Discussion Paper Series

A Kernel Technique for Forecasting the Variance-Covariance Matrix

By

Ralf Becker, Adam Clements and Robert O'Neill

Centre for Growth and Business Cycle Research, Economic Studies,
University of Manchester, Manchester, M13 9PL, UK

November 2010
Number 151

A Kernel Technique for Forecasting the Variance-Covariance Matrix.

Ralf Becker[†], Adam Clements[‡] and Robert O'Neill[†]
[†]Economics, School of Social Sciences, University of Manchester
[‡]School of Economics and Finance, Queensland University of Technology

October 28, 2010

Corresponding author
Ralf Becker
Economics, School of Social Sciences
University of Manchester

email ralf.becker@manchester.ac.uk
Ph +44 (0)161 275 4807.

1 Introduction

The forecasting of variance-covariance matrices (VCMs) is an important issue in finance, having applications in portfolio selection and risk management as well as being directly used in the pricing of several financial assets. In recent years an increasing body of literature has developed multivariate models to forecast this matrix, these include the DCC of Engle and Sheppard (2001), the VARFIMA model of Chiriac and Voev (2009) and Riskmetrics of J.P. Morgan (1996). All of these models can be used to forecast the VCM of a portfolio and all do so using data only relating to the performance of the stocks under consideration.

Previous studies, focusing on modelling the volatility of single assets, have identified economic variables that may influence the variance of returns and attempted to utilise such variables in forecasting. For example Aït-Sahalia and Brandt (2001) investigate which of a range of factors influence stock volatility. In this paper we introduce this approach to a multivariate setting as we introduce a technique for utlilising macroeconomic variables when forecasting the VCM of a number of stocks.

Advances in multivariate volatility modelling are complicated by the requirement that all forecasts of VCMs must be positive-definite and symmetric, restrictions which have meant it has also been difficult for models to incorporate macroeconomic information. Multivariate models have also encountered estimation problems as the dimensions of the VCM are allowed to grow, the technique presented here is able to satisfy all of the required constraints and can be apply it to a relatively large portfolio of stocks.

In this paper we introduce a non-parametric approach to forecasting that is similar in technique to the Riskmetrics approach. However, while the latter delivers a weighted avearge of past realized VCM with the weights determined by the distance of obervations to the time at which the forecast is made, we introduce a weighting approach that allows for a wider range of variables to determine the weights. We allow statistics that measure how similar matrices are and macroeconomic information (such as interest rates) to be used in addition to the time information used in the Riskmetrics approach. Technically we use a multivariate kernel to obtain such weights.This approach builds on the work by Clements and Becker (2010), who show in a univariate setting that employing kernels to determine weight structures dependent on the similarity of volatility observations through time can improve forecast accuracy when compared to more established methods. As the method essentially calculates VCM forecasts as weighted averages of past VCM it guarantees symmetry and positive-definitness by construction.

We apply our proposed methods to a large real life dataset. As the method makes use of a potentially large set of exogenous information it is impossible to devise a representative simulation setup that could serve to establish the usefullness of the proposed forecasting tool. Therefore we provide a careful forecasting experiment in which we compare our method with other models. The results of this forecasting experiment are promising in that they establish that our nonparametric approach is able to produce forecasts of the VCM that can be statistically superior. Interestingly we can demonstate that macroeconomic data are critical to this improvement, providing evidence that these variables provide important information for predicting the behaviour of the VCM.

The rest of this paper is organised as follows, Section 2 introduces some notation and assumptions. Section 3 reviews the literature surrounding multivariate modelling, nonparametric econometrics and the relationship between macroeconomic variables and stock return volatility. Section 4 describes how Riskmetrics, a popular volatility forecasting tool, can be described as a kernel approach based on time. Section 5 outlines how our model uses kernel techniques to obtain forecasts of the VCM using a wider range of data while section 6 introduces the variables we include in our model. Section 7 outlines our forecasting experiment and provides a discussion of our results. Finally section 8 concludes and notes further areas of interest.

2 Notations and Assumptions

The model we present is used to forecast the volatility of stock returns in an n stock portfolio. For any given day, t, the $(n \times 1)$ vector of returns is denoted by $\mathbf{r}_t = (r_{1t}, ..., r_{nt})'$, where r_{it} is the return on stock i on day t, and we assume that given all information available at time $t-1$, $\mathcal{F}_{t-1}$, the mean is unforecastable, i.e. $E(\mathbf{r}_t|\mathcal{F}_{t-1}) = 0$. The object of interest is the $(n \times n)$ positive-definite variance-covariance matrix of returns, $Var(\mathbf{r}_t|\mathcal{F}_{t-1}) = \mathbf{\Sigma}_t$, which we assume to be time-varying, predictable and, although unobserved, consistently estimated by a realized variance-covariance matrix $\mathbf{V}_t$. Generally in this paper $\mathbf{\Sigma}$ represents the actual VCM, $\mathbf{V}$ is an observed realized value of the VCM, calculated from intraday data, and $\mathbf{H}$ is used to denote a forecast of the matrix.

3 Literature Review

In this paper we use nonparametric econometrics to produce forecasts of a variance-covariance matrix. This approach has previously been used when forecasting univariate volatility in Clements *et al.* (2010), in which forecasts are a weighted average of historical values of realized volatility. The weights are increased when the pattern of historical volatility behaviour is similar to that around the time at which the forecast is made. Clements *et al.* (2010) show that at a 1 day forecast horizon such an approach performs well against competing volatility forecasting techniques.

What distinguishes this approach to many other VCM forecasting models is the role played by the time between a past VCM observation and the period in which a forecast is made. In general the weight given to past observations decreases for observations further in the past. An important example of this is RiskMetrics, described in J.P. Morgan (1996), a popular method of forecasting the VCM using an exponentially weighted moving average (EWMA) approach. Gijbels, Pope and Wand (1999) show that this approach can be interpreted as a kernel approach in which weights on historical observations are determined by the lag at which a realization was observed. Fleming, Kirby and Ostdiek (2003) also used an EWMA weighting scheme, similar to that used in

Riskmetrics, to show that weighted averages of realized covariance matrices can improve forecasts of volatility, based on economic performance, compared to methods using daily returns.

The performance of the model proposed in this paper is compared to currently available multivariate forecasting models, which have proliferated in number in recent years. The most popular multivariate model is the dynamic conditional correlation model (DCC) of Engle (2002) and Engle and Sheppard (2001), which allows correlations between stocks to vary according to a GARCH type process. Other interesting recent models include the regime switching dynamic correlation model (RSDC) of Pelletier (2006) which assumes that correlations/covariances change depending on the state of the world. Colactio, Engle and Ghysels (2007) introduce a model which models long run correlation behaviour using a MIDAS approach and short term behaviour via the DCC, however such an approach requires a complex system of restrictions in order to ensure positive definiteness of the VCM matrix. Engle and Kelly (2008) introduce the dynamic equicorrelation model which assumes that correlations between all stocks have the same value, but that this changes through time. All of these models share the fact that they must make restrictions, either on paramater values or in their setup, to ensure that estimates and forecasts obtained from them are positve-definite and symmetric.

An interesting new approach to VCM modelling is that of Chiriac and Voev (2009) who introduce a VARFIMA model which models the behaviour of elements of the Cholesky decomposition of the VCM, however this suffers from the problem that interpretation of the elements of the decomposition is difficult and incorporating additional variables is far from straightforward.

In this paper we introduce a nonparametric technique for obtaining VCM forecasts using a multivariate kernel approach that encompasses the Riskmetrics method as a special case. This extends the univariate approach of Clements *et al.* (2010) to the multivariate context. Importantly we show how variables, other but time delay, can be used in a multiplicative kernel weighting.

Our contribution heavily depends on contributions made in nonparametric, kernel estimation literature. It is well known that when estimating densities or conditional expectations by means of kernel methods the properties of the resulting estimators will depend on the choice of kernels and, more importantly the choice of bandwidth used in the kernel estimators. Plug-in bandwith rules have been proposed but it was also recognised that these may be inadequate when data did not meet strict assumptions (Silverman, 1986, Bowman, 1997). Here we will use cross-validation methods (Bowman, 1984, and Rudemo, 1982) to find optimal bandwidths.

Application of cross-validation will also facilitate the selection of relevant variables to be used in the kernel weighting algorithm. One class of variables considered for the kernel weighting algorithm are scalar transformations of matrices as they can be used to establish the *closeness* of matrices. The idea is to give higher weight to past observations that relate to times when the VCM was similar to the current VCM (regardless of how distant that observation is). Moskowitz (2003) proposes three statistics to evaluate the closeness of VCMs. The first metric compares the matrices' eigenvalues, the second looks at the relative differences between the individual matrix elements and the third considers how many of the correlations have the same sign in the matrices. Taken together these three metrics can be used to determine the level of similarity between two VCMs. Other functions used to compare matrices, often called loss functions, have been discussed in the literature (Laurent, Rombouts and Violante, 2009). One such loss function is the Stein distance, also known as the MVQLIKE function. This loss function is shown to perform well in discriminating between VCM forecasts in Clements, Doolan, Hurn and Becker (2009) and Laurent, Rombouts and Violante (2010) and represents another useful tool for comparing VCMs.

