A Comparison of Covariance Forecasts from High-Frequency, Daily and Option Data

Dudley Gilder *

January 14, 2010

Abstract

The relative merits of forecasting volatility using high-frequency, daily and option data have received much attention. The objective of this paper is to extend this research to the covariance matrix. Forecasts are compared for covariance matrices composed of 17 stocks over 4 horizons, 1-, 30-, 90- and 180-days. Since it is not possible to invert equity option prices to obtain a covariance estimate, a market model assumption is made and forecasts are made based on implied betas and variances. Consequently, these are denoted option-factor-implied forecasts. An additional 6 models are included in the comparisons which can be categrorised into multivariate GARCH models, models estimated directly on the realised covariances (RC) and historical averages. Multiple forecast evaluation criteria are used to obtain a more complete understanding of each model's performance. These include a statistical loss function, optimlaity tests based on Mincer-Zarnowitz regressions and an economic loss function. On balance, it is very difficult to identify a single model as superior. The option-factor-implied forecasts appear to perform poorly under the statistical loss function whilst they are one of the more informative according to the R^2 from Mincer-Zarnowitz regressions. A persistent positive bias appears to be the source of the conflicting results. However, a simple historical average computed from high-frequency data performs well under all evaluation criteria, bringing into question the benefits of dynamic models in forecasting covariance matrices.

EFM Classification Codes: 330, 370, 410

Keywords: High-Frequency Data, Option Prices, Realised Covariance, Multivariate GARCH, Covariance Forecasting

*Department of Accounting and Finance, Lancaster University Management School, Lancaster, LA1 4YX, UK. Tel: (+44)01524 592445. e-mail: d.gilder@lancaster.ac.uk.

1 Introduction

Understanding the dependence between financial assets is important. This stems from individual investors and institutions holding portfolios of assets, the risk of which can only be appreciated with a measure of their dependence. Although many measures exist, correlation is the most prevelant and represents an important metric in many applications. For example, in risk management, the correlation between different business units must be identified in order for hedging strategies to be implemented correctly. For banks, correlation has been formally incorporated into the regulatory framework of the Basel II accords which stipulate rules on computing capital adequacy (Das, 2007). Correlations are also important in hedging certain derivative securities, such as differential swaps, where the payoffs cannot be replicated from the separate underlying assets (Mahoney, 1997). They have become central to the market convention in pricing credit derivatives, one example being Credit Default Obligations (CDO), where traders quote implied correlations (Collin-Dufresne, 2009). Then, of course, there is the mean-variance portfolio theory of Markowitz (1952), which forms the basis for a large proportion of finance theory.

Although correlation is known to have disadvantages in some contexts, especially where extreme, tail-events, are the object of interest (Embrechts, 2009), it remains an important tool in constructing diversified equity portfolios. A mean-variance efficient portfolio still provides considerable diversification benefits with an accurate forecast of the correlation, or covariance, matrix being essential (Ledoit and Wolf, 2003) [1]. Indeed, there is active research investigating the value of equity covariance matrix forecasts for a range of applications, including Value at Risk (Skintzi and Xanthopoulos-Sisinis, 2007), portfolio hedging (Lien et al. (2002); Alexander and Barbosa (2008)), basket option pricing (Gibson and Boyer (1998); Byström (2002)) and asset allocation (Fleming et al., 2003). The focus here is to analyse the forecasting performance of covariance matrix models based on three different data types: low-frequency daily data, high-frequency intraday data and option price data.

Covariance matrix modelling has received much less attention than volatility modelling and innovations have concentrated on developing multivariate GARCH models that use daily data. One of the earliest was the VEC model of Bollerslev et al. (1988), which modelled each element of the covariance matrix as a GARCH process, and highlighted two fundamental problems: The necessity to guarantee positive definiteness, or, at the least, positive semi-definiteness, and the need to devise a parsimonious model which also provides an accurate representation of the covariance dynamics, otherwise known as the curse of dimensionality. Neither of these requirements is met by the VEC model. Subsequent models have largely been developed to address these problems, the constant correlation and BEKK models of Bollerslev (1990) and Engle and Kroner (1995), respectively, being examples in this succession. However, the most popular multivariate GARCH model is the dynamic conditional correlation (DCC) model of Engle (2002), which can genuinely claim to have solved the positive definiteness and curse of dimensionality problems[2].

