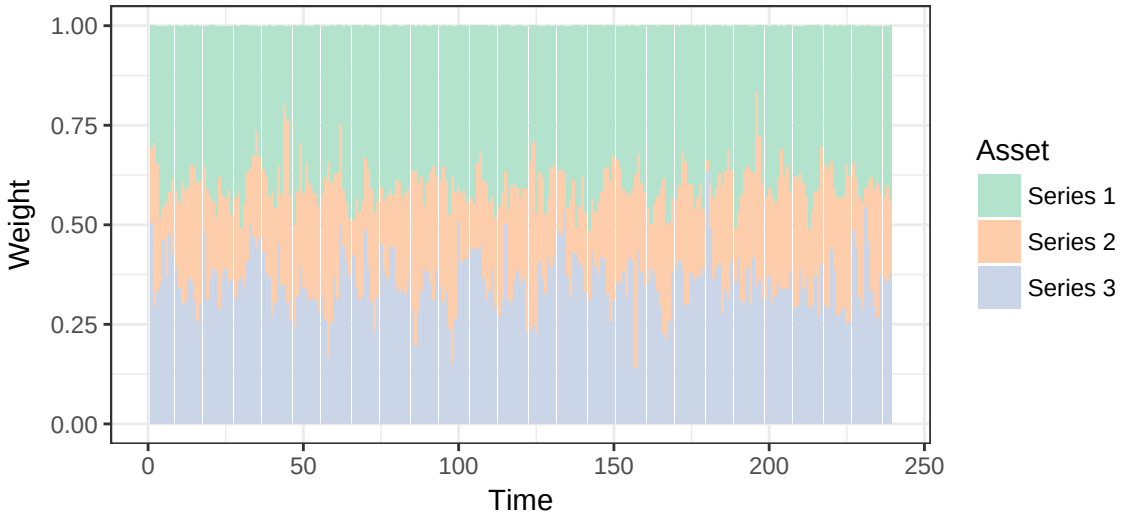


Figure 3 – Weights for the minimum variance portfolio using a Wishart MSV estimated using EIS



while the first presents risk equal to 2.72 and return equal to 0.0894% (figure 4), the latter presented an "optimum" at a risk equal to 1454 and return equal to 0.0508% (figure 5).

Table 2 – Descriptives for the returns obtained from the minimum variance portfolio for each model

	DCC	EIS Wishart MSV
Mean return	0.0001	-0.0004
Min. Return	-0.081	-0.107
Max. Return	0.103	0.115
Return std. dev.	0.025	0.032
Return Kurtosis	1.291	1.132
Return Skewness	0.217	0.196

Figure 4 – Efficient frontier and optimal portfolio using a multivariate GARCH

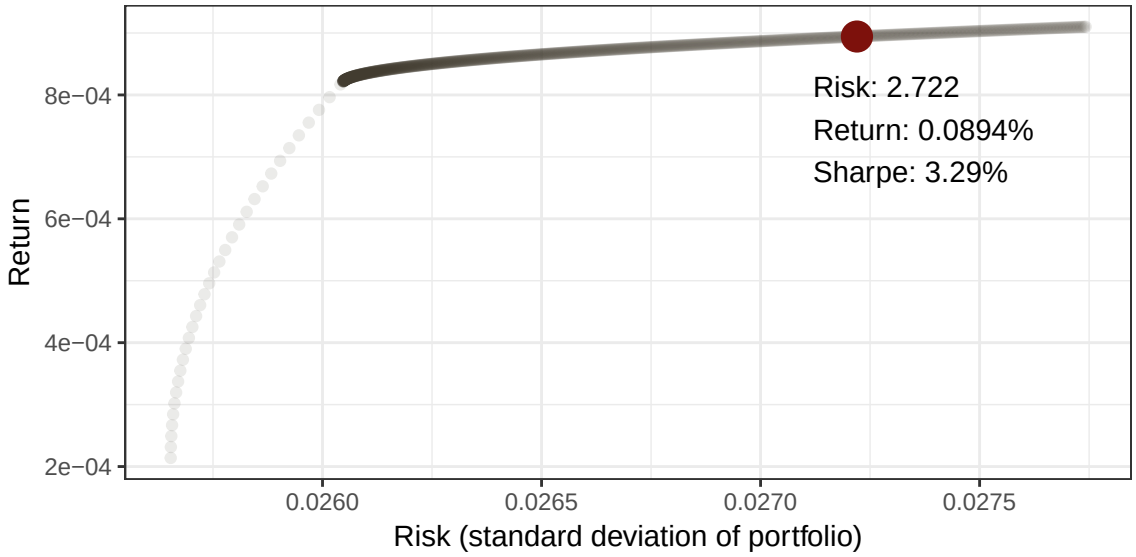
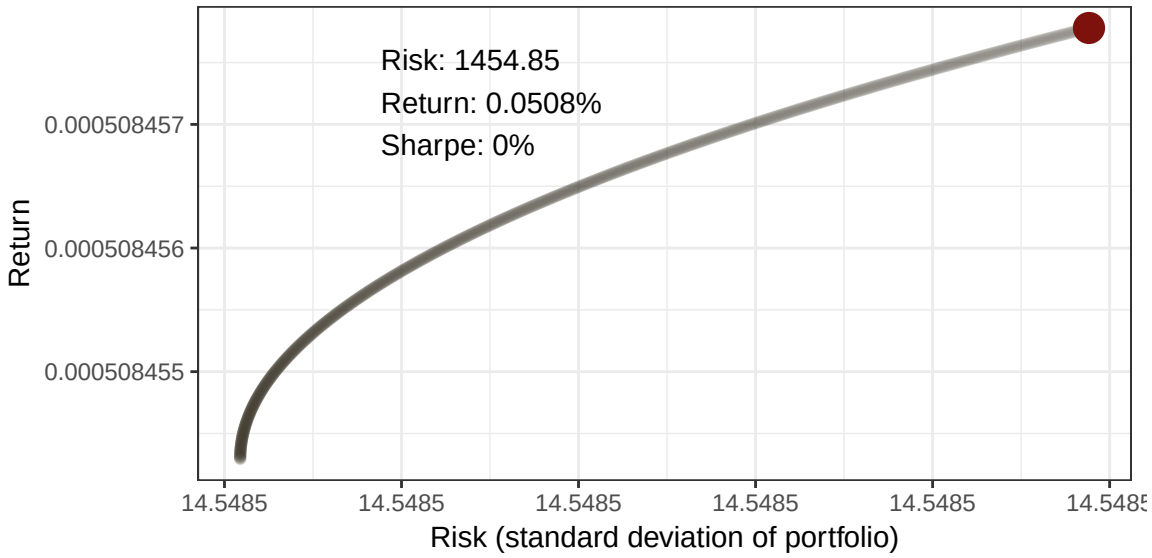