The second class of variables used in the kernel weighting algorithm are variables carrying information on the state of the economy. The basic idea is to give past VCM larger weights in the forecast if the macroeconomic conditions are similar to those prevalent at the time of the forecast formation. Aït-Sahalia and Brandt (2001) investigate factors influencing stock volatility and propose dividend yield, default spreads and term spreads as factors. This builds on existing work which identifies term spreads and default spreads as potential drivers of stock volatility processes. Campbell (1987), Fama and French (1989) and Harvey (1991) investigate the relationship between term spreads and volatility while Fama and French (1989), Whitelaw (1994) and Schwert (1989) consider a volatiliy-default spread relationship. In addition Harvey (1989) considers the impact of default spreads on covariances. Hence there is an established literature relating these variables to the behaviour of elements of a VCM.

Empirical evidence in Schwert (1989), Hamilton and Lin (1996) and Campbell, Lettau, Malkiel & Xu (2001) suggests that during market downturns/recessions stock return volatility can be expected to increase. Based on these findings we propose to use an algorithm, such as that detailed in Pagan and Sossounov (2003), to identify periods in which the stock market is upbeat, as VCMs in such periods may have common characteristics. Commodity prices, such as gold (Sjaastad and Scacciavillani, 1996) and oil (Sadorsky, 1999, and Hamilton, 1996) prices, have also been linked to stock market volatility and are therefore considered here as potential variables to contribute to the kernel weighting functions.

The final variable used in this paper is implied volatility, namely the VIX index of the Chicago Board of Exchange. This is often interpreted as a market's view on future stock market volatility. This measure has been used in the context of univariate volatility forecasting (Poon and Granger, 2003, Blair, Poon and Taylor, 2001) and is here considered as another variable in the multivariate kernel weighting scheme.

4 Riskmetrics as a Kernel Approach

In this section we restate a result by Gijbels, Pope and Wand (1999) that establishes that a Riskmetrics type, exponential smoothing forecast can be represented as a univariate kernel forecast in which weights vary with time. This will provide a special case of the more general methodology introduced in Section 5 in which we introduce a multivariate kernel which potentially utilises the variables listed in the previous Section.

In a multivariate setting, the variance-covariance matrix forecast $\mathbf{H}_{T+1}$ at time T, given by the standard Riskmetrics equation is

$$\mathbf{H}_{T+1} = \lambda \mathbf{H}_T + (1-\lambda)\mathbf{r}_T \mathbf{r}_T' \tag{1}$$

when observations are equally spaced in time and λ is a smoothing paramater, $0 < \lambda < 1$, commonly set at a value recommended in J.P. Morgan (1996). From recursive substitution and with $\mathbf{H}_1 = \mathbf{r}_1' \mathbf{r}_1$, the forecast of the VCM can be expressed as

$$\mathbf{H}_{T+1} = (1-\lambda) \sum_{j=0}^{T-1} \lambda^j \mathbf{r}_{T-j} \mathbf{r}_{T-j}' \tag{2}$$

The sum of the weights is thus equal to $1-\lambda^T$ and as noted in Gijbels, Pope and Wand (1999) this approaches 1 as we allow T to approach infinity. However in order to normalise the sum of the weights to be exactly 1 we restate the Riskmetrics model as

$$\mathbf{H}_{T+1} = \frac{\sum_{j=0}^{T-1} \lambda^j \mathbf{r}_{T-j} \mathbf{r}_{T-j}'}{\sum_{j=0}^{T-1} \lambda^j} \tag{3}$$

We can now reformulate (3) as a kernel[1], defining $h = 1/\log(\lambda)$ and $K(u) = \exp(u)\mathbf{1}_{u \leq 0}$, we

[1] More accurately this is a half kernel as it is zero for $T+1, T+2, ...etc.$.

can restate (3) as

$$\mathbf{H}_{T+1} = \frac{\sum_{j=1}^{T} K\left(\frac{t-T}{h}\right)\mathbf{r}_t\mathbf{r}_t'}{\sum_{j=1}^{T} K\left(\frac{t-T}{h}\right)} = \sum_{j=1}^{T} W_{rm,t}\ \mathbf{V}_{rm,t} \tag{4}$$

From this we replicate the conclusion of Gijbels, Pope and Wand (1999) that Riskmetrics is a zero degree local polynomial kernel estimate with bandwidth h. From a practial point of view the Riskmetrics kernel determines weights, $W_{rm,t} = K\left(\frac{t-T}{h}\right) / \left(\sum_{j=1}^{T} K\left(\frac{t-T}{h}\right)\right)$, based on how close observations of $\mathbf{V}_{rm,t} = \mathbf{r}_t\mathbf{r}_t'$ are to time T, the period at which a forecast is being made. The largest weight is attached to the observation at time T and the weights decrease according to an exponentially weighted smoothing average pattern. In the remainder of this paper we aim to expand such an approach by including factors other than time in our estimation of kernel weights.

5 Multivariate Kernel Methodology

In this section we present the method by which we obtain the kernel and subsequent forecasts of the VCM. The inputs to our model are a set of p variables, which we believe to contain information relevant to forecasting the VCM, and a time series of realized variance covariance matrices. Calculation of the $n \times n$ realized variance-covariance matrix, $\mathbf{V}_t$, is a non-trivial issue. Here we compute it using standard methods from the realized (co)variance literature and we assume that $\mathbf{V}_t$ is positive definite. The method used to calculate the matrices used in the rest of this paper is described in section 7.2.

At time T we wish to obtain a forecast of the d-step ahead VCM, which is the matrix describing variances and covariances over the time period $T+1$ to $T+d$, denoted by $\mathbf{H}_{T+d}^{(d)}$. We obtain our forecast by taking a weighted combination of historical VCMs, hence

$$\mathbf{H}_{T+d}^{(d)} = \sum_{t=1}^{T-d} W_t \mathbf{V}_{t+d}^{(d)}. \tag{5}$$

As our forecast is a weighted combination of symmetric, positive definite matrices, $\mathbf{H}_{T+d}$ also has these properties and so is a valid covariance matrix. Ensuring that forecasts of the variance covariance matrix are positive definite is rarely so straightforward and models usually have to employ parameter restrictions or decompositions of $\mathbf{V}_t$ in order to ensure this.

The focus of much of the remainder of this section is the method by which we determine the optimal weights, to use in (5). In order to ensure that the weights sum to one we impose the following normalisation:

$$W_t = \frac{\omega_t}{\sum_{i=1}^{T-d} \omega_i}. \tag{6}$$

This allows Equation (5) to be interpreted as a weighted average, ensuring an appropriate scaling for $\mathbf{H}_{T+d}^{(d)}$.

We now explain how to determine ω_t using kernel estimation techniques. The idea underpinning the approach is to determine which of the past time periods had conditions most similar to those at the time we make the forecast, T. We then place more weight on the VCMs that occured over the d periods following the dates that were most similar to time T.

We determine the similarity of other time periods to time T using p variables and employ a multivariate kernel to calculate the raw weight applicable to day t, hence

$$\omega_t = \prod_{j=1}^{p} K_j(\Phi_{t,j}, \Phi_{T,,j}, h_j) \tag{7}$$

where $\Phi_{T,,j}$ is the element from the T^{th} row and j^{th} column of the data matrix Φ which has dimensions $T \times p$ and h_j is the bandwidth for the j^{th} variable.

For continuous variables $K_j(\Phi_{t,j}, \Phi_{T,,j}, h_j)$ is the standard normal density kernel[2] (Silverman 1986 and Bowman 1997) defined as

$$K_j(\Phi_{t,j}, \Phi_{T,,j}, h_j) = (2\pi)^{-0.5} \exp\left[-\frac{1}{2}\left(\frac{\Phi_{T,,j} - \Phi_{t,j}}{h_j}\right)^2\right].$$

In the case of a discrete dummy variable, such as a bull/bear market dummy we use the discrete univariate kernel proposed in Aitchison and Aitken (1976). The form of the kernel is

$$K_j(\Phi_{t,j}, \Phi_{T,,j}, h_j) = \begin{cases} 1 - h_j & \text{if } \Phi_{t,j} = \Phi_{T,,j} \\ h_j/(s_j - 1) & \text{if } \Phi_{t,j} \neq \Phi_{T,,j} \end{cases} \tag{8}$$

where s_j is the number of possible values the discrete variable can take ($s_j = 2$ in the case of the bull/bear market variable). In the two state discrete case $h_j \in [0, 0.5]$. If $h_j = 0.5$ the value of the discrete variable has no impact on the forecast, while if $h_j = 0$ we disregard data points which do not share the same discrete variable value as $\Phi_{T,j}$.

[2] We normalise continuous variables before applying the kernel function.

In addition to the discrete and continuous kernels we use a third approach when we include time as one of the p variables. In that case

$$K_j(\Phi_{t,j}, \Phi_{T,j}, h_j) = \frac{h_j^{T-t}}{\sum_{q=1}^{T-h} h_j^{T-q}} \tag{9}$$

which has the same structure as the Riskmetrics approach in Equation (3). However, here we allow a flexible bandwidth, $h_j \in [0, 1]$, as opposed to a prespecified value as in J.P. Morgan (1996).

