

A Bayesian Approach to Measuring Risk on Portfolios with Many Assets

Samuel Bonello ¹, David Suda ¹, and Monique Borg Inguanez ¹

¹ Department of Statistics and Operations Research, University of Malta, Msida, Malta.
(E-mails:samuel.bonello.17@um.edu.mt,david.suda@um.edu.mt,
monique.inguanez@um.edu.mt)

Abstract. Hedge fund companies typically deal with huge liquid multi-asset portfolios, and modelling the risk of these investments can be challenging. Furthermore, their susceptibility to global market crashes makes modelling their risk even more important. Fitting multivariate models to such portfolios can be challenging given their size, while modelling them univariately runs the risk of ignoring dependencies between the different assets. In this study, a three-stage method for measuring risk on a hedge fund portfolio with many assets is proposed. The first step is that of performing dimension reduction using dynamic principal component analysis which yields orthogonal components that can then be modeled separately avoiding the need to consider multivariate models. This is followed by volatility modelling and forecasting of the individual principal components using a Bayesian generalized autoregressive conditional heteroscedastic (GARCH) model with t -distributed innovations. This allows one to construct a posterior predictive distribution for the whole portfolio. Finally, from this posterior predictive distribution, direct estimation of the risk of the portfolio is obtained using value at risk and expected shortfall. To determine the optimal balance between dimension reduction and accurate forecasts, this method is applied on 4, 11, and 36 dynamic principal components cut-off points determined by the elbow method and the total variation accounted for. Cross-validation over 135 trading days of the different modelling approaches is performed using log pseudo-maximum likelihood as measure of predictive ability. In this case study, it is found that the model with 11 dynamic principal components yields the most accurate forecasts, while the model with 4 principal components yields the least favourable ones.

Keywords: Dimension reduction, dynamic principal components analysis, Bayesian GARCH models, risk measures.

1 Introduction

Over the past few decades, regulations surrounding financial markets have gone through substantial changes to match the pace of technological advancements as well as globalization. Nowadays, a country's stock market can no longer be considered to function independently since a crash in one country has the potential to cause a significant domino effect all around the world. The US subprime mortgage crash in 2008, and China's 'Black Monday' crash in 2015 are just two

7th SMTDA Conference Proceedings, 7-10 June 2022, Athens, Greece

© 2022 ISAST



of the many examples of such global market chain reactions that for the financial industry have been eye openers to the importance of employing an effective model to measure risk. One of the most commonly used tools for measuring the potential losses incurred by financial services as a result of market risk is the Value at Risk (VaR). However, VaR is often criticized for its lack of subadditivity and for the fact that it fails to say anything about the tail-behaviour of the distribution of losses. To overcome these shortfalls, Artzner *et al.* [5] introduced the concept of coherent measures. The simplest and most popular coherent measure is the expected shortfall (ES), sometimes referred to as conditional VaR. Hedge fund companies typically deal with huge liquid multi-asset portfolios and modelling the risk of these investments can be challenging. Furthermore, their susceptibility to global market crashes makes modelling their risk even more important. Fitting multivariate models to such portfolios can be challenging given their size, while modelling them univariately runs the risk of ignoring dependencies between the different assets. Many authors have extended GARCH models to multivariate GARCH (MGARCH) models (see Tsay [18]). When applied to multi-asset portfolios, given the large number of parameters, the likelihood function becomes flat and its optimization becomes increasingly difficult (Orskaug [17]). These difficulties can be overcome using a Bayesian approach. An extensive overview of several Bayesian MGARCH models is provided by Virbickaite *et al.* [19]. Galeano and Ausín [12] note that most of the proposed multivariate models aim at describing the correlation of several return series, with features such as leverage getting very little attention.

Since the univariate GARCH model has been extensively adapted to account for many different phenomena, many analysts prefer to operate in the univariate space. For this reason, Alexander [1] introduced the orthogonal GARCH (O-GARCH) model. The O-GARCH model makes use of principal component analysis (PCA) to reduce the dimensions of the dataset, and then the univariate GARCH model is implemented on each principal component (PC) individually. This method contains a clear drawback - when performing PCA on a time series, although the covariance matrix of the PCs will be diagonal, it has been shown that the PCs still display lagged cross correlations, and thus, treating them as independent is inaccurate (Hörmann *et al.* [16]). This problem can be circumvented by applying dynamic principal component analysis (DPCA) instead of PCA. Contrary to classical PCA, which operates in the time domain, DPCA operates in the frequency domain and obtains the PCs using spectral analysis. As a result, once the PCs are transformed back into the time domain using the inverse discrete Fourier transform, they are uncorrelated across all lags. It is important to acknowledge that like many other techniques, performing DPCA has some disadvantages; the main one being its complexity. Classical PCA and more specifically, the time domain, tends to be much more comprehensible than DPCA and the frequency domain, especially for those that may not have a mathematical background. It is likely the reason why DPCA is not as popular as its classical counterpart. In this study, a method for measuring risk on a hedge fund portfolio with many assets consisting of three main steps is proposed: (1) perform

dimension reduction using DPCA which yields orthogonal components that can then be modeled separately avoiding the need to consider multivariate models; (2) model and forecast volatility of the individual principal components using a Bayesian GARCH model with t-distributed innovations; (3) reconstruct a posterior predictive distribution for the whole portfolio. From this posterior predictive distribution in the third step, direct estimation of the risk of the portfolio can be obtained using value at risk (VaR) and expected shortfall (ES). Finally, a method that mimics the ideas of the popular method of cross-validation is applied to compare the predictive ability of fitted models. Apart from the results given in this paper, supplementary results and material is found in Bonello et al. [7].