Figure 5 – Efficient frontier and optimal portfolio using a Wishart MSV estimated using EIS



4 Discussion

In this paper we presented the Wishart MSV model as proposed by Philipov & Glickman (2006b) and developed the EIS procedures in order to obtain maximum likelihood estimates. An empirical application using simulated data were performed, using the multivariate DCC GARCH model as benchmark in an optimization portfolio problem framework. The estimated returns from the EIS Wishart MSV model were worse than the ones obtained using the benchmark model, probably due to bad specification of the problem. Further developments include optimizing the EIS routines to decrease computational processing and applications using real data from the stock market.

Bibliography

ASAI, M.; MCALEER, M. The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, Elsevier, v. 150, n. 2, p. 182–192, 2009. Cited 2 times on pages 2 and 3.

ASAI, M.; MCALEER, M.; YU, J. Multivariate stochastic volatility: a review. *Econometric Reviews*, Taylor & Francis, v. 25, n. 2-3, p. 145–175, 2006. Cited on page 2.

BOLLERSLEV, T.; ENGLE, R. F.; WOOLDRIDGE, J. M. A capital asset pricing model with time-varying covariances. *Journal of political Economy*, The University of Chicago Press, v. 96, n. 1, p. 116–131, 1988. Cited on page 1.

CHIB, S.; NARDARI, F.; SHEPHARD, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, Elsevier, v. 134, n. 2, p. 341–371, 2006. Cited on page 2.

CHIB, S.; OMORI, Y.; ASAI, M. Multivariate stochastic volatility. In: *Handbook of Financial Time Series*. [S.l.]: Springer, 2009. p. 365–400. Cited on page 2.

COGLEY, T.; SARGENT, T. J. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics*, Elsevier, v. 8, n. 2, p. 262–302, 2005. Cited on page 6.

DEGROOT, M. H. *Optimal statistical decisions*. [S.l.]: John Wiley & Sons, 2005. v. 82. Cited on page 11.

DURBIN, J.; KOOPMAN, S. J. *Time series analysis by state space methods*. 2. ed. [S.l.]: Oxford University Press, 2012. Cited on page 8.

ENGLE, R. F.; KRAFT, D. Multiperiod forecast error variances of inflation estimated from arch models. *Applied time series analysis of economic data*, Washington, DC:, p. 293–302, 1983. Cited on page 1.

GEWEKE, J. Bayesian inference in econometric models using monte carlo integration. *Econometrica*, v. 57, n. 6, 1989. Cited on page 8.

LIESENFELD, R.; RICHARD, J.-F. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, Elsevier, v. 10, n. 4, p. 505–531, 2003. Cited 3 times on pages 2, 4, and 7.

MOURA, G. V. *Efficient importance sampling in applied econometrics*. Tese (Doutorado), 2010. Cited 2 times on pages 8 and 10.

PHILIPOV, A.; GLICKMAN, M. E. Factor multivariate stochastic volatility via wishart processes. *Econometric Reviews*, Taylor & Francis, v. 25, n. 2-3, p. 311–334, 2006. Cited 5 times on pages 2, 3, 4, 6, and 12.

PHILIPOV, A.; GLICKMAN, M. E. Multivariate stochastic volatility via wishart processes. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 24, n. 3, p. 313–328, 2006. Cited 5 times on pages 2, 3, 5, 6, and 17.

PRIMICERI, G. E. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, Wiley-Blackwell, v. 72, n. 3, p. 821–852, 2005. Cited 2 times on pages 4 and 5.

RICHARD, J.-F.; ZHANG, W. Efficient high-dimensional importance sampling. *Journal of Econometrics*, Elsevier, v. 141, n. 2, p. 1385–1411, 2007. Cited 4 times on pages 1, 2, 7, and 9.

ROBERT, C.; CASELLA, G. *Introducing Monte Carlo Methods with R*. [S.l.]: Springer New York, 2010. Cited on page 8.

TSAY, R. S. *An introduction to analysis of financial data with R*. [S.l.]: John Wiley & Sons, 2014. Cited on page 14.

UHLIG, H. Bayesian vector autoregressions with stochastic volatility. *Econometrica*, v. 65, n. 1, p. 59, 1997. Cited on page 2.