As we are using a multiplicative kernel the time kernel suggested in Equation (9) is problematic as $K_j(\cdot)$, with increasing $(T-t)$, will quickly decline towards a value of zero which will make the value of ω_t in (7) approach zero. This implies that in effect observations with sufficiently large $(T-t)$ will be ignored regardless of how similar the macroeconomic and VCM characteristics are to the point of forecast. We therefore propose an alternative scheme where

$$K_j(\Phi_{t,j}, \Phi_{T,j}, h_j) = \frac{h_j^{T-t}}{\sum_{q=1}^{T-h} h_j^{T-q}} + 1. \tag{10}$$

The minor alteration ensures that the weights decline to a value of one. There is still an increased weight on more recent observations however less recent time periods are not ignored because of the effects of the time kernel. We present results using both versions of the time kernel in order to demonstrate the impact of such an approach.

While the general approach presented through Equations (5), (6) and (7) has the Riskmetrics approach as a special case[3], it introduces a significant amount of additional flexibility, by allowing the weights W_t to be determined from a set of p variables.

5.1 Choice of Bandwidth

The choice of bandwidth is a non-trivial issue in nonparametric econometrics, however a common rule of thumb quoted for multivariate density estimation is

$$h_j = \left\{ \frac{4}{(p+2)\,T} \right\}^{\frac{1}{p+4}} \sigma_j$$

where σ_j is the standard deviation of the j^{th} variable. Although this rule of thumb provides a simple method for choosing bandwidths, as noted in Wand and Jones (1995) these bandwidths may be sub-optimal.

[3] Using $\mathbf{V}_t = \mathbf{r}_t \mathbf{r}_t'$ rather than a realized VCM.

Importantly, if one was to optimise (using cross-validation) the bandwidth parameters the optimised values, h_j, will contain information on whether the jth element in Φ does contribute significant information to the optimal weights W_t. As noted in Li and Racine (2007, pp. 140-141), irrelevant (continuous) variables are associated with $h_j = \infty$. For binary variables (and kernel as in Equation (8)) and a time variable (and a kernel as in Equation (10)) the bandwidths $h_j = 0.5$ and $h_j = 1$ respectively represent irrelevant variables.

Cross-validation is a bandwidth optimisation strategy introduced in Rudemo (1982) and Bowman (1984). It selects bandwidths to minimise the mean integrated squared error (MISE) of density estimates and is generally recommended as the method of choice in the context of nonparmetric density and regression analysis (Wand and Jones, 1995, Li and Racine, 2007). As we are interested in forecast performance rather than density estimation, we obtain bandwidths which minimise the MVQLIKE of our forecasts rather than the MISE.

MVQLIKE is a robust loss function for the comparison of matrices, where $\mathbf{H}_t^{(1)} = \mathbf{H}_t$ is the $(n \times n)$ dimensional 1 period ahead forecast of the VCM at time t and $\mathbf{V}_t^{(1)} = \mathbf{V}_t$ is the realized VCM at time t[4]. The loss function is calculated as

$$MVQLIKE(\mathbf{H}_t) = tr(\mathbf{H}_t^{-1}\mathbf{V}_t) - \log\left|\mathbf{H}_t^{-1}\mathbf{V}_t\right| - n. \tag{11}$$

This is the criterion function to be minimised in our cross-validation approach. Consider that we have data available up to and including time period T and we aim to forecast the VCM for $T+1$. The available data over time periods 1 to T can be used to identify the optimal bandwidths for use in forecasting. This is done by evaluating $K\,(< T)$ forecasts for periods $T-K+1$ to T. The initial $T-K$ observations[5] are used to produce the first forecast $\mathbf{H}_{T-K+1}$. For any period τ, $T-K+1 \leq \tau \leq T$, the forecast $\mathbf{H}_\tau$ is based on observations of variables in Φ available at time $\tau - 1$.

Having obtained these forecasts we select bandwidths to minimise the mean of MVQLIKE over these in-sample forecasts

$$CVMVQ\,(h) = \frac{1}{K}\sum_{\tau=T\_K+1}^{T} MVQLIKE(\mathbf{H}_\tau\,(h)) \tag{12}$$

[4]The following argument is, for notational ease, made for 1 period ahead forecasts but the extension to d period forecasts is straightforward. Initially we shall also suppress the dependence of $\mathbf{H}_t$ on h, the $(p \times 1)$ vector of bandwidths.

[5]We set $T-K = 300$, which means that every forecast is based on a minimum of 300 observations.

where dependence on h, the $(p \times 1)$ vector of bandwidths is now made explicit. The bandwidths that minimise (12) are then used in Equations (5), (6) and (7) in order to forecast $\mathbf{H}_{T+1}$.

The optimised bandwidth values should carry information on which of the p variables included in Φ contribute significantly to the determiniation of the weights in Equation (5). Li and Racine (2007) suggest that a cross-validation approach in the context of a multivariate kernel regression should, asymptotically, deliver bandwidth estimates that approach their "irrelevant" values discussed above ($h_j = \infty$, $h_j = 0.5$ and $h_j = 1$ respectively for continuous, binary and time variables). They suggest that, therefore, there is no need to eliminate irrelevant variables.

When following this strategy we encountered significant difficulties in the optimisation process, in particular our nonlinear bandwidth optimisation is unable to identify an optimum. We therefore recommend an alternative strategy which eliminates irrelevant variables and identifies optimal bandwidths only for the remaining variables. The elimination of variables is achieved as follows. Each variable is used as the only variable in Φ to determine kernel weights. We find the optimal bandwith, $\widetilde{h}_j$, for each variable by minimising the criterion in (12). The optimal $CVMVQ\left(\widetilde{h}_j\right)$ is then compared to the $CVMVQ_R$ which is obtained by forming VCM forecasts from averaging all available past VCMs. The rationale is that a relevant variable should deliver improvements compared to a simple average. This is illustrated in Figure 1 for a forecast horizon of $d = 1$[6]. The dashed lines indicate $CVMVQ_R$ and the solid lines represent $CVMVQ\left(h_j\right)$. The minima on the latter illustrate the bandwidth $\widetilde{h}_j$ that minimises $CVMVQ\left(h_j\right)$. Weighting variables that do not improve on the $CVMVQ_R$ by at least 1% are then eliminated.

In order to obtain a handle on the size of this threshold we simulated 1000 random variables which were subsequently considered as potential weighting variables (and there $CVMVQ\left(\widetilde{h}_{rv}\right)$) calculated. As it turns out a threshold of 1% would eliminate virtually all of these irrelevant random variables[7]. Despite this the threshold is essentially ad-hoc and it is envisaged that future reserach may improve on this aspect of the proposed methodology.

[6] Similar illustrations for 5 and 22 day forecast horizons offer no additional insight and so are not presented here.

[7] We also applied a more conservative threshold of 2% but results remained virtually unchanged and are therefore not reported.