Yet the univariate volatility forecasting literature, on balance, suggests that option-implied volatility forecasts may be superior relative to those based on daily data. For example, Jorion (1995) shows that volatility forecasts implied by the Black-Scholes (BS) model outperform those generated by GARCH models, even in-sample, for three currencies. For the S&P 100, Christensen and Prabhala (1998) and Fleming (1998) redress the problem of spurious regression resulting from the evalutation of temporally overlapping forecasts to find implied volatility provides more accurate forecasts relative to historical based estimators. Similarly, Blair et al. (2001) and Koopman et al. (2005) use the VIX index to form their implied volatility forecasts, also finding these to be superior to those based on low-frequency daily data. A more recent development has been the introduction

[1]Note, the terms correlation matrix and covariance matrix are deemed interchangeable.

[2]An almost identical model is suggested by Christodoulakis and Satchell (2002).

of model-free volatility by Britten-Jones and Neuberger (2000) and Bakshi et al. (2003) in which the computation of implied volatility does not depend on a specific option pricing model. Jiang and Tian (2005) demonstrates its superiority to both BS implied and historical volatility forecasts. In a much wider review of the literature, Poon and Granger (2003, 2005) conclude that option-implied volatility probably provides the best forecasts. Therefore the extension of option-implied metrics to the covariance matrix is important to investigate.

Another innovation in volatility forecasting has been the use of high-frequency realised variance (RV), where the reduction in sample variation, relative to that in the alternative proxy of squared daily returns, is sufficient enough to render volatility observable and allows standard time series tools to be employed in its modelling (Andersen et al., 2001b). Based on the observations of Andersen et al. (2001b) and Andersen et al. (2001a) that RV is likely to follow a long memory process in the foreign exchange and equity markets, Martens and Zein (2004), Pong et al. (2004) and Koopman et al. (2005) have shown fractionally integrated ARMA (ARFIMA) models of RV to be able to forecast either the S&P 500, Yen-USD or Sweet Crude Oil prices, the GBP-USD, DM-USD and Yen-USD exchange rates and the S&P 100 index, respectively, more accurately than low-frequency GARCH models. In fact, these forecasts are able to compete with implied volatility, leading Koopman et al. (2005) to conclude that long-memory S&P 100 RV forecasts are superior to the VIX index. The use of RV also improves the accuracy with which volatility forecasts can be assessed. Substituting daily squared returns for RV is shown by Andersen and Bollerslev (1998) to theoretically increase R^2 in Mincer-Zarnowitz regressions of GARCH volatility forecasts. More importantly, they show that using an inaccurate forecast target can spuriously reduce the assessed forecasting accuracy of GARCH models, even when such models represent the true returns process.

Given the performance of volatility forecasts based on RV, it is important that the set of competing covariance matrix forecasting models compared here includes those based on high-frequency data. In analogy to RV, Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) show high-frequency data can be used to estimate realised covariance (RC) by summing cross-products of intraday returns and Chiriac and Voev (2007, 2008) have developed a long-memory, vector ARFIMA (VARFIMA) model to forecast covariance matrices based on its Cholesky decomposition.

In this study, the performance of 7 models in forecasting the covariance matrix of 17 stocks is examined. Data is available from January 2002 to September 2008 and is divided into in-sample, from 02/01/2002 to 04/01/2006 , and out-of-sample, from 05/01/2006 to 25/09/2008, periods for the purposes of forecasting. The models include the DCC, the Factor Double ARCH, a Cholesky Factor ARMA (CF-ARMA) model, a Factor ARMA model, an option-factor-implied model and the historical estimates of covariance computed from low-frequency daily data and high-frequency intraday data. Since it is not possible to infer implied covariances directly from individual equity option prices, covariance matrices were estimated indirectly through the assumption of a one-factor market model. The implied betas necessary to implement this strategy were computed using the methodology of Chang et al. (2009), hence the denotation option-factor-implied model. The Factor Double ARCH, discussed in Engle (2009a), and the Factor ARMA also assume a one-factor market model and are included to separate the effect of the market model assumption on forecasting performance from that of the data type. In order to simplify estimation, the long-memory model of Chiriac and Voev (2007, 2008) applied to Cholesky factors was replaced by an ARMA(2,1), which has been shown to provide almost indistinguishable properties to that of a true long-memory ARFIMA model when applied to RV (Pong et al., 2004), giving the CF-ARMA model. The historical estimators follow the usual formulae for covariances and RC.