2 Dynamic Principal Component Analysis

DPCA is applied in the frequency domain rather than the time domain and hence considers the spectral representation of a time series. The spectral representation of a stationary process X_t decomposes the process into a sum of sinusoidal components with uncorrelated random coefficients, see e.g. Brockwell and Davis [10]). In this text it is assumed that the reader is familiar with the basic concepts of the frequency domain. Let $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{p,t})'$ denote a p -variate asset vector at time t belonging to a stationary multivariate time series with mean vector $\mathbf{0}$ and covariance matrix function $\mathbf{P}_{\mathbf{X}\mathbf{X}}(\cdot)$. The function $f_{mn}(\omega) = \sum_{h=-\infty}^{\infty} \rho_{mn}(h) e^{-i2\pi\omega h}$, $-\frac{1}{2} \leq \omega \leq \frac{1}{2}$ is called the cross spectrum or cross spectral density of the time series $X_{m,t}$ and $X_{n,t}$ where $m, n \in \{1, 2, \dots, p\}$ and $m \neq n$, and $\rho_{mn}(h)$ refers to the covariance between the series $X_{m,t}$ and $X_{n,t}$ at lag h . The $p \times p$ matrix

$$\mathbf{F}_{\mathbf{X}\mathbf{X}}(\omega) = \sum_{h=-\infty}^{\infty} \mathbf{P}_{\mathbf{X}\mathbf{X}}(h) e^{-i2\pi\omega h} = \begin{bmatrix} f_{11}(\omega) & f_{12}(\omega) & \dots & f_{1n}(\omega) & \dots & f_{1p}(\omega) \\ f_{21}(\omega) & f_{22}(\omega) & \dots & f_{2n}(\omega) & \dots & f_{2p}(\omega) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_{m1}(\omega) & f_{m2}(\omega) & \dots & f_{mn}(\omega) & \dots & f_{mp}(\omega) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_{p1}(\omega) & f_{p2}(\omega) & \dots & f_{pn}(\omega) & \dots & f_{pp}(\omega) \end{bmatrix}$$

is the spectral density matrix of \mathbf{X}_t at frequency ω . For a fixed frequency ω , it is desired to find a complex-valued univariate process $Y_t(\omega)$ such that the univariate spectral density $f_Y(\omega)$ is maximized. In other words the aim of DPCA is to find complex vector $\mathbf{c}(\omega)$ of unit length such that:

$$\max_{\mathbf{c}(\omega) \neq \mathbf{0}} \mathbf{c}(\omega)' \mathbf{F}_{\mathbf{X}\mathbf{X}}(\omega) \mathbf{c}(\omega) \quad (1)$$

where $f_Y(\omega) = \mathbf{c}(\omega)' \mathbf{F}_{\mathbf{X}\mathbf{X}}(\omega) \mathbf{c}(\omega)$. Brillinger [8] can be viewed as the pioneer of DPCA. The method that he proposed makes use of discrete Fourier transform (DFT) to jump from the time domain to the frequency domain and the inverse DFT to transform a time series that is represented in the frequency domain into the time domain. Further detail can be found in Brillinger [9]. The following is a brief overview. Let $\{(\lambda_1(\omega), \mathbf{e}_1(\omega)), \dots, (\lambda_p(\omega), \mathbf{e}_p(\omega))\}$ denote the eigenvalue-eigenvector pairs $\mathbf{F}_{\mathbf{X}\mathbf{X}}(\omega)$. Then, the solution to equation (1) is

choosing $\mathbf{c}(\omega) = \mathbf{e}_1(\omega)$, in which case $Y_t(\omega) = \mathbf{e}_1(\omega)' \mathbf{X}_t$ maximises the spectral density $f_Y(\omega) = \lambda_1(\omega)$. The first PC series is then defined as $Y_{1,t} = \sum_{h=-\infty}^{\infty} \mathbf{e}_{1,h}' \mathbf{X}_{t-h}$ where $\mathbf{e}_1(\omega) = \sum_{h=-\infty}^{\infty} \mathbf{e}'_{1,h} e^{-i2\pi h\omega}$. This process can be repeated for k series such that $k \leq p$, where the k^{th} dynamic PC series is formulated as follows:

$$Y_{k,t} = \sum_{h=-\infty}^{\infty} \mathbf{e}_{k,h}' \mathbf{X}_{t-h} \quad (2)$$

and the coherency between $Y_{i,t}$ and $Y_{q,t}$ for $i \neq q$ is zero. In this way, $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{q,t})'$ has spectral density $\mathbf{F}_{\mathbf{Y}\mathbf{Y}}(\omega) = \text{diag}(\lambda_1(\omega), \dots, \lambda_q(\omega))$.

So far, DPCA has been reviewed in the context of a population. Next, a brief overview of the method for performing DPCA when dealing with observed data is provided. Given observations $\mathbf{x}_t = (x_{1,t}, \dots, x_{p,t})'$, equation (3) determines the empirical lagged covariances between $x_{m,t}$ and $x_{n,t}$. More precisely, it determines $\hat{\rho}_{mn}(h)$ for h lags. For a sample of size v , set $\hat{\mu}_m = \frac{1}{v} \sum_{t=1}^v x_{m,t}$ and $\hat{\mu}_n = \frac{1}{v} \sum_{t=1}^v x_{n,t}$ such that

$$\hat{\rho}_{mn}(h) = \frac{1}{v} \sum_{t=1}^{v-h} (x_{m,t+h} - \hat{\mu}_m)(x_{n,t} - \hat{\mu}_n)' \quad (3)$$