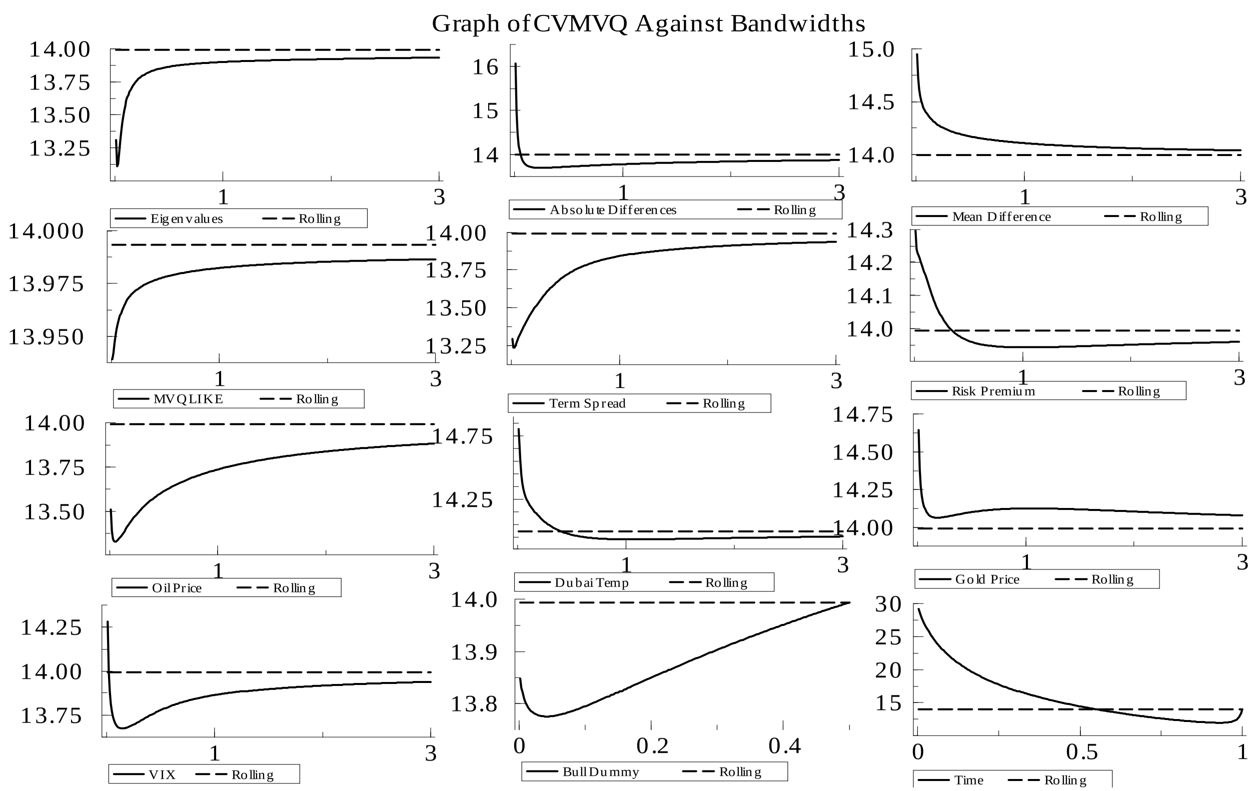

Figure 1: Graphs of 1 day ahead CVMVQ against bandwidth values for 12 variables. The dashed line represents the CVMVQ from a rolling average forecastor. Describtions of the variables used are provided in Section 6.

In short the process of variable elimination and bandwidth optimisation can be summarised in the following three step procedure:

1. For each of the p variables considered for inclusion in the multivariate kernel, apply cross validation to obtain the optimal bandwidth when only that variable is included in the kernel estimator. We refer to these as univariate optimised bandwidths $\tilde{h}_j$ $j = 1, ..., p$.

2. Compare the forecasting performance of the univariate optimised bandwidths from Step 1, $CVMVQ\left(\widetilde{h}_j\right)$, against $CVMVQ_R$ from a simple average forecasting model. Any of the p variables that fails to improve on the rolling average forecast performance by at least 1% is eliminated at this stage as it is considered to have little value for forecasting. We are left with $p^* \leq p$ variables used as weighting variables.

3. Estimate the multivariate optimised bandwidths h_j^* for the p^* variables that are not eliminated in Step 2 by minimising the cross validation criterion in Equation (12). As opposed to Step 1 this optimisation is done simultaneously over all p^* bandwidths.

Having obtained the optimised bandwidths from Step 3, we then forecast the VCM for the d day-ahead time period ending at $T + d$ using (7).

6 Potential Variables

The approach outlined in the previous Section illustrates how the p^* relevant variables are to be identified from a list of p variables initially considered to be potentially relevant. The p^* selected variables then contribute to the calculation of the weights used in Equation (5). Here we describe the set of p variables from which we select p^* variables considered to be relevant. The variables used can be classified in three categories. First, the time variable, as it is used in the Riskmetrics approach. This assumes that that VCM observations close to the time period T at which the forecast is made, are more relevant than observations further back in time. The second class of potential weighting variables are measures of matrix closeness. In essence, the more similar a VCM matrix at any time $t < T$ is to the VCM at time T, the larger should be the weight given to the associated observed VCM over the subsequent d days in Equation (5). These measures of matrix closeness are discussed in Section 6.1. Finally we consider variables that can broadly be categorised as variables describing the prevelant economic circumstances at time t. Larger weight is to be given to VCM if the associated macroeconomic conditions are similar to those prevelant at the time of forecast formation, T. Variables that fall into this category are described in Section 6.2.

6.1 VCM Comparison Variables

Moskowitz (2003) discusses a number of summary statistics that measure the difference between two matrices. We consider three of these statistics here. The first is the ratio of the eigenvalues of the (squared) VCM at time t relative to those of the (squared) VCM at time T:

$$\frac{\sqrt{trace\,(\mathbf{V}_t'\mathbf{V}_t)}}{\sqrt{trace\,(\mathbf{V}_T'\mathbf{V}_T)}} \tag{13}$$

Values close to 1 indicate that matrices are similar to each other.

The second statistic adopted from Moskowitz (2003) evaluates the absolute elementwise differences between two matrices $\mathbf{V}_t$ and $\mathbf{V}_T$. The sum of all absolute differences is standardised by the sum of all elements in $\mathbf{V}_T$. The statistic is defined as

$$\frac{\iota'|\mathbf{V}_T - \mathbf{V}_t|\iota}{\iota'\mathbf{V}_T\iota} \tag{14}$$

where ι is a ι is a $n \times 1$ vector of ones. For identical matrices this statistic will take a value of 0.

A third metric suggested in Moskowitz (2003) is based on the realized correlation matrices $\mathbf{C}_t$ and $\mathbf{C}_T$[8]. The statistic compares how similar $\mathbf{C}_t$ and $\mathbf{C}_T$ are in relation to the average realized correlation matrix $\bar{\mathbf{C}}$. Specifically we are concerned about the relative position of a particular correlation relative to its long-run average. $sign(vech(\mathbf{C}_t - \bar{\mathbf{C}})_i)$ delivers a positive (negative) sign if the realized correlation (of the ith unique element) at time t is larger (smaller) than the relevant average correlation. The statistic considered here essentially calculates the proportion of the m unique elements in $\mathbf{C}_t$ that have identical deviations from the long-run correlations as those in $\mathbf{C}_T$:

$$\frac{1}{m}\sum_{i=1}^{m} I\left\{sign(vech(\mathbf{C}_t - \bar{\mathbf{C}})_i){=}sign\left(vech(\mathbf{C}_T - \bar{\mathbf{C}})_i\right)\right\}. \tag{15}$$

$I\{\}$ is an indicator taking the value of 1 when the statement inside the brackets is true and 0 otherwise and $m = \frac{n}{2}(n-1)$ is the number of unique correlations in the $n \times n$ correlation matrix. If matrices are identical with respect to this measure this statistic will take the value 1.

[8] The realized correlation matrices are calculated from $\mathbf{C}_t = \mathbf{D}_t^{-1}\mathbf{V}_t\mathbf{D}_t^{-1}$ where $\mathbf{D}_t$ is a $(n \times n)$ diagonal matrix with $\sqrt{V_{iit}}$ on the ith diagonal element and V_{iit} is the (i,i) element of $\mathbf{V}_t$.

We also compare VCM matrices using the MVQLIKE loss function (Laurent, Rombouts and Violante, 2009) due to it being a robust multivariate loss function, where it is defined as

$$tr\left(\mathbf{V}_t^{-1}\mathbf{V}_T\right) - \log\left|\mathbf{V}_t^{-1}\mathbf{V}_T\right| - n \tag{16}$$

such that matrices which are identical will deliver a statistic of value 0.

These four statistics will determine the level of similarity between the VCMs at time t and time T. The variable selection and bandwidth estimation strategy described previously will determine which of these variables are relevant for VCM forecasting.

6.2 Economic Variables

The variables introduced in this section attempt to identify economic variables that are potentially significant for VCM behaviour, based on findings in the existing literature. The first variable we introduce to our model is the term spread as it was used in Aït-Sahalia and Brandt (2001), Campbell (1987), Fama and French (1989) and Harvey (1991) when investigating the time varying volatility of asset returns. This variable, as defined in Aït-Sahalia and Brandt (2001) is the difference between 1 and 10 year US government bond yields and so if we define the yield on an x year US government bond at time t as $Y_{Gx,t}$, the term spread variable is

$$Y_{G10,t} - Y_{G1,t}. \tag{17}$$

Aït-Sahalia and Brandt (2001), Fama and French (1989), Whitelaw (1994) and Schwert (1989) investigated the relation between return volatility and default spread. The default spread measures the difference in yield between Moody's Aaa and Baa rated corporate bonds. Hence defining the yields on Aaa and Baa rated bonds at time t as $Y_{Aaa,t}$ and $Y_{Baa,t}$ respectively the default spread variable is

$$Y_{Baa,t} - Y_{Aaa,t}. \tag{18}$$

Both oil prices (Oil_t) and gold prices ($Gold_t$) have been shown to influence stock return volatility (Sjaastad and Scacciavillani, 1996, Sadorsky, 1999, and Hamilton, 1996), based on this we include both prices in our investigation.

Schwert (1989), Hamilton and Lin (1996) and Campbell *et al.* (2001) demonstrate that volatility increases during economic downturns. We therefore include a dummy variable identifying bull and bear market periods as described in Pagan and Sossounov (2003)[9]. When applying this to data we use only the information available up until time T in determing turning points between states of the market. We define the variable $Bull_t$ as having a value of one when the market is bullish and 0 otherwise.

As we are interested in the volatility of a stock portfolio it may also be useful to include a market measure of volatility in our list of potential variables. In order to do this we use the volatility index (VIX) quoted by the Chicago Board of Exchanges, VIX_t. This provides a measure of volatility implied by market prices and so may be useful as a guide to the level of volatility expected by the market.