In evaluating forecasts, several criteria are emloyed. The first is based on a statistical loss function, the mean squared error (MSE), which is commonly used in forecast evaluation. It also has the advantage of being robust in the sense that the ranking of forecast accuracy is not affected by the choice of (unbiased) proxy for the true covariance matrix, as defined in Patton (2008) and

Patton and Sheppard (2009). The second is the optimality of the forecasts, tested in the usual manner by Mincer-Zarnowitz regressions. This criterion assesses both whether the forecasts are unbiased and whether they are orthogonal to information available at the time the forecasts were made. In this case, the information set is composed of the forecasts from each model, so that it should not be possible to improve the forecast generated by one model by including information contained in an alternative model. Lastly, an economic loss function is introduced. Following Engle and Colacito (2006), hedging portfolios are formed, where weights are selected to minimise the variances of portfolios designed to hedge the returns of one stock in the sample, resulting in a total of 17 hedging portfolios for each model. In addition to evaluating the accuracy of the forecasts, this criterion allows an economically meaningful interpretation to be given to their performance. By evaluating forecasts over these multiple criteria, a more complete assessment of the merits and demerits of each model can be made, something that cannot be achieved by limiting the evaluation to any single criterion.

The remainder of this paper is organised as follows. Section 2 describes the models employed and the forecast evaluation criteria. Section 3 discusses the data and selection of the stocks whilst Section 4 and Section 5 discuss the in-sample and out-of-sample forecasting results, respectively. Section 6 concludes.

2 Forecasting Procedure

2.1 High-Frequency Covariance

Like RV, the empirical application of RC suffers from imperfections in high-frequency data which can be broadly categorised into microstructure noise and price asynchronicity. Microstructure noise arises as returns are sampled over increasingly shorter intervals and can be attributed to phenomena such as bid-ask bounce and price discreteness whilst price asynchronicity refers to the asynchronicity between the times at which prices of different assets are observed. Due to the importance of RC, both as a forecast target and to the models that directly model its dynamics, three forms of RC are estimated[3]: An unadjusted form (standard realised covariance), an adjusted form, through the inclusion of lead and lag terms (lead-lag adjusted realised covariance) and a subsampled version. Definitions of each form given below are for the bivariate case.

Let high-frequency prices be sampled at intraday times $t_q = t + \Delta q$, for $q = 0, \ldots, M$ and $M = 1/\Delta$, where Δ is a fixed time interval and the length of a day is standardised to take the value $t_M - t_0 = 1$. Therefore, the high-frequency return over the intraday interval from $\Delta(q-1)$ to Δq for asset i on day t is computed from,

$$r_{t,q,i} = y_{t,q,i} - y_{t,q-1,i},$$

where $y_{t,q,i} = \log p_{t,q,i}$, the log of the intraday price for asset i at time t_q. Standard RC can then be computed from,

$$RC_t^{snd} = \sum_{q=1}^{M} r_{t,q,i} r_{t,q,j}. \qquad (1)$$

[3] Many alternative estimators have been posited to remove the distortions introduced. For example see Scholes and Williams (1977), Cohen et al. (1983), Bandi and Russell (2005), Hayashi and Yoshida (2005), Sheppard (2006), Zhang (2006), Barndorff-Nielsen et al. (2008a).

An implicit assumption in computing RC_t^{snd} is that intraday prices are observed on the sample grid comprised of the times $t_q = t + \Delta q$, $q = 0, \ldots, M$ and, hence, the interval over which high-frequency returns are measured is Δ. Despite high-frequency data providing a complete record of all ticks, there is no guarantee they will occur at these times and empirically this is rarely the case[4]. Consequently, an imputation procedure is required to match observed prices to the sample grid, two examples of which are the previous tick and interpolated tick methods (see Dacarogna et al., 2001).

To represent prices at the times on the sample grid, the previous tick method involves taking prices observed immediately prior to or at these times. Specifically, if $p_{t,s,i}$ represents the price of asset i observed at the irregularly spaced times t_s, corresponding to observation times, then prices at the grid points are set according to,

$$p_{t,q,i} = \{p_{t,s,q}; \max(t_s \leq t_i)\}.$$

In contrast, if a price is not observed on a grid point, the interpolation method requires two prices to be sampled; one either side of the grid point. Therefore, if t_s and t_{s+1} represent, respectively, the times at which the price is observed immediately prior to and after the grid point, t_q, so that $t_s \leq t_q \leq t_{s+1}$, then the price, $p_{t,q,i}$ can be computed by linearly interpolating between $p_{t,s,i}$ and $p_{t,s+1,i}$.