The spectral density f is often estimated by a function of the form:

$$\hat{f}_{mn}(\omega) = \frac{1}{2\pi} \sum_{|h| < q} K\left(\frac{h}{q}\right) \hat{\rho}_{mn}(h) e^{-i2\pi h\omega} \quad (4)$$

where $\hat{\rho}(\cdot)$ is the sample autocovariance function, and $K(x)$ is a kernel smoother with window size q . By default, in the *freqdom* package in R which shall be used for DPCA (Hörmann and Kidzinski [15]), $K(x)$ is the Bartlett kernel. This is calculated for all $m, n \in \{1, 2, \dots, p\}$ and $m \neq n$, to estimate the $p \times p$ cross-spectral matrix, $\hat{\mathbf{F}}_{\mathbf{X}\mathbf{X}}(\omega)$. (4) is repeated for all $\omega_j = \frac{j}{n}$, $j = 1, \dots, n$. Next, using singular value decomposition (SVD), one can retrieve the dynamic eigenvectors and eigenvalues of $\hat{\mathbf{F}}_{\mathbf{X}\mathbf{X}}(\omega)$ for each ω_j . Denote the estimated dynamic eigenvectors as $\boldsymbol{\phi}_1(\omega), \dots, \boldsymbol{\phi}_p(\omega)$. To obtain the linear filter, Hörmann *et al.* [16] use the Fourier inverse as follows:

$$\boldsymbol{\phi}_{l,h} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{\phi}_l(\omega) e^{ih\omega} d\omega \quad (5)$$

for $|h| < q$ and $1 \leq l \leq p$. To reduce the dimensions, take $k < p$ PCs. and calculate (5) for $1 \leq l \leq k$. In this way, k vector-valued filters $\{\boldsymbol{\phi}_{l,h}\}$ of dimension $p \times 1$ are estimated for each h , where $-q \leq h \leq q$. Compute the estimated scores for each PC, $\hat{Y}_{1,t}, \dots, \hat{Y}_{p,t}$, using (2) with $\hat{Y}_{k,t} = \sum_{h=-q}^q \boldsymbol{\phi}_{l,h}' \mathbf{x}_{t-h}$, $k = 1, \dots, p$. Finally, using these estimated scores the reconstruction of \mathbf{X}_t is performed using the dynamic Karhunen–Loève expansion:

$$\hat{\mathbf{X}}_t = \sum_{l=1}^k \sum_{h=-q}^q Y_{l,t+h} \boldsymbol{\phi}_{l,h} \quad (6)$$

Ultimately, the aim is to model the estimated $\hat{Y}_{k,t}$'s using the Bayesian GARCH(1,1) model.

3 The Bayesian GARCH(1,1) Model

Nowadays modelling and forecasting the volatility of the returns of a financial asset is crucial given that numerous asset-pricing models use volatility estimates to measure risk, and it is also known that financial time series are typically conditionally heteroscedastic in nature. In this section, the usefulness of the GARCH model in modelling portfolio risk shall be discussed. A detailed discussion is provided on how to infer the parameters of a GARCH(1,1) model with t -distributed innovations using Bayesian methods. Since the GARCH(1,1) model shall be used on the dynamic PCs, in the notation Y_t shall be used to denote the observed series. The classical t -distributed GARCH(1,1) model (Bollerslev, [16]) has the following formulation:

$$\begin{aligned} Y_t &= Z_t \sigma_t; Z_t \stackrel{iid}{\sim} S(0,1, \nu) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \quad (7)$$

where $S(0,1, \nu)$ refers to a scaled Student t -distribution with mean zero, unit variance, and ν degrees of freedom. It is important to note that the standard Student t -distribution is not what is being used here, as that would have a variance of $\frac{\nu}{\nu-2}$ which would skew the conditional variance. Ardia (2008) explains that the model described in (7) has been shown to be problematic when considering Bayesian inference since it is hard to find proposal densities for the parameters when applying the Metropolis Hastings (MH) algorithm. To address this issue, Ardia and Hoogerheide [4] propose the following reparameterization:

$$\begin{aligned} Y_t &= Z_t \left(\frac{\nu-2}{\nu} \omega_t \sigma_t^2 \right)^{\frac{1}{2}}; Z_t \stackrel{iid}{\sim} N(0,1), \omega_t \stackrel{iid}{\sim} IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \quad (8)$$

where $\alpha_0 > 0, \alpha_1, \beta \geq 0$ and $\nu > 2$; $N(0,1)$ denotes the standard normal distribution and $IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ denotes the inverted gamma distribution with shape and scale parameter equal to $\frac{\nu}{2}$. A derivation of the equivalence of (7) and (8) can be found in Bonello *et al.* [7].

To begin to make inferences, define the observed values as $\mathbf{y} := (y_1, \dots, y_T)'$, and the vectors of parameters to be estimated $\boldsymbol{\omega} := (\omega_1, \dots, \omega_T)'$ and $\boldsymbol{\alpha} := (\alpha_0, \alpha_1)'$. The parameters are then grouped such that $\boldsymbol{\theta} := (\boldsymbol{\alpha}', \beta, \nu)'$. Ultimately, it is desired to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. In other words, one wishes to find the posterior density $P(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y})$:

$$P(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}) = \frac{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}) \cdot P(\boldsymbol{\theta}, \boldsymbol{\omega})}{\int \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{y}) \cdot P(\boldsymbol{\theta}, \boldsymbol{\omega}) d\boldsymbol{\theta} d\boldsymbol{\omega}} \quad (9)$$

where the denominator is a marginal likelihood. The priors used are the following: $P(\boldsymbol{\alpha}) \propto \tilde{\phi}_{N_2}(\boldsymbol{\alpha} | \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha), \boldsymbol{\alpha} \in \mathbb{R}^{+2}; P(\beta) \propto \tilde{\phi}_{N_1}(\beta | \mu_\beta, \Sigma_\beta), \beta \in \mathbb{R}^+; P(\boldsymbol{\omega} | \nu) = \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-T} \left(\prod_{t=1}^T \omega_t\right)^{-\frac{\nu}{2}-1} e^{-\frac{1}{2}\sum_{t=1}^T \frac{\nu}{\omega_t}}; P(\nu) = \lambda e^{-\lambda(\nu-\delta)}; \nu > \delta$ - see Ardia