In addition to these variables we include two variables which we expect to be irrelevant for the purpose of VCM forecasting. The spurious variables we use are the temperature in Dubai[10] ($DUBAITEMP_t$) and a random variable generated using a standard normal distribution random number generator ($RANDOM_t$). In the absence of a sensible simulation strategy that can evaluate the "size" or "power" of our approach, these variables are included as a sensibility check for the results produced. Any sensible methodology should eliminate such variables. As it turns out, the proposed methodology does indeed eliminate these two irrelevant variables at all forecasting horizons.

7 Forecasting Competition

The empirical application presented in this paper is designed to answer the following two questions. *First*, does the forecasting approach introduced in Section 5 compare favourably to more established forecasting techniques for high dimensional VCMs? *Second*, and more specifically, do the economic indicators discussed in Section 6.2 add valuable information into the process of VCM forecasting?

While one could think of a sensible Monte-Carlo setup to establish the answer to the first question, this seems an impossible task with respect to the second. One would have to devise a large multivariate system that jointly modelled stock returns and macroeconomic variables. It is likely that any results would be highly specific to the system devised and therefore this paper

[9] The algorithm identifies bull and bear periods based on monthly data, daily data is often too noisy to support identification of broad trends. As a result once the algorithm identifies a month as belonging to a bull/bear period all of the constituent days are also assumed to belong to this period.

[10] Temperature data was obtained from the University of Daytons daily temperature archive. See http://www.engr.udayton.edu/weather/.

is restricted to an empirical analysis of the questions posed above. In order to obtain sufficient information to address the two issues raised we will apply the Multivariate Kernel approach to two sets of potential weighting variables. In one set of forecasts (MVK) the variable elimination and bandwidth optimisation strategy described in Section 5.1 is applied to the entire set of potential weighting variables. In a second set of forecasts ($MVKne$) the weighting variables are restricted to come from a set that includes the time variable and the variables describing matrix similarities from Section 6.1. If the latter set, that includes no economic variables, does significantly worse than the first, we will consider this evidence that economic variables contain useful information in the context of VCM forecasting.

In Section 7.1 alternative forecasting models are introduced and Section 7.2 discusses data estimation setup issues. The model confidence set methodology used to establish the statistical significance of our results, is reviewed in Section 7.3. Results are presented in Section 8.

7.1 Models Included In the Forecast Comparisons

In addition to the forecasting model we propose above we include the Dynamic Conditional Correlation (DCC) model of Engle and Sheppard (2001), and versions of the RiskMetrics method of J.P. Morgan (1996) in our MCS evaluations of forecast accuracy. Here we provide a brief summary of the models we use.

The DCC model is perhaps the most popular of recent models focused on the VCM of stock returns. In general it models the variances of individual stocks using a GARCH process and then apples a similar process to the correlation matrices. Hence, assuming returns are non forecastable, the vector of returns $\mathbf{r}_t$ is distributed as

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\
\boldsymbol{\epsilon}_t &= \mathbf{H}_t^{1/2}\mathbf{z}_t \\
\mathbf{z}_t &\sim IID(\mathbf{0}_n, \mathbf{I}_n)
\end{aligned}$$

Where $\mathbf{0}_n$ is a $(n \times 1)$ vector of zeroes and $\mathbf{I}_n$ is a $(n \times n)$ identity matrix and $\mathbf{H}_t$ is the $(n \times n)$ variance covariance matrix. Each of the variances, which form the diagonal of $\mathbf{H}_t$ are then modelled using a GARCH(1,1) process and the DCC models the underlying correlation matrix. Define $\mathbf{D}_t$

as $(n \times n)$ diagonal matrix with the $(n \times 1)$ vector of standard deviations of returns at time t on the diagonal. We can restate $\mathbf{H}_t$ as in (19)

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t. \tag{19}$$

Here $\mathbf{R}_t$ is the $(n \times n)$ correlation matrix describing how stocks move together on day t with 1's on the diagonal and off diagonals between -1 and 1. A further transformation is performed on this matrix in order to ensure these properties, the transformation is

$$\begin{aligned} \mathbf{R}_t &= \mathbf{Q}_t^* \mathbf{Q}_t \mathbf{Q}_t^* \\ \mathbf{Q}_t &= (1-\alpha-\beta)\bar{\mathbf{Q}} + \alpha \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}^{'} + \beta \mathbf{Q}_{t-1} \\ \mathbf{Q}_t^* &= diag(\mathbf{Q}_t)^{1/2} \end{aligned} \tag{20}$$

where $\bar{\mathbf{Q}}$ is the long run correlation matrix and α and β are parameters determining how the correlations move through time. In order to ensure that correlations are stationary the restrictions $a, \beta > 0$ and $\alpha + \beta < 1$ are imposed.

In order to obtain muliple horizon forecasts from the DCC we utilise one of a pair of suggestions from Engle and Sheppard (2001) which yields the h step forecast equation

$$\begin{aligned} \hat{\mathbf{H}}_{t+h} &= \hat{\mathbf{D}}_{t+h} \hat{\mathbf{R}}_{t+h} \hat{\mathbf{D}}_{t+h} \\ \hat{\mathbf{R}}_{t+h} &= \hat{\mathbf{Q}}_{t+h}^* \hat{\mathbf{Q}}_{t+h} \hat{\mathbf{Q}}_{t+h}^* \\ \hat{\mathbf{Q}}_{t+h} &= \sum_{i=0}^{h-2} (1-\alpha-\beta)\bar{\mathbf{Q}}(\alpha+\beta)^i + (\alpha+\beta)^{h-1} \hat{\mathbf{Q}}_{t+1} \end{aligned}$$

where the elements of $\hat{\mathbf{D}}_{t+h}$ are forecast from the models used for univariate volatilities.

The DCC model can be difficult to estimate as n increases. In the context of this paper $(n = 20)$ we utilise the composite likelihood approach of Engle, Shephard and Sheppard (2008) to obtain estimates of a and β (see Appendix B for details).

Another popular method for forecasting the VCM of stock returns is the Riskmetrics forecasting model (J.P. Morgan, 1996) as describen in Section 4. The weighting paramater λ in (2) determines the weighting scheme and it is recommended in in J.P. Morgan (1996) that this be set at a value of 0.94 for daily data and 0.97 for monthly data. These values are used for 1 and 22 day forecast comparisons respectively. There is no guidance for what λ should be set to when using weekly data and so we follow Laws and Thompson (2005) in setting $\lambda = 0.95$ when obtaining forecasts for a five day ahead period. Forecasts from this model will be labeled *RM*.

The recommended values are the result of averaging optimal λs over several different economic time-series models, not all of which will be representative for the data at hand. We therefore expect an optimised value of λ to outperform the fixed recommendation. As well as adopting the above recommendations for λ we, therefore, also include a version of Riskmetrics for which we optimise λ, choosing a value of λ that minimses $CVMVQ$ for in sample estimates of the VCMs (*RMopt*).

We introduce one further adjustment to the Riskmetrics methodology. The new information entering the Riskmetrics forecast at time T is the cross product of daily returns $\mathbf{r}_T\mathbf{r}_T'$ (see Equation 1). This can be interpreted as a very noisy proxy for the variance covariance structure at day T. It is well known that a less noisy proxy is the realized VCM $\mathbf{V}_T$, and hence we propose (similar to Fleming, Kirby and Ostdiek, 2003) the following forecasting model[11]

$$\mathbf{H}_{T+1} = \lambda\mathbf{H}_T + (1-\lambda)\mathbf{V}_T. \tag{21}$$

As the series of $\mathbf{V}_T$ has different properties compared to $\mathbf{r}_T\mathbf{r}_T'$ and it is, therefore, apparent that the fixed values for λ recommended for the latter should not be applied here. We use the cross validation approach proposed above to find optimal values for the weighting parameter. In what follows forecasts from this model are labeled *RMvcm*.

7.2 Data and Setup

The VCM forecasts included in our analysis all relate to a portfolio of 20 stocks listed on the NYSE over the period 28/11/1997-31/8/2006. A full list of the stocks used can be found in Appendix A at the end of this paper. The macroeconomic information used covers the same period. the information on term spreads (GVUS05(CM10)~U\$, GVUS05(CM01)~U\$), default spreads (DAAA, DBAA) , oil prices OILBREN) and gold prices (GOLBLN) were obtained from Datastream[12]. The bull and bear dummy variables were calculated using the algorithm suggested in Pagan and Sossounov (2003) based on monthly S&P 500 index prices. The algorithm was adjusted so that the values of the dummies were that which would have been calculated using the data available at the point in time at which we make our forecast. The VIX data was obtained

[11] Illustrated for a $d = 1$ day forecating horizon. The generalisation is straightforward.
[12] The information in brackets are the datastream codes/names for the data series used to construct these variables.

from the Chicago Board of Exchange (CBOE) website[13].

In versions of the model in which we make multiple day forecasts we take averages of the data over periods of the same length, with the exception of the bull and bear day variable for which we take the value on the first day of the period.