Although the imputation methods provide consistent procedures to place irregularly observed prices onto a regular sample grid, asynchronicity between the times at which the prices of different assets are observed will exist. In turn, the intraday returns of the assets will be measured over different time intervals which may introduce cross-correlation between the returns[5]. In order to account for the induced cross-correlation, Cohen et al. (1983) suggest adjusting RC_t^{snd} by including the cross-products of the contemporaneous and lagged returns of the two assets, giving the adjusted form of RC_t,

$$RC_t^{adj} = \sum_{q=1}^{M} \sum_{l=-L}^{L} r_{t,q,i} r_{t,q-1,j}. \tag{2}$$

The form of RC_t^{adj} in (2) ignores the impact of the opening and close of the market (end effects). For example, at $q = 1$ it is not possible to compute the cross-product between the return of asset i and any lagged returns of asset j. Although rendering RC_t^{adj} inconsistent, in application end effects can be safely ignored. An issue, investigated in Martens (2004), involves the optimal choice of L. For simplicity, the value of L is set to $L = 1$ in empirical applications.

The final form of RC involves subsampling and is best described through an example. Take the RC_t^{snd} computed using a grid of 5-min-returns. For a trading day covering the 09:30 to 16:00 period, there would be 78 5-min-returns, where the sample grid points lie at 09:30, 09:35,. . . ,16:00. However, 5 min subgrids may also be formed. If the first price observation is taken at 09:31 instead, then a subgrid of 5-min-returns can be formed with grid points occurring at 09:31, 09:36, . . . , 15:56. The next subgrid can be formed by taking the first price observation at 09:32 and the process may continue until an additional 4 subgrids have been formed. The RC_t^{snd} may then be computed on each of these subgrids, where it is multiplied by an adjustment factor $M/(M - 1)$ to account for the loss of one intraday return. For G subgrids,

$$RC_t^{sub} = \frac{1}{G} \sum_{g=1}^{G} RC_{t,g}^{snd}, \tag{3}$$

[4]The term tick refers to a price observation, taking the form of either a transaction or quote. In future references, the precise category a tick belongs to will be clear from the context. For now, the term is used for its generality.

[5]Cross correlation refers to the correlation between the contemporaneous return of one asset and the lagged return of another.

where RC_t^{sub} is the subsampled RC and $RC_{t,g}^{snd}$ is the RC_t^{snd} computed on subgrid g. An additional benefit is a reduction in the sample variance of RC. This has been shown theoretically for realised volatility by Zhang et al. (2005) and the concept has been applied to RC_t in Martens (2004) and de Pooter et al. (2008).

2.2 Covariance Models and Forecasting

2.2.1 The Dynamic Conditional Correlation Model

The DCC model of Engle (2002) is a popular multivariate GARCH model applied to daily returns, providing a parsimonious representation of the dynamics of the covariance matrix and guarantees positive definiteness[6]. To understand the model, use the following decomposition of the N asset covariance matrix,

$$\mathbf{H_t} = \mathbf{D_t \Gamma_t D_t}, \tag{4}$$

where, $\mathbf{H_t}$ is the $N \times N$ conditional covariance matrix, $\mathbf{D_t}$ are $N \times N$ diagonal matrices of conditional variances and $\mathbf{\Gamma_t}$ is the $N \times N$ conditional correlation matrix. Modelling the conditional covariance matrix is then a matter of specifying the dynamics of the conditional variances and correlations, both of which may be selected independently of one another. Here, the asymmetric GJR-GARCH model of Glosten et al. (1993) is chosen as the model for the conditional variances,

$$\begin{aligned} r_{i,t} &= \mu_i + \theta_i e_{i,t-1} + e_{i,t}, \\ e_{i,t} &= \sqrt{h_{i,t}}\varepsilon_{i,t}, \\ h_{i,t} &= \omega_i + \alpha_i e_{i,t}^2 + \gamma_i \mathbb{I}_{\{e_{i,t}<0\}} e_{i,t}^2 + \beta_i h_{i,t-1}, \\ \varepsilon_{i,t} &\sim \mathcal{N}(0,1), \end{aligned} \tag{5}$$

where $i = 1, \ldots, N$, $r_{i,t}$ is the daily log return of the i^{th} asset, $e_{i,t}$ is the return innovation to the i^{th} asset, $h_{i,t}$ is the conditional variance of the i^{th} asset and $\mathbb{I}_{e_t<0}$ is an indicator function taking the value 1 when $e_{i,t} < 0$ and 0 otherwise. To construct the dynamic model for conditional correlations, there are three steps: Forming and modelling quasi-correlations, correlation targeting and rescaling. First, take the standardised return innovations $\boldsymbol{\varepsilon}_t = \mathbf{D}_t^{-1}\mathbf{y}_t$ and model their conditional covariance matrix, the quasi-correlations, $\mathbf{Q}_t$, as,

$$\mathbf{Q}_t = \mathbf{\Omega}_{qc} + \alpha_{qc}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta_{qc}\mathbf{Q}_{t-1}, \tag{6}$$