& Hoogerheide [4]. Here $\tilde{\phi}_{N_d}$ is the d -dimensional truncated normal density on \mathbb{R}^d , $\boldsymbol{\mu}_\omega$, μ_β , $\boldsymbol{\Sigma}_\alpha$ and Σ_β are hyperparameters, and $P(v)$ is a translated exponential distribution with parameters $\lambda > 0$ and $\delta \geq 2$. Finally, the joint prior $P(\boldsymbol{\theta}, \boldsymbol{\omega}) = P(\boldsymbol{\alpha}).P(\beta).P(\boldsymbol{\omega} | v).P(v)$ is obtained. The posterior in (9) is estimated empirically using a MCMC simulator with the MH algorithm. Note that, the MCMC sampler for this model was created by Ardia [3], and consists of an algorithm where the GARCH parameters are not updated in one single block. Rather, the GARCH parameters are updated in two separate blocks: one block for $\boldsymbol{\alpha}$ and another block for β . Further details on the MCMC procedure can be found in Ardia [3]. In the next section, the extraction of orthogonal dynamic PC components and the use of Bayesian Student- t GARCH model is discussed. Each of these dynamic PC components are modelled univariately and independently.

4 Modelling a Portfolio with Many Assets

The dataset being used in this study is comprised of 106 assets that are traded in various European exchanges. The original dataset is based on a typical portfolio provided to the authors by a risk management professional in an unnamed financial services company, with the intent of designing a procedure for quick and accurate measurements of risk on portfolios with many assets. The daily adjusted closing prices of each asset were observed for 5 years, from 10/09/2015 to 10/09/2020. All data has been downloaded using Yahoo Finance. The prices are represented by (\mathbf{P}_t) . One of the issues encountered was that since these assets operate in different exchanges, the assets experienced inconsistent trading days, which led to the observations of different time series being asynchronous in time. To overcome this problem, the trading days of all exchanges were compared and only those which were common to all were considered. Another issue encountered was that the assets included in the portfolio are being traded in multiple different currencies (EUR, GBX, SEK, NOK, CHF and DKK). It was opted to standardize the currencies of all assets by converting the value of all assets to EUR using the relevant exchange rates at the time. DPCA was applied to (\mathbf{X}_t) , the log returns of these assets. Fig. 1 displays a scree plot, which provides a summary of the proportion of variation explained by each component in the data. A rule of thumb is to analyze the scree plot and determine the point from which the variance explained seems to level off, known as the ‘‘elbow’’. The components to the left of this point should then be retained. For the given dataset, this seems to happen at around the 4th PC. However, when taking 4 PCs, these are only accounting for 54.66% of the total variation in the original dataset. Thus, apart from the results obtained when using 4 PCs, results for when retaining 11 and 36 PCs shall also be obtained, which explain a minimum of 70% and 90% of the total variation in the original dataset respectively (71.46% and 90.27%, to be precise) shall be obtained for comparison purposes. It would have been ideal to analyse models every k in terms of predictive ability but alas this would have resulted in a very computationally expensive exercise. Given k , the resulting output will be k independent univariate time series represented by $\hat{\mathbf{Y}}_t = (\hat{Y}_{1,t}, \hat{Y}_{2,t}, \dots, \hat{Y}_{k,t})'$, where

$\hat{Y}_{i,t}$ refers to the score of the i^{th} dynamic PC at time t for $i = 1, \dots, k$. As a comparison, one must note that the first PC explains 38.21% of the total variation in \mathbf{X}_t , while the 36th PC explains only 0.40% of the total variation in \mathbf{X}_t . After performing DPCA, one can use the dynamic Karhunen–Loève expansion illustrated in equation (6) to recover an estimation of \mathbf{X}_t from $\hat{\mathbf{Y}}_t$. One can then transform the estimated log returns, $\hat{\mathbf{X}}_t$, into the estimated prices, $\hat{\mathbf{P}}_t$.

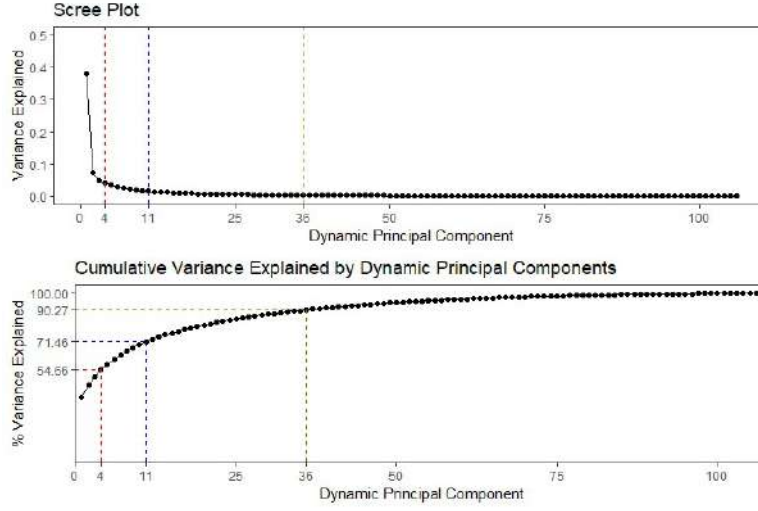


Fig. 1. Scree plot (top) and the cumulative variance (bottom) to display the variance captured by each dynamic PC

When using DPCA on time series data, the dynamic PC obtained after reducing the dimensions are uncorrelated over time. As a result, each dynamic PC can be treated as a univariate time series, allowing one to model the volatility independently. The Bayesian GARCH model is now implemented on all the first 36 dynamic PCs, and results are presented for $k = 1, 4, 11$ and 36. Two chains of length 30000 with a burn-in period of 25000 are run, using the average of the two chains to estimate the empirical probability density function of the parameters. The credibility intervals presented are the highest posterior density (HPD) intervals. For diffuse priors on α and β , we take $\mu_\alpha = (0, 0)'$, $\mu_\beta = 0$, $\Sigma_\alpha = \text{diag}\{1000, 1000\}$ and $\Sigma_\beta = 1000$. On the other hand, for $P(v)$ we take $\lambda = 0.01$ and $\delta = 2$. The parameter α_0 is an intercept term which assures that the volatility is never equal to zero.