The non-parametric approach described in this paper utilise realized variance covariance matrices compiled from intra-day price quotes. In order to compile our realized VCMs we use the following method. We obtain vectors of returns over the period between the market closing on day $t-1$ and opening on day t, we denote these as $\mathbf{r}_{Ot}$. We also obtain vectors of returns over every 5 minute period during the time the market is open[14], hence as the stocks are traded over the period 9:30-16:00 we obtain 55 intraday return vectors $\mathbf{r}_{it}$ $i = 1, ..., 55$. In order to calculate a VCM for an entire 24 hour time period we use one of the methods for such calculations proposed in Hansen and Lunde (2005). We calculate the realized variance-covariance matrix for day t as

$$\mathbf{V}_t = \mathbf{r}_{Ot}\mathbf{r}_{Ot}^{'} + \sum_{i=1}^{55} \mathbf{r}_{it}\mathbf{r}_{it}^{'}. \tag{22}$$

As the close to open returns on a stock represent a significant part of the risk of holding stocks it seems appropriate to include these in our forecasting approach. When forecasting over multiple days ($d = 5, 22$) we require $\mathbf{V}_{t+d}^{(d)}$ which can be obtained from $\mathbf{V}_{t+d}^{(d)} = \sum_{\tau=t+1}^{t+d} \mathbf{V}_{\tau}$.

The initial estimation period for all time horizons in the forecast competition results below consists of the first 936 datapoints. All forecasting periods are non-overlapping, leading to 1,266, 253 and 57 forecast periods for the 1, 5 and 22 day forecast horizons respectively. As our model makes use of instances in the past when conditions are similar to the forecast point it seems logical to allow the model access to as much data as possible and so we allow expanding estimation samples to be used in the compilation of forecasts. In order to ensure that the DCC is not hampered by data restrictions we also employ the expanding dataset in the estimation of the DCC paramaters.

The cross-validation procedure for eliminating variables that do not contribute to improved VCM forecasts is very computing intensive and therefore is performed after every 200 days, seven times throughout our sample period. On each of these seven occasions a variable is either included in the model or not. Significant variables are then retained for the following 200 days. On each

[13] See http:\\www.cboe.com/micro/vix/historical.aspx.

[14] As all 20 stocks are very liquid and frequently traded we do not anticipate any microstructure or non-synchronicity issues at a 5 minute sampling interval.

day, however, a new multivariate bandwidth optimisation (as described in Section 5.1), for the fixed set of retained variables, is performed.

7.3 Analysis of Results Model Confidence Sets (MCS)

In our forecast competition we want to determine which of the six models provide the best forecasts. In order to do this we use the MCS, introduced in Hansen, Lunde and Nason (2003) which analyses forecasting performance in order to distill that group of models which contains the best forecasting model with a given confidence level. This collection of forecasting models is called the model confidence set (MCS). The models that remain in the MCS at the end of the process are assumed to have equal predictive power.

We begin the process of forming the MCS with a set of forecasting models Γ_0. The first stage of the process tests the null hypothesis that all of the these models have equal predictive accuracy (EPA) when their performance is measured against a set of ex-post observations. If $\mathbf{H}_{it}$ is the i^{th} forecast of the VCM at time t and Σ_t is the observed VCM (or a consistent estimate[15]) for the same period then the value of a loss function based on comparison of these is denoted $L(\mathbf{H}_{it}, \Sigma_t)$. The evaluation of the EPA hypothesis is based on loss differentials between the values of the loss functions for different models where the loss differential between forecasting models i and j for time t, $d_{ij,t}$, is defined as

$$d_{ij,t} = L(\mathbf{H}_{it}, \Sigma_t) - L(\mathbf{H}_{jt}, \Sigma_t).^{16} \tag{23}$$

If all of the forecastors are equally accurate then the loss differentials between all pairs of forecastors should not be significantly different from zero. The null hypothesis of EPA is then

$$H_0 : E\left(d_{ij,t}\right) = 0 \ \forall i > j \in \Gamma \tag{24}$$

and failure to reject H_0 implies all forecasting models have equal predictive ability. We test (24) using the semi-quadratic test statistic described in Hansen and Lunde (2007). If the null hypothesis is rejected at an $\alpha\%$ confidence level, we remove the model with the worst loss functions and begin the process again with the reduced set of forecasting models, Γ_1. This process is iterated until

[15] In the below forecast experiments we use the realized VCM, V_t, in place of Σ_t as it is a consistent estimator of the unobserved VCM.

the test of equal predictive accuracy cannot be rejected, or a single model remains. The model(s) which survive form the MCS with $\alpha\%$ confidence.

The loss function we use to analyse the performance of our VCM forecasts is the MVQLIKE (Stein distance) function described above in (16). This is a robust loss function, as described in Laurent, Romouts and Violante (2009). Clements, Doolan, Hurn and Becker (2009) and Laurent *et al.* (2010) established that this loss function, compared to other loss functions, identifies a correctly specified forecasting model in a smaller MCS, hence it is more discriminatory than, say, the mean square forecast error criterion.

8 Forecast Comparison - Results

Here we present the results of our forecasting competitions for 1,5 and 22 day forecast horizons. The results analysed here are the mean of MVQLIKE loss functions for forecasts and the MCS p-values. If an MCS p-value is greater than the significance level then a model is included in the MCS, otherwise it is omitted.

We present two sets of results which differ in the way in which the time variable is included into the multivariate kernel forecast. It was discussed in Section 5 that it could be introduced (see Equation 9) in such a way that observations far distant from the time at which the forecast is made are heavily penalised (Table 1). Alternatively (see Equation 10) it could be specified such that close observations can obtain a higher weight, but observations in the long past would not be excluded from attracting significant positive weights (Table 2).

Referring to the results presented in Table 1 relating to 1 day ahead forecasts, it can be concluded that the *MVK* model is the only model surviving in the MCS. All other models have a p-value smaller than 5% and are hence exluded from a 95% confidence level MCS. The forecasting model with the 2nd largest p-value is the *MVKne* forecasting model that excludes the economic variables. This allows the conclusion that for short-term forecasts the inclusion of economic variables adds value to the VCM forecasts in our setup. It is also interesting to note that the standard Riskmetrics approach (*RM* - with fixed λ), and to a lesser extend *RMopt*, deliver VCM forecast with loss functions that exceed those of other forecasting models by a large margin (considering the variation in loss functions between the other models). A further interesting conclusion can be drawn from comparing the results between *MVK* and the Riskmetrics approach that uses realized

1 Day Forecasts	$\overline{MVQ}$	p-value	5 Day Forecasts	$\overline{MVQ}$	p-value	22 Day Forecasts	$\overline{MVQ}$	p-value
MVK	13.03	1.0000	*RMvcm*	6.78	1.0000	*RMvcm*	4.43	1.0000
MVKne	13.12	0.0033	*MVKne*	6.94	0.1235	*MVK*	5.49	0.1033
RMvcm	13.31	0.0000	*MVK*	7.10	0.1235	*MVKne*	5.68	0.1033
DCC	14.65	0.0000	*DCC*	8.80	0.0004	*DCC*	6.39	0.0597
*RM*opt	17.05	0.0000	*RM*opt	13.04	0.0000	*RM*opt	21.93	0.0006
RM	43.35	0.0000	*RM*	25.35	0.0000	*RM*	26.22	0.0077

Table 1: MCS Results 1. Uses time kernel as in Equation 9. The table reports the MCS results for the multivariate kernel (with macro variables - *MVK*; without macro variables - *MVKnm*), the DCC, the Riskmetrics (*RM*), the Riskmetrics with cross-validated λ and the Riskmetrics forecasting model using realized VCM (*RMvcm*). Forecasts are for 1, 5 and 22 days horizons. $\overline{MVQ}$ is the average loss function and p-value represents the MCS p-value.

VCM and optimised λs (*RMvcm*). Essentially the latter is a special case of the former that excludes all potential weighting variables but for the time variable. The fact that *RMvcm* is excluded from the MCS for 1 day ahead forecasts indicates that matrix comparison and macroeconomic variables can add significant information to the VCM forecasting process at short forecast horizons.

The finding that the *RM* and *RMopt* deliver inferior VCM forecasts can be generalised to longer forecast horizons. In none of the forecasts comparisons looked at in this paper is any of these two forecasting models close to being included in a MCS. As the forecast horizon is increased to 5 and 22 days the relative performance of the *RMvcm* model improves significantly. For both these horizons it has the smallest loss function although it shares membership in the MCS with both the *MVK* and the *MVKnm* forecast models. This appears to indicate that the value of the matrix closeness measures and the economic variables is largest for very short-term forecasts. It is also notable that for the longest forecast horizon the *DCC* model is included in a 95% (but not a 90%) confidence level MCS.

In Table 2 we present results based on a kernel weighted VCM forecasts that use the modified time kernel proposed in Equation 10. This kernel ensures that that observations are not automatically discarded just because they occur long before the forecast period. While the non kernel forecasting methods remain unchanged[17] the MCS methodology will have to be reapplied as the MCS p-values are conditional on the initial set of forecasts used.

Some of the basic findings discussed above remain unchanged. *RM* and *RMopt* are still inferior

[17]The average loss functions for these forecasting models are identical in Tables 1 and 2.