where $\mathbf{\Omega}_{qc}$ is an $N \times N$ matrix of intercept parameters $\omega_{i,j}$ and α_{qc} and β_{qc} govern the dynamics of the process[7]. To perform correlation targeting, take the expectation of (6) to get[8],

$$\mathbf{\Omega_{qc}} = (1 - \alpha_{qc} - \beta_{qc})\mathbf{\Gamma},$$

which can be approximated by,

$$\begin{aligned} \widehat{\mathbf{\Omega}} &= (1 - \alpha_{qc} - \beta_{qc})\bar{\mathbf{R}}, \\ \bar{\mathbf{R}} &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'. \end{aligned}$$

[6] As the name suggests, DCC models correlations, but it is simple to retrieve covariances.

[7] There are only $N(N-1)/2$ unique parameters in $\mathbf{\Omega}_{qc}$ since the diagonal elements can be restricted to be equal to 1.

[8] See Engle and Mezrich (1996) for an account of volatility targeting in GARCH models.

The matrix $\mathbf{Q}_t$ can now be represented by,

$$\mathbf{Q}_t = \bar{\mathbf{R}} + \alpha_{qc}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} - \bar{\mathbf{R}}\right) + \beta_{qc}\left(\mathbf{Q_{t-1}} - \bar{\mathbf{R}}\right). \quad (7)$$

Therefore correlation targeting reduces the number of parameters in (6) to two. Positive definiteness is guaranteed when $\{\alpha_{qc}, \beta_{qc}, (1-\alpha_{qc}-\beta_{qc})\} > 0$ and the initial $\mathbf{Q}_t$ is positive definite. In the last step, the quasi-correlations must be rescaled. Although $\mathbf{Q}_t$ is positive definite, the elements may violate the condition that correlations lie between -1 and 1. The correlations can be retrieved from,

$$\rho_{i,j,t} = \frac{Q_{ij,t}}{\sqrt{Q_{ii,t}Q_{jj,t}}},$$

where $\rho_{i,j,t}$ is the $(i,j)^{\text{th}}$ element of $\mathbf{\Gamma_t}$.

Estimation of the full model is straightforward and can be conducted by maximum likelihood. Given the conditional normality of returns, the full log likelihood can be written as,

$$LL = -\frac{1}{2}\sum_{t}^{T}\left(n\log(2\pi) + 2\log|\mathbf{D}_t| + \mathbf{r}'_t\mathbf{D}_t^{-2}\mathbf{r}_t - \boldsymbol{\varepsilon}'_t\boldsymbol{\varepsilon}_t + \log|\mathbf{R}_t| + \boldsymbol{\varepsilon}'_t\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t\right).$$

Note, the first three terms contain the data and the variance parameters, whilst the last three terms contain the variance adjusted data and the correlation parameters allowing a two-step procedure to be employed. In the first step, the univariate variance are estimated from which standardised returns are formed and used in the second step to estimate the parameters governing the correlations. If $\boldsymbol{\xi} = \left\{\boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\theta}, \bar{\mathbf{R}}, \alpha_{qc}, \beta_{qc}\right\}$, then it is known,

$$\sqrt{T}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}\right), \quad (8)$$

where $\hat{\boldsymbol{\xi}}$ is the vector of estimated parameters, $\boldsymbol{\xi}_0$ is the vector of true parameters, $\mathbf{A}_0$ is the Hessian matrix and $\mathbf{B}_0$ is the expected value of the outer product of the score vector. In the results reported below these are estimated using numerical derivatives[9].

The DCC model naturally generates one-step-ahead correlation and variance forecasts, which can be used to form a covariance forecast. In order to forecast k steps ahead, where $k > 1$, the following assumption is used (see Engle and Sheppard, 2001),

$$\mathbb{E}_t\mathbf{R}_{t+k} = \mathbb{E}_t\mathbf{Q}_{t+k}.$$

Thus, k-step-ahead forecasts are obtained from,

$$\mathbb{E}_t\mathbf{R}_{t+k} = \bar{\mathbf{R}} + (\alpha_{qc} + \beta_{qc})^{k-1}\left(\mathbb{E}_t\mathbf{R}_{t+1} - \bar{\mathbf{R}}\right).$$

Similarly, k-step-ahead conditional variance forecasts can be obtained from,

$$\mathbb{E}_t h_{i,t+k} = \sigma_i^2 + \left(\alpha_i + \frac{1}{2}\gamma_i + \beta_i\right)^{k-1}(h_{i,t+1} - \sigma_i^2), \quad (9)$$

where,

$$\sigma_i^2 = \frac{\omega_i}{1 - \alpha_i - \frac{1}{2}\gamma_i - \beta_i}.$$

[9] Estimation was conducted using the Oxford MFE Toolbox written by Kevin Sheppard, available from http://www.kevinsheppard.com/wiki/MFE_Toolbox.