In Table 1, one can discern that for larger k , the mean of α_0 shifts closer to 0. This is expected, as for larger k , $Y_{i,t}$ accounts for less and less of the total variation in the data. The role of α_1 , on the other hand, is to approximate the degree to which the conditional volatility reacts to market shocks. According to Alexander [2] (see Section III.4.5.6), when α_1 is around 0.1 or larger, then the volatility is said to be sensitive to market events. In Table 1 it can be seen that at least until the 11th PC, the mean for α_1 has a HPD interval which is greater than 0.1,

indicating that the earlier PCs are rather sensitive to market events. However, it can be seen that for $k = 36$, the density of α_1 shifts closer to zero, indicating that for larger k , dynamic PCs lose their sensitivity. In fact, from the 31st PC onwards, (see Bonello *et al.* [7]) the HPD interval for α_1 lies completely below 0.1. Similar to the previous discussion, since the later PCs explain less of the total variation in the data, the variation due to market events becomes miniscule. In fact, for dimension reduction purposes, one hopes that such important variation is fully explained by a small number of the earlier PCs. β represents the degree to which the volatility of the last period affects the next period's volatility. As can be seen from Table 1, the HPD intervals for $k = 4$ and $k = 11$ are similar, with relatively high numbers. Interestingly, the HPD intervals for β have similar ranges up until the 21st PC. The parameter ν is responsible for the heavy tails of the distribution and is a crucial component in accounting for the leptokurtosis that is present in financial data. In the model, the role of the degrees of freedom parameter ν is to magnify the effect of the conditional variance. In Table 1, it can be seen that for larger k , the HPD intervals become elongated and the corresponding empirical probability density function indicates more dispersion. For further inspection, trace plots and EPDF plots for these parameters can be found in Bonello *et al.* [7]. From the higher order PCs, it was found that the chains for β do not converge, and so the resulting EPDF plot is not accurate. The chains also indicate instability for higher order PCs when it comes to ν .

Table 1. Mean, standard deviation, and HPD interval for the 5,000 realisations generated from the posterior distribution of $\alpha_0, \alpha_1, \beta$ and ν for the PCs $\hat{Y}_{4,t}, \hat{Y}_{11,t}$, and $\hat{Y}_{36,t}$. Full results for all PCs can be found in Bonello *et al.* [7].

	4 th PC	11 th PC	36 th PC
Mean (α_0)	15.87×10^{-5}	6.73×10^{-5}	6.67×10^{-5}
Standard Deviation (α_0)	4.11×10^{-5}	2.22×10^{-5}	1.80×10^{-5}
95% HPD L.B. (α_0)	8.46×10^{-5}	2.98×10^{-5}	3.07×10^{-5}
95% HBD U.B. (α_0)	23.68×10^{-5}	11.06×10^{-5}	9.60×10^{-5}
Mean (α_1)	0.1605	0.1286	0.0327
Standard Deviation (α_1)	0.0279	0.0287	0.0254
95% HPD L.B. (α_1)	0.1132	0.0777	1.07×10^{-5}
95% HPD U.B. (α_1)	0.2210	0.1878	0.0819
Mean (β)	0.7578	0.7654	0.2671
Standard Deviation (β)	0.0382	0.0541	0.1834
95% HPD L.B. (β)	0.6834	0.6598	0.0002
95% HPD U.B. (β)	0.8271	0.8579	0.6174
Mean (ν)	10.9650	24.2921	23.1302
Standard Deviation (ν)	2.8084	23.4420	17.8266
95% HPD L.B. (ν)	6.5518	6.6321	6.8372
95% HPD U.P. (ν)	16.7609	76.1065	61.0999

Now that an MCMC sample for each parameter has been generated, it is possible to move on to constructing forecasts for the given portfolio. First, however, the theoretical framework for these forecasts is first provided in Section 5.

5 Forecasting and Risk Estimation

In this section, the procedure for obtaining forecasts for each PC is introduced, which are then transformed into forecasts for log returns of each asset. Finally, these are converted to asset prices and ultimately, the value of the portfolio. One can then apply the ‘direct’ approach on these forecasts to estimate VaR and ES. Consider G estimated MCMC realizations. Then for each $\alpha_0^{(g)}$, $\alpha_1^{(g)}$, $\beta^{(g)}$, and $\nu^{(g)}$, one can obtain the following forecasts for the dynamic PCs: $\hat{Y}_{i,T+m}^{(g)} = Z_{i,T+m}^{(g)} \left(\frac{\hat{\nu}_{(i)}^{(g)-2}}{\hat{\nu}_{(i)}^{(g)}} \omega_{i,T+m}^{(g)} \hat{\sigma}_{i,T+m}^{2(g)} \right)^{\frac{1}{2}}$, $Z_{i,T+m}^{(g)} \stackrel{iid}{\sim} N(0,1)$, $\omega_{i,T+m}^{(g)} \stackrel{iid}{\sim} IG\left(\frac{\hat{\nu}_{(i)}^{(g)}}{2}, \frac{\hat{\nu}_{(i)}^{(g)}}{2}\right)$ and $\hat{\sigma}_{i,T+m}^{2(g)} = \hat{\alpha}_{0(i)}^{(g)} + \hat{\alpha}_{1(i)}^{(g)} \hat{Y}_{i,T+m-1}^{(g)2} + \hat{\beta}_{(i)}^{(g)} \hat{\sigma}_{i,T+m-1}^{2(g)}$ for $g = 1, 2, \dots, G$ and $m = 1, 2, \dots, M$, where $\hat{Y}_{i,T}^{(g)} \equiv \hat{Y}_{i,T}$. Furthermore, from $\hat{Y}_{i,T+m}^{(g)}$, the forecasts $\hat{X}_{i,T+m}^{(g)}$, $m = 1, \dots, M$ of the log returns of the original assets are retrieved for a specified number of dynamic PCs using the Karhunen–Loève expansion in (6). Consequently, from these, the forecasts of the prices $\hat{P}_{i,T+m}^{(g)}$, $m = 1, \dots, M$ are easily recovered. Finally, one can obtain forecasts for the value of the portfolio by assigning the relevant weightings, τ_1, \dots, τ_p , to each asset:

$$\hat{V}_{T+m}^{(g)} = \tau_1 \hat{P}_{1,T+m}^{(g)} + \tau_2 \hat{P}_{2,T+m}^{(g)} + \dots + \tau_p \hat{P}_{p,T+m}^{(g)}$$

for $m = 1, \dots, M$. The simulated sample of size G of the given portfolio’s value at time $T + m$ are combined into a vector i.e. $\hat{\mathbf{V}}_{T+m} := (\hat{V}_{T+m}^{(1)}, \hat{V}_{T+m}^{(2)}, \dots, \hat{V}_{T+m}^{(G)})'$ for $m = 1, \dots, M$, with which one can calculate the m -step ahead Bayesian estimation of VaR and ES. The direct approach in Hoogerheide and van Dijk [14] is used. This approach involves obtaining the forecasted loss at time $T + m$. In order to do this one may simply calculate $\hat{L}_{T+m}^{(g)} = \hat{V}_{T+m}^{(g)} - V_T$ for all $g = 1, \dots, G$. Thus, at confidence level $\alpha \in (0,1)$, the m -step ahead estimate for the VaR of the given portfolio will be obtained through the sample α -quantile q_α obtained from $\hat{\mathbf{L}}_{T+m} = (\hat{L}_{T+m}^{(1)}, \hat{L}_{T+m}^{(2)}, \dots, \hat{L}_{T+m}^{(G)})'$: $\widehat{\text{VaR}}_\alpha(\mathbf{L}_{T+m}) = -q_\alpha(\hat{\mathbf{L}}_{T+m})$. The m -step ahead estimate for the ES of the given portfolio will be obtained as follows:

$$\widehat{\text{ES}}_\alpha(\mathbf{L}_{T+m}) = -\frac{1}{\alpha} \left\{ \frac{\sum_{g=1}^G \hat{L}_{T+m}^{(g)} \mathbb{1}_{\{\hat{L}_{T+m}^{(g)} \leq q_\alpha(\hat{\mathbf{L}}_{T+m})\}}}{\sum_{g=1}^G \mathbb{1}_{\{\hat{L}_{T+m}^{(g)} \leq q_\alpha(\hat{\mathbf{L}}_{T+m})\}}} \right\}.$$

The implementation of the forecast of the value of the portfolio for $k = 4, k = 11$ and $k = 36$ shall now be seen. Consequently, the value at risk and expected shortfall for different time horizons is estimated. Due to space limitations, the forecasts of $\hat{\sigma}_{i,T+m}^{2(g)}$ and $\hat{P}_{i,T+m}^{(g)}$ shall not be displayed neither here nor on

Bonello *et al.* [7], but can be requested to the authors if needed. The typical portfolio weightings which were provided to the authors by a hedge fund company also included the number of units to invest into each asset, and these can also be accessed in the same link. These weightings are a combination of buy and short positions. A short position means that one is borrowing the asset from a financial institution such that a fall in the price of the asset which one has shorted will increase the value of the given portfolio. In total, on 10/09/2015, -€111,992,327 have been invested in this portfolio. To gain some insight on the portfolio, the reader is referred to Fig. 2 which portrays how the value of the portfolio varied over the 5 years. Note that a huge spike was experienced in the value during the beginning of 2016, which is a direct result of the fluctuations of one particular asset which has a very large weighting in the given portfolio, namely SPMR.MI in which a short position of -750,003 units is held, however it is relatively stable otherwise.

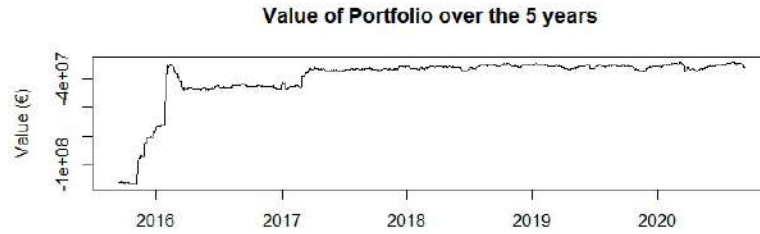


Fig. 2. The value of the given portfolio over the past 5 years.

Table 2. Mean, standard deviation, and HPD interval for the 5,000 realisations of the 1 and 5 step ahead value of the given portfolio when taking $k = 4$, $k = 11$ and $k = 36$ dynamic PCs.