1 Day Forecasts			5 Day Forecasts			22 Day Forecasts		
	$\overline{MVQ}$	p-value		$\overline{MVQ}$	p-value		$\overline{MVQ}$	p-value
MVK	13.06	1.0000	$RMvcm$	6.78	1.0000	$RMvcm$	4.43	1.0000
$RMvcm$	13.31	0.0110	MVK	6.90	0.2209	MVK	5.55	0.0010
DCC	14.65	0.0000	DCC	8.80	0.0000	DCC	6.39	0.0005
RMopt	17.05	0.0000	$MVKnm$	11.09	0.0000	$MVKnm$	8.87	0.0000
$MVKnm$	17.97	0.0000	RMopt	13.04	0.0000	RMopt	21.93	0.0000
RM	43.35	0.0000	RM	25.35	0.0000	RM	26.22	0.0000

Table 2: MCS Results 2. Uses time kernel as in Equation 10. The table reports the MCS results for the multivariate kernel (with macro variables - MVK; without macro variables - $MVKnm$), the DCC, the Riskmetrics (RM), the Riskmetrics with cross-validated λ and the Riskmetrics forecasting model using realized VCM ($RMvcm$). Forecasts are for 1, 5 and 22 days horizons. $\overline{MVQ}$ is the average loss function and p-value represents the MCS p-value.

to all other forecasting methodologies used here. The value of the kernel VCM forecasting method is more apparent for shorter forecasting horizons and hence the value of matrix closeness measures and economic variables disappears with increasing forecast horizon. As the time variable will have a less dominating impact on the kernel forecasts it is not surprising to find larger differences between the kernel forecasts that include macroeconomic variables (MVK) and those that do not ($MVKnm$). On all occasions the former has clearly superior average loss measures and for the 1 and 5 day forecast horizon MVK is included in the MCS while $MVKnm$ is not. At the 1 day horizon the MVK is unambiguously the best model, at the 5 day horizon it is in the MCS together with the $RMvcm$, while the latter is the unique surviving model at a 22 day forecasting horizon.

Before highlighting some more aspects of the multivariate kernel forecasts it should be noted that the results for the $RMvcm$ forecasts are rather impressive allowing for the limited information set utilised in these forecasts. While Fleming *et al.* (2003) used essentially the same model they estimated the optimal decay parameter in a slightly different way. It is apparent that this extension to the traditional Riskmetrics approach should be seriously considered in the context of high dimensional VCM forecasting.

It is interesting to compare the MVK forecast performance for the two different time kernels. When using the the kernel that converges to 0 and therefore virtually eliminates observations with large $(T-t)$ the difference in loss functions between the forecasting model that uses and that which does not use the economic variables is fairly small, although statistically significant at the 1 day forecasting horizon. When applying the time kernel that converges to 1, and hence does not penalise observations with large $(T - t)$, the difference between the two sets of weighting variables becomes larger. This result is best interpreted in combination with the observation that the loss function values for the *MVK* forecasts remain almost unchanged for both kernel types, whereas that for the *MVKnm* deteriorates as one moves from kernel (9) to (10). The obvious interpretation to this result is that evaluating similarity merely on the basis of matrix closeness measures is not a good strategy. The kernel forecast will then give weight to past observations of $\mathbf{V}_t$ that do not positively contribute to the forecasting performance. Past observations appear relevant when they are similar in terms of $\mathbf{V}_t$ and $\mathbf{V}_T$ having similar characteristics and if the prevailing macroeconomic conditions at time T are close to those at time t.

In both Table 1 and Table 2 we find evidence that the macroeconomic information and VCM based similarity statistics significantly improve on a kernel based purely on time at a 1 day forecast horizon[18]. We investigate which of the potential kernel weighting variables considered tend to survive the variable elimination strategy described in Section 5.1.

The survival probabilities are reported in Table 3, these report the percentage of times that a variable is included in the kernel for forecasting horizon, as mentioned above the inclusion/exclusion decision is made every 200 days, or seven times for each forecast horizon. This Table refers to results generated with the time kernel that converged to 0 (Equation (9))[19]. It is apparent that a good number of the variables considered remained in the weighting mechanism in the vast majority (if not all) instances. Variables that were eliminated more often than not are the correlation comparison variable and the MVQLIKE measure of closeness. There is little variation between the different forecast horizons with the exception that the comparison of absolute differences is dropped and the correlation comparison included more often for the 22 day horizon. In general the inclusion frequencies drop somewhat for the 22 day forecast horizon. The exception are the term spread as well as oil and gold prices which are never dropped at any forecasting horizon.

[18] We also ran versions of the model in which the 5 and 22 day periods in the estimation of the bandwidths were non-overlapping. We found that this marginally improved mean MVQs but did not alter the qualitative nature of the MCS analysis.

[19] Note that the results for all variables other than time are identical regardless of the time variable we use as the inclusion/exclusion is based on a univariate kernel.

Variable	Percentage Inclusion		
	1 Day	5Day	22 Day
Ratios of Eigenvalues of $\mathbf{V}_T$ and $\mathbf{V}t$	100%	100%	85.7%
Absolute relative differences of elements of $\mathbf{V}_T$ and $\mathbf{V}t$	100%	100%	0%
Proportion of correlations with the same sign at time t and T.	0%	0%	42.9%
MVQLIKE of $\mathbf{V}t$ when compared to $\mathbf{V}_T$	28.6%	71.4%	57.1%
Term Spread of government bonds.	100%	100%	100%
Default spread on corporate debt	100%	85.7%	85.7%
Oil Price	100%	100%	100%
Gold Price	100%	100%	100%
VIX index	85.7%	100%	42.9%
Time	100%	100%	85.7%
Bull and Bear Market Phases	100%	100%	85.7%

Table 3: Variables Included in the Kernel Model.

9 Conclusion

This paper presents a flexible kernel model which can be used to forecast symmetric, positive definite variance covariance matrices for large scale portfolio stocks while being able to incorporate a wide array of data. This is in contrast to many of the more popular approaches to VCM modelling which have to make simplifications to their estimation processes/paramatisations in order to handle large scale covariance matrices. The model relies on techniques well established in the nonparametric econometrics literature. Importantly the scale of the computational task scales with the number of variables we wish to use in determining our kernel weights rather than the dimension of the covariance matrix. Our model is flexible, capable of using a wide range of economic information and can be as easily used for small and large scale matrices. It does, however, depend on the availability of positive definite VCM estimates which may not be easily available for vast dimensional problems.

This paper establishes the feasibility of the proposed forecasting approach and further demonstrates that using a larger set of information (matrix closeness measures and economic variables) can have statistically significant advantages. These advantages seem to be strongest for very short forecast horizons. We also found that a version of the popular Riskmetrics model, using VCMs based on high frequency data and cross-validated decay parameter, provided extremely useful. While the kernel method dominated on the very short horizons, the modified Riskmetrics approach performed best for 5 and 22 day ahead forecasts. This very simple forecasting method has not attracted much attention in the empirical literature and it is suggested that its merits should be reevaluated.

A number of issues for further research are beyond the scope of this paper. While we evaluated forecast performance with statistical measures, future research will establish whether the proposed forecasting methodology will deliver economically significant improvements. We also anticipate that the ability of incorporating exogenous information into the VCM forecasting process will allow researchers to re-evaluate the type of variables considered in the context of VCM forecasting.

10 Bibliography

References

Aït-Sahalia, Y. & Brandt, M.W. (2001) "Variable selection for portfolio choice", *The Journal of Finance,* Vol. 56, no. 4, pp. 1297-1351.

Aitchison, J. & Aitken, C.G.G. (1976) "Multivariate binary discrimination by the kernel method", *Biometrika,* Vol. 63, no. 3, pp. 413-420.

Bowman, A.W. (1984) "An alternative method of cross-validation for the smoothing of density estimates", *Biometrika,* Vol. 71, pp. 353-360.

Bowman, A. W. (1997) *Applied smoothing techniques for data analysis : the kernel approach with S-Plus illustrations,* Clarendon Press, Oxford.

Blair, B. J., Poon, S. H., & Taylor, S. J. (2001), "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns", *Journal of Econometrics*, vol. 105, no. 1, pp. 5-26

Campbell, J.Y. (1987) "Stock returns and the term structure", *Journal of Financial Econometrics,* Vol. 18, pp. 373-399.

Campbell, J.Y., Lettau, M., Malkiel, B.G. & Xu, Y. (2001) "Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk" *The Journal of Finance,* Vol. 56, No. 1, pp1-43.

Clements, A., Doolan, M., Hurn, S. & Becker, R. (2009) "Evaluating multivariate volatility forecasts", NCER Working Paper Series 41, National Centre for Econometric Research.

Clements, A., Hurn, S. & Becker, R. (2010) "Semi-Parametric Forecasting of Realized Volatility", to appear in: *Studies in Nonlinear Dynamics & Econometrics.*

Chiriac, R. & Voev, V. (2008) "Modelling and forecasting multivariate realized volatility", CoFE Discussion Paper 08-06, Center of Finance and Econometrics, University of Konstanz.

Colacito, R., Engle, R.F. & Ghysels, E. (2007), "A component model for dynamic correlations", Unpublished

Engle, R.F (2002) "Dynamic conditional correlation - a simple class of multivariate GARCH models", *Journal of Business and Economic Statistics*, Vol. 20, no. 3, pp. 339-350.