2.2.2 Factor Double ARCH

The Factor Double ARCH model assumes returns are generated by the market model and that the variance of the idiosyncratic component follows a GARCH specification (see Engle, 2009a, chap. 8). Therefore,

$$\begin{aligned} r_{i,t} &= \mu_i + \beta_{m,i}^{ARCH} r_{m,t} + e_{i,t}, \\ e_{i,t} &= \sqrt{h_{i,t}} \varepsilon_{i,t}, \\ h_{i,t} &= \omega_i + \alpha_i e_{i,t}^2 + \gamma_i \mathbb{I}_{\{e_{i,t}<0\}} e_{i,t}^2 + \beta_i h_{i,t-1}, \\ \varepsilon_{i,t} &\sim \mathcal{N}(0,1), \end{aligned}$$

where $\beta_{m,i}^{ARCH}$ is the beta of the i^{th} asset and $r_{m,t}$ is the market return, otherwise all other variables are as defined in (5). Conditional variances and covariances can then be computed from,

$$\begin{aligned} V_{t-1}(r_{i,t}) &= \left(\beta_{m,i}^{ARCH}\right)^2 V_{t-1}(r_{m,t}) + V_{t-1}(e_{i,t}), \\ Cov_{t-1}(r_{i,t} r_{j,t}) &= \beta_{m,i}^{ARCH} \beta_{m,j}^{ARCH} V_{t-1}(r_{m,t}). \end{aligned}$$

To implement these measures the variance of market returns is modelled using the GJR-GARCH in equation (5). The models are estimated using maximum likelihood and robust standard errors are computed as described in equation (8). Finally, k-step-ahead forecasts can be obtained from k-step-ahead forecasts of idiosyncratic and market conditional variance, both of which are given by equation (9).

2.2.3 Cholesky Factor ARMA Model

The realised covariance metrics described in Section 2.1 in effect render the covaraince matrix an observable random variable, leading to the possibility of modelling covarainces directly with well-established time series tools. However, any model must ensure covariance estimates are positive definite. In this respect, the solution of Chiriac and Voev (2007, 2008) to model the Cholesky factors is utilised, with the modification here being that an ARMA(2,1) model is estimated in place of a long memory vector fractionally integrated ARMA (VARFIMA). Chiriac and Voev (2007, 2008) argue that the long memory observed in covariance autocorrelations is preserved in the autocorrelations of the Cholesky factors, justifying the application of VARFIMA[10]. It is believed the use of an ARMA(2,1) is equally suitable, as there is considerable evidence in the volatility literature that long memory can be replicated by a two-factor model, one factor being highly persistent and the other more transient (Gallant et al., 1999; Alizadeh et al., 2002). Indeed, Pong et al. (2004) and Pong et al. (2008) have shown there is little, if any, benefit in modelling RV as an ARFIMA process over an ARMA(2,1). Moreover, the ARMA(2,1) offers a flexible short memory specification in the absence of long memory in the Cholesky factors.

More precisely, the Cholesky factors of RC on day t can be represented by,

$$RC_t^{\mathcal{E}} = \mathbf{L}_t' \mathbf{L}_t,$$

where $\mathcal{E} = \{snd, adj, sub\}$ and $\mathbf{L}_t$ is an upper triangular matrix containing the Cholesky factors of $RC_t^{\mathcal{E}}$. Since recovering the covariance matrix requires taking the product of the Cholesky factors, it is clear that modelling the Cholesky factors guarantees positive definiteness. The ARMA(2,1) model used is given by,

$$Vech\left(\mathbf{L}_t\right) = \boldsymbol{\mu} + \mathbf{\Phi}_1 Vech\left(\mathbf{L}_{t-1}\right) + \mathbf{\Phi}_2 Vech\left(\mathbf{L}_{t-2}\right) + \mathbf{\Psi} \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (10)$$

[10] See Andersen et al. (2001a) for evidence of long memory in equity correlations.

where $Vech(\cdot)$ is the column stacking operator so that $Vech(\mathbf{L}_t)$ is an $N(N+1)/2$ vector of Cholesky factors, $\boldsymbol{\mu}$ is an $N(N+1)/2$ vector of mean parameters and $\boldsymbol{\Phi}_1$,$\boldsymbol{\Phi}_2$ and $\boldsymbol{\Psi}$ are $N \times N$ diagonal parameter matrices. Hence, in (10) the Cholesky factors are modelled as univariate series. Although interaction between the dynamics of different Cholesky factors is feasible, the complexity of estimating such a model becomes prohibitively expensive and, in any case, Chiriac and Voev (2007) find relaxing the dynamics in an VARFIMA model to allow independence between the factors in all but the long memory parameter has no material effect.