\hat{V}_{1186+1}	$k = 4$	$k = 11$	$k = 36$
Mean	-30,598,530	-30,598,530	-30,598,530
Standard Deviation	87,127.01	142,472.90	226,486.40
95% HPD L.B.	-30,766,041	-30,670,470	-30,744,850
95% HPD U.B.	-30,425,646	-30,112,297	-29,866,495
\hat{V}_{1186+5}	$k = 4$	$k = 11$	$k = 36$
Mean	-30,826,255	-30,459,983	-30,320,170
Standard Deviation	321,099	473,519	609,058
95% HPD L.B.	-31,491,554	-31,356,762	-31,498,444
95% HPD U.B.	-30,222,764	-29,515,981	-29,141,408

Table 2 displays the descriptive statistics for the distributions of \hat{V}_{1186+} and \hat{V}_{1186+5} , respectively. EPDF plots for \hat{V}_{1186+m} for $m = 1, \dots, 5$ when taking $k = 4$, $k = 11$ and $k = 36$ are provided in Bonello *et al.* [7]. Furthermore, as one takes more PCs, more variation in the dataset is being accounted for. As a result, the standard deviation of the forecasted portfolio value increases with the number of

PCs, capturing the true variation of the portfolio. Finally, in Table 3, measures of risk at different time horizons are estimated. To do this, the losses/gains experienced for each MCMC realization when compared to V_{1186} are extracted, and $\widehat{VaR}_\alpha(\mathbf{L}_{T+m})$ and $\widehat{ES}_\alpha(\mathbf{L}_{T+m})$ are extracted. When analysing Table 3, a clear discrepancy between the estimations using different k 's is noticed, however this is not surprising. Models with a higher number of dynamic PCs account for more variation of the original portfolio, which consequently increases VaR and ES.

Table 3. Estimated 1-, 5-, 10-, 50-, and 100-day ahead $VaR_{0.05}$ and ES when taking $k = 4$, $k = 11$ and $k = 36$ dynamic PCs. The percentages reflect the estimated fluctuations from the last observed value of the portfolio, i.e. V_{1186} .

	No. of PCs	Estimated $VaR_{0.05}$	Estimated ES
one-day ahead	4	-4.64%	-4.79%
	11	-5.65%	-5.91%
	36	-6.62%	-7.03%
five-day ahead	4	-5.31%	-5.85%
	11	-7.55%	-8.26%
	36	-9.30%	-10.31%
10-day ahead	4	-6.01%	-6.85%
	11	-9.49%	-10.71%
	36	-11.45%	-13.08%
50-day ahead	4	-10.44%	-12.48%
	11	-16.31%	-19.37%
	36	-21.97%	-25.93%
100-day ahead	4	-13.26%	-16.06%
	11	-21.11%	-25.06%
	36	-28.37%	-33.65%

In the following section, the different models shall be put to the test by comparing them with actual future values, and checking the performance between difference number of dynamic PCs.

6 Measuring Predictive Ability

Since the main aim of the analysis here is to find a model that can adequately predict future risk one needs to evaluate the predictive ability of the Bayesian models fitted. This was done by applying a method which mimics the idea of cross-validation methods popularly applied in non-time series statistical models. Recall from the previous section that we have fitted the different models on a portfolio whose assets were observed between 10/09/2015 to 10/09/2020 we shall refer to these series as the training set. As the validation set, the observations between 11/09/2020 to 04/05/2021 shall be taken, where all assets in the portfolio were simultaneously observed for a total of 135 days. The models for $k = 4$, $k = 11$ and $k = 36$ shall be forward-tested and compared. The models fitted are used

to obtain simulated paths, denoted by $\hat{\mathbf{V}} = \{\hat{V}_{1186+1}, \dots, \hat{V}_{1186+135}\}$, and these are compared to the actual values making up the validation set, denoted by $\mathbf{V} = \{V_{1186+1}, \dots, V_{1186+135}\}$. Fig. 6 displays $\hat{\mathbf{V}}$ and \mathbf{V} for all models.

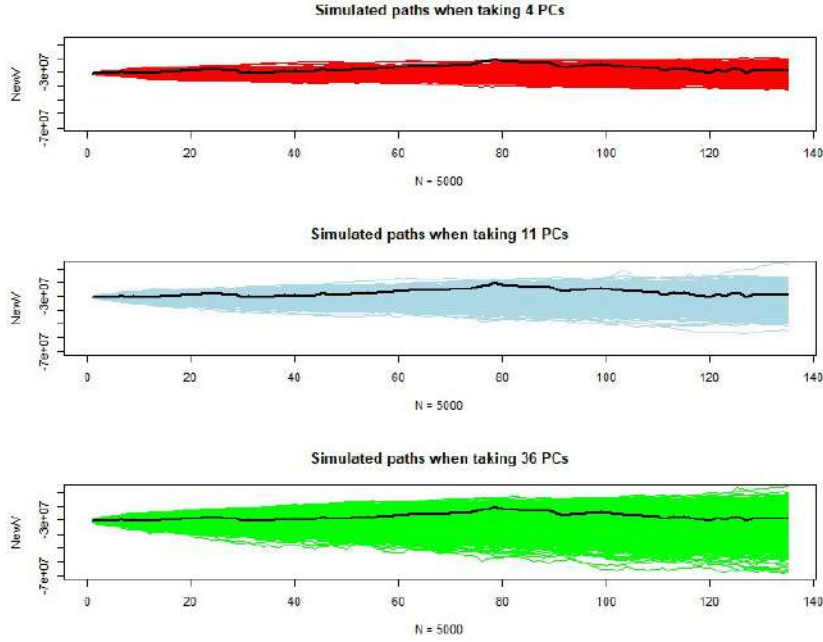


Fig. 6. The red, blue and green coloured paths represent the 5000 simulated paths forecasted till 04/05/2021 when taking $k = 4, k = 11$ and $k = 36$ dynamic PCs. The black path represents the true value of the given portfolio from 10/09/2020 to 04/05/2021

Table 4. Displaying the log PsML for the model when taking taking $k = 4, k = 11$ and $k = 36$ dynamic PCs.