Engle, R.F. & Kelly, B.T. (2008), "Dynamic equicorrelation", NYU Working Paper FIN-08-038.

Engle, R. F. & Sheppard, K. (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH", NBER Working Paper No. 8554.

Engle R.F. Shephard N. and Sheppard K. (2008) “Fitting vast dimensional time-varying covariance models”, unpublished mimeo, http://www.oxford-man.ox.ac.uk/~nshephard/.

Fama, E.F. & French, K.R. (1989) "Business conditions and expected returns on stocks and bonds", *Journal of Financial Economics,* Vol. 25, pp. 23-49.

Fleming, J., Kirby, C. & Ostdiek, B. (2003) "The economic value of volatility timing using ’realized’ volatility", *Journal of Financial Economics,* Vol. 67, no. 3, pp 473-509.

Gijbels, I., Pope, A. & Wand, M.P. (1999) "Understanding exponential smoothing via kernel regression", *Journal of the Royal Statistical Society,* Vol. 61, pp. 39-50.

Hamilton, J.D. & Lin, G. (1996), "Stock market volatility and the business cycle", *Journal of Applied Econometrics,* Vol. 11, No. 5, pp. 573-93.

Hamilton, J.D. (1996), "This is what happened to the oil price-macroeconomy relationship", *Journal of Monetary Economics,* Vol. 38, pp. 215-220.

Hansen, P.R. A test for superior predictive ability. Brown University, Department of Economics Working Paper 2003-09 . 2001.

Hansen, P.R. & Lunde, A. (2005) "A realized variance for the whole day based on intermittent data", *Journal of Financial Econometrics*, Vol 3 (4), pp525-554.

Hansen, R.P. & Lunde, A. (2007) "MULCOM 1.00, Econometric toolkit for multiple comparisons" (Packaged with Mulcom package)

Hansen, P. D., Lunde, A., & Nason, J. M. (2003), "Choosing the best volatility models: the model confidence set approach", *Oxford Bulletin of Economics and Statistics*, vol. 65, no. Supplement, pp. 839-861.

Hansen, P.R., Lunde, A. & Nason, J.M. (2004), 'Model confidence sets for forecasting models',Federal Reserve Bank of Atlanta Working Paper No. 2005-7 http:\\ssrn.com\\paper=522382.

Harvey, C.R. (1989) "Time-varying conditional covariance in tests of asset pricing models", *Journal of Financial Economics,* Vol. 24, pp. 289-317.

Harvey, C.R., 1991, The specification of conditional expectations, Working paper, Duke University.

J.P. Morgan (1996) *Riskmetrics Technical Document 4th Edition*, J.P. Morgan, New York.

Krolzig, H.M. & Hendry, D.F. (2001) "Computer automation of general-to-specific model selection procedures", *Journal of Economic Dynamics & Control,* Vol. 25, pp. 831-866.

Laurent, S., Rombouts, J.V.K. & Violante, F. (2009) "On loss functions and ranking forecasting performances of multivariate volatility models," CIRPEE Working Paper 09-48.

Laurent, S., Rombouts, J.V.K. & Violante, F. (2010), "On the forecasting accuracy of multivariate GARCH models", CIRPÉE Working Paper 10-21.

Laws, J. & Thompson, J. (2005) "Hedging effectiveness of stock index futures", *European Journal of Operational Research,* Vol. 163, pp. 171-191.

Li, Q. & Racine, J.S. (2007) *Nonparametric Econometrics Theory and Practice,* Princeton University Press, Oxford.

Moskowitz, T.J. (2003) "An analysis of covariance risk and pricing anomalies", *The Review of Financial Studies*, Vol. 16, pp 417-457.

Pagan, A.R. & Sossounov, K.A. (2003) "A simple framework for analysing bull and bear markets", *Journal of Applied Econometrics,* Vol. 18, No. 1, pp. 23-46.

Pelletier, D. (2006) "Regime switching for dynamic correlations", *Journal of Econometrics*, Vol. 131, no. 1-2, pp. 445-473.

Poon S-H. and Granger C.W.J. (2003) "Forecasting volatility in financial markets: a review" *Journal of Economic Literature*, 41, 478-539.

Rudemo, M. (1982) "Empirical choice of histograms and kernel density estimators", *Scandinavian Journal of Statistics,* Vol. 9, No. 2, pp65-78.

Sadorsky, P. (1999) "Oil price shocks and stock market activity", *Energy Economics,* Vol. 21, pp. 449-469.

Schwert, G.W. (1989) "Why does stock market volatility change over time", *The Journal of Finance,* Vol. 44, no. 5, pp 1115-1153.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Sjaastad, L.A. & Scacciavillani, F. (1996) "The price of gold and the exchange rate", *Journal of international Money and Finance,* Vol. 15, No. 6, pp. 879-97.

Vasilellis, G.A. & Meade, N. (1996), "Forecasting volatility for portfolio selection", *Journal of Business, Finance and Accounting,* Vol. 23, pp. 125-143.

Wand, M.P. & Jones, M.C. (1995) *Kernel Smoothing,* Chapman & Hall, London.

Whitelaw, R.F. (1994) "Time variations and covariations in the expectation and volatility of stock market returns", *The Journal of Finance,* Vol. 49, No.2, pp. 515-541.

11 Appendix A

Table 4 provides a full list of the stocks used in the analysis in section XX.

	Ticker symbol	Company Name
1	AA	Alcoa Inc
2	AXP	American Express Inc
3	BA	Boeing Co.
4	BAC	Bank of America Corp.
5	BMY	Bristol Myers Squibb Co.
6	CL	Colgate Palmolive
7	DD	El DuPont de Nemours co.
8	DIS	Walt Disney Corp.
9	GD	General Dynamics Corp
10	GE	General Electric Co.
11	IBM	IBM
12	JNJ	Johnson & Johnson
13	JPM	JP Morgan Chase Co.
14	KO	Coca Cola Corp.
15	MCD	Mcdonald's Corp
16	MER	Merril Lynch Co. Inc
17	MMM	3M co.
18	PEP	Pepsico Inc.
19	PFE	Pfizer Inc
20	TYC	Tyco International ltd.

Table 4: Stocks included in the forecasting experiment in section X.

12 Appendix B

In the text of the article reference is made to the composite-likelihood DCC calculation method suggested in Engle, Sheppard and Shephard 2008 in order to make estimation fo the DCC paramaters feasible for a large scale problem.

In a DCC model we assume that

$$\begin{aligned} \mathbf{r}_t &= \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &= \mathbf{H}_t^{1/2}\mathbf{z}_t \\ \mathbf{z}_t &\sim IID(\mathbf{0}_n, \mathbf{I}_n) \end{aligned}$$

and $\mathbf{H}_t$ is governed by a process dependent on a paramater vector θ, which contains the values of α, β and the off unique off diagonal elements of $\bar{\mathbf{Q}}$ from (20). The paramater vector is estimated by maximising the likelihood equation

$$\begin{aligned} \log L(\theta, F_T) &= \sum_{t=1}^{T} \mathcal{L}_t(\theta) \\ \mathcal{L}_t(\theta) &= -\frac{1}{2} \log |\mathbf{H}_t| - \frac{1}{2} \mathbf{r}_t^{\prime} \mathbf{H}_t \mathbf{r}_t, \end{aligned}$$

however it is difficult to maximise this if we have a high dimensional VCM.

The method suggested by Engle, Sheppard and Shephard 2008 to circumvent this problem is composite-likelihood estimation. The first step in this is to estimate a likelihood for several pairs of stocks. For example if we have three stocks we may compute the likelihoods for DCC models using the pairs of stocks $(1,2), (2,3)$ and $(1,3)$ and we denote these likelihoods as $\mathcal{L}_{j,t}(\theta_j)$ $j = 1, 2, 3$ respectively. The same equations as above can be calculated using the equation for $\mathcal{L}_t(\theta)$ above, however the dimension if $\mathbf{H}_t$ and $\mathbf{r}_t$ in this case are 2x2 and 2x1 respectively. The composite likelihood approach then suggests that we sum these likelihoods time and average over the Z pairwise combinations included. We then choose the paramaters to minimise the sum

$$CL(\theta) = \frac{1}{Z} \sum_{t=1}^{T} \sum_{z=1}^{Z} \mathcal{L}_{j,t}(\theta_j).$$

Hence we find the DCC paramaters which minimise the sum of likelihoods. By doing this it is possible to estimate the DCC paramaters using only pairwise calculations, however the estimated paramaters will not be the same unless the pairs are independent. The paramater vector θ_j varies over $j = 1, 2, 3$ only in the long run correlation paramater for the pair of stocks, obtained from the standardised GARCH residuals from the first step of DCC estimation . As in the aggregate models these are replaced by the full set of long run correlations in θ there is a difference between θ and each θ_j.

Engle, Sheppard and Shephard 2008 note that the loss in efficiency from using less than all available pairs is only very small and so we include each of our stocks in only one pairing of stocks for calculation of the composite-likelihood DCC.