Estimation of the univariate series in (10) is conducted by maximum likelihood using the error decomposition form of the log likelihood, where ε_t is assumed to be distributed as a standard normal random variable. Standard errors are computed according to (8).

2.2.4 Factor ARMA Model

Under the factor ARMA model, log returns are again assumed to follow a market model. Conditional covariances are therefore estimated by,

$$Cov_{t-1}(r_{i,t}, r_{j,t}) = \beta^{realised}_{m,i,t} \beta^{realised}_{m,j,t} V_{t-1}(r_{m,t}), \tag{11}$$

where,

$$\beta^{realised}_{m,i,t} = \frac{\sum_{t-B}^{t} RC_{m,i,t}/B}{\sum_{t-B}^{t} RV_{m,t}/B}. \tag{12}$$

Note, the $RC_{m,i,t}$ is the realised covariance between the market portfolio and the i^{th} stock on day t, $RV_{m,t}$ is the realised variance of the market portfolio on day t and B is the lag length used in computing the realised beta, $\beta^{realised}_{m,i,t}$, of stock i. To forecast covariances, only a model of $RV_{m,t}$ is required, where, as above, an ARMA(2,1) specification is used, but with a log transformation taken of $RV_{m,t}$,

$$\ln RV_{m,t} = \mu_m + \lambda_{m,1} \ln RV_{m,t-1} + \lambda_{m,2} \ln RV_{m,t-2} + \eta_m \varepsilon_{t-1} + \varepsilon_t. \tag{13}$$

By taking the log transform of RV, the necessity to place restrictions on parameters during estimation is alleviated and results in a distribution closer to the normal improving parameter estimates and forecasts (see Andersen et al., 2001a). Estimation proceeds as outlined in Section 2.2.3.

However, because Andersen et al. (2001a) show accurate variance forecasts can be obtained by directly applying ARMA models to RV, the variances of individual stocks are also forecasted using (13), with the RV of the individual stocks, $RV_{i,t}$, substituted in for $RV_{m,t}$. To incorporate the information contained in these forecasts and maintain the positive semi-definiteness of the covariance matrices constructed from (11), the decomposition in (4) is used. Correlation matrices are formed from the forecasts based on (11) and then pre- and post-multiplied by diagonal matrices with the square roots of the individual variance forecasts comprising the diagonal elements.

2.2.5 Option-Factor-Implied Covariance

Currently, it is not possible to estimate a model-free option-implied covariance for equities. However, recently Chang et al. (2009) have shown option-implied betas may be extracted. Employing the assumption that returns follow the market model then allows covariances to be estimated from,

$$IC_{i,j} = \beta_i^{imp} \beta_j^{imp} IV_m, \quad \text{for } i \neq j,$$

where $IC_{i,j}$ is the implied covariance estimate between assets i and j, β_i^{imp} and β_j^{imp} are the option-implied betas for stocks i and j, respectively, and IV_m is the option-implied market variance. Since the covariances cannot be estimated directly from option prices, these covariance estimates will be referred to as option-factor-implied covariances.

The option-implied betas of Chang et al. (2009) are given by,

$$\beta_i^{imp} = \left(\frac{SKEW_i}{SKEW_m}\right)^{\frac{1}{3}} \left(\frac{VAR_i}{VAR_m}\right)^{\frac{1}{2}},$$

where $SKEW_i$ and $SKEW_m$ are, respectively, the risk-neutral implied skewness of asset i and the market and VAR_i and VAR_m are, respectively, the risk-neutral implied kurtosis of asset i and the market. Bakshi et al. (2003) show risk-neutral implied skewness and variance can be obtained from[11],

$$\begin{aligned} SKEW &= \frac{e^{r\tau}Cubic - 3\mathbb{E}^Q[R]e^{r\tau}Quad + 2\mathbb{E}^Q[R]^3}{VAR^{\frac{3}{2}}}, \\ VAR &= e^{r\tau}Quad - \mathbb{E}^Q[R]^2, \end{aligned}$$

where r is the risk-free rate, τ is the time to maturity, R is the log return over the period τ and $\mathbb{E}^Q[\cdot]$ is the expectation operator under the risk neutral measure. The *Cubic* and *Quad* terms refer to the discounted risk-neutral expectations of squared and cubed returns over the period τ. These can be computed from,