	4 PCs	11 PCs	36 PCs
log PsML	-2338.27	-2210.67	-2213.38

Using the terminology given in Gelfand *et al.* [13], the posterior probability of observing V_{1186+m} is referred to as the conditional predictive ordinate (CPO). One can denote the CPO as $p(V_{1186+m} | \hat{V}_{1186+m})$ for $m = 1, \dots, 135$. Higher CPO values are an indication of a better predictive ability. To measure the predictive ability of the entire model, Congdon [11] suggests to look at the product of the CPOs $\hat{p}(\mathbf{V}) = \prod_{m=1}^{135} p(V_{1186+m} | \hat{V}_{1186+m})$. This is referred to as the pseudomarginal likelihood (PsML). It may sometimes be easier to calculate the log PsML, given by $\log \hat{p}(\mathbf{V}) = \sum_{m=1}^{135} \log p(V_{1186+m} | \hat{V}_{1186+m})$ where the larger the log PsML, the better the predictive ability. By comparing the spread of the paths when taking different PCs in Fig. 6, it is observed that the more PCs one

takes, the more conservative the model becomes, with the paths taking on a larger spread. As a result, during stable market conditions, one should expect a lower PC model to still perform well. Table 4 displays the log PsML for the different models. Since the log PsML is lowest when 4 PCs are taken, one may conclude that it has the worst predictive ability. The fact that the 4 PC model performed the worst on the validation set (not used in the model fit) even though the value of the portfolio did not experience any extreme variation allows one to conclude that not enough variation of the original dataset is being accounted for when taking few PCs. Interestingly, the 11 PC model slightly outperforms the 36 PC model, when compared to the true observations. While this may simply be due to the fact that the 135 observation validation set does not represent any unstable market conditions, it is a possibility that the additional PCs did not represent a significant source of variation and may have therefore caused only a slight detrimental effect to the model. However, it must also be noted that the difference between log PsML when $k = 11$ and log PsML when $k = 36$ is marginal, compared to the considerably inferior log PsML when $k = 4$ case. Of course, this result is valid only for this particular validation set, and considering different time periods and different portfolios may yield different outcomes. However, it is encouraging that one may still get higher predictive ability with a much-reduced dimension.

7 Conclusion

In this study, a Bayesian approach to measuring risk on portfolios with many assets is taken. The idea behind this study is that of going for a dimension reduction approach whereby the orthogonal series can be modelled independently, rather than deal with the dependency dynamics of multivariate approaches which can become unwieldy when dealing with a large number of assets. Inevitably, however, when dealing with a model in reduced dimension, some of the variation of the original portfolio may be lost. We have seen that this may not necessarily lead to an inferior model – indeed, it was found that for the validation dataset, the best predictive ability was that of the model considering 11 dynamic PCs. The model considering 36 dynamics PCs, on the other hand, was only marginally inferior. Furthermore, the MCMC algorithm was found to be less stable when modelling dynamic PCs of a higher order, though this was not expected to have a major impact on the results, as higher order PCs account for a very small proportion of the variation. Further research within this direction that can be undertaken is to consider Bayesian models (of GARCH type or otherwise) which take into account asymmetry in the distribution when it comes to positive and negative shocks, as the Bayesian GARCH(1,1) does not, and to look into whether the above methodology of modelling many assets can be extended to the portfolio optimization problem.

Acknowledgements We would like to thank Mr. George Camilleri, risk management professional, for his contributions to this study.

References

1. Alexander, C. (2003). Principal Component Models for Generating Large GARCH Covariance Matrices. *Economic Notes*, 31, 337–359. <https://doi.org/10.1111/1468-0300.00089>
2. Alexander, C. (2008). *Market Analysis III: Pricing, Hedging and Trading Financial Instruments*. John Wiley & Sons.
3. Ardia, D. (2008). *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*. Springer-Verlag.
4. Ardia, D., and Hoogerheide, L. (2010). Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations. *Tingbergen Institute Discussion Papers*, 2. <https://doi.org/10.32614/RJ-2010-014>
5. Artzner, P., Delbaen, F., Jean-Marc, E., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9, 203–228. <https://doi.org/10.1111/1467-9965.00068>
6. Bollerslev, T.. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 69(3), 542-547. <https://doi.org/10.2307/1925546>
7. Bonello, S., Suda, D. and Borg Inguanez, M. – Supplementary material for paper ‘A Bayesian Approach to Measuring Risk on Portfolios with Many Assets.’ https://github.com/davidsuda80/bonello_suda_inguanetz. Last accessed 5/7/2022.
8. Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, R&W.
9. Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory (Expanded Edition)*. Holden-Day.
10. Brockwell, P. J., and Davis, R. A. (1991). *Time Series: Theory and Methods (2nd ed.)*. Springer-Verlag.
11. Congdon, P. (2007). *Bayesian Statistical Modelling*. John Wiley & Sons.
12. Galeano, P., and Ausín, C. (2010). The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection. *Journal of Business & Economic Statistics*, 28, 559–571. <https://doi.org/10.1198/jbes.2009.07238>
13. Gelfand, A., Dey, D., and Chang, H. (1992). Model Determination using Predictive Distributions with Implementation via Sampling-based Methods. *Bayesian Statistics: Proceedings of the Fourth Valencia International Meeting*, April 15-20, 147-167.
14. Hoogerheide, L., and van Dijk, H. K. (2010). Bayesian Forecasting of Value at Risk and Expected Shortfall using Adaptive Importance Sampling. *International Journal of Forecasting*, 26(2), 231-247. <https://doi.org/10.1016/j.ijforecast.2010.01.007>.
15. Hörmann, S., and Kidziński, Ł. (2022). Package ‘freqdom’. CRAN. <https://doi.org/10.1111/rssb.12076>
16. Hörmann, S., Kidziński, Ł., and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 77(2), 319–348. <https://doi.org/10.17877/DE290R-4881>
17. Orskaug, E. (2009). *Multivariate DCC-GARCH Model: -With Various Error Distributions (Master’s Thesis)*. Retrieved from Norwegian Computing Center (Accession Order No. SAMBA/19/09).
18. Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons.
19. Virbickaite, A., Ausín, M. C., and Galeano, P. (2015). Bayesian Inference Methods for Univariate and Multivariate Garch Models: A Survey. *Journal of Economic Surveys*, 29(1), 76–96. <https://doi.org/10.1111/joes.12046>