$$\begin{aligned} Quad &= \int_S^\infty \frac{2(1 - \ln\left[\frac{K}{S}\right])}{K^2} C(\tau, K)dK + \int_0^S \frac{2(1 + \ln\left[\frac{K}{S}\right])}{K^2} P(\tau, K)dK, \\ Cubic &= \int_S^\infty \frac{6\ln\left[\frac{K}{S}\right] - 3\ln\left[\frac{K}{S}\right]^2}{K^2} C(\tau, K)dK + \int_0^S \frac{6\ln\left[\frac{K}{S}\right] - 3\ln\left[\frac{K}{S}\right]^2}{K^2} P(\tau, K)dK, \end{aligned}$$

where $C(\tau, K)$ and $P(\tau, K)$ are the market prices of European calls and puts with strike prices K expiring in τ periods. Lastly, the risk-neutral expected return can be approximated by,

$$\mathbb{E}^Q[R] = e^{r\tau} - 1 - \frac{e^{r\tau}}{2}Quad - \frac{e^{r\tau}}{2}Cubic.$$

It can be seen that to compute the value of *Quad* and *Cubic* and hence the risk-neutral implied moments, a continuum of option prices are required. In order to obtain this continuum, a quadratic function is fitted to the implied volatility smile; the U-shape, or, more frequently for equities, smirk-shape, observed when implied volatilities are plotted against their strike prices. Following Malz (1997a,b) and Taylor et al. (2009), the function is estimated in implied volatility-delta space, instead of implied volatility-moneyness space, where a quadratic function appears appropriate. Given the limited number of strike prices traded on individual stock options and that the most liquid options are those with strike prices close to or ATM, this approach has the advantage of requiring a minimum of 3 observed option prices and placing the largest weight on options with ATM strikes. This can be seen by looking at the expression for the delta on the x^{th} strike price of asset i, $\delta_{x,i} = \mathcal{N}(d1_{x,i})$, where $\mathcal{N}(\cdot)$ is the cumulative normal ditribution function and,

$$d1_{x,i} = \frac{\ln\frac{S_i}{K_{x,i}} + (r + \frac{\bar{\sigma}_i^2}{2})\tau}{\sqrt{\bar{\sigma}_i\tau}},$$

where S_i is the spot price and $\bar{\sigma}_i$ the ATM implied volatility of the i^{th} underlying asset option. Most of the dispersion in $\delta_{x,i}$ is generated by options with strikes close to or ATM, meaning less

[11] Note, the subscripts identifying the individual stocks and the market are dropped from this point.

liquid out-of-the-money (OTM) option prices will exert less influence on the estimated quadratic function[12]. To be precise, the following weighted least squares regression is fitted to the implied volatility,

$$w_{x,i}\sigma_{x,i} \quad = \quad w_{x,i}a + w_{x,i}b\delta_{x,i} + w_{x,i}c\delta_{x,i}^2 + w_{x,i}\epsilon, \tag{14}$$

where $\sigma_{x,i}$ is the implied volatility of the x^{th} strike price, w_x is the weight applied in the case of the x^{th} strike price, a, b and c are parameters to be estimated and ϵ represents the error in estimation. Hence, the influence of OTM options are mitigated further by using weighted least squares, where weights are derived from each option's vega, ν_x,

$$w_{x,i} \quad = \quad \frac{\nu_{x,i}}{\sum_{x=1}^{X}\nu_{x,i}},$$

and

$$\nu_{x,i} \quad = \quad S_i\sqrt{\tau}\frac{1}{\sqrt{2\pi}}\exp\{-d1_{x,i}^2/2\}.$$

The expiry dates of exchange traded options are set according to the January, February and March expiration cylcles. This means option expiry dates do not often coincide with the horizon over which implied betas and variances would like to be computed. To estimate their values over the desired horizon, implied betas and variances are computed for option expiry dates immediately preceding and following the date marking the forecast horizon and linearly interpolated. Finally, the covariance decomposition in (4) is used once again to incorporate the information contained in the implied variances computed for individual stocks. The procedure is identical to that used for the Factor ARMA model in Section 2.2.4.

2.2.6 Historical Covariances

In addition to the dynamic (DCC, Factor Double ARCH, CF-ARMA and Factor ARMA) models and option-factor-implied forecasts, covariance forecasts based on historical measures are included as benchmarks. Both a historical high-frequency based covariance forecast, $\widehat{\Sigma}_{i,j,t}^{HF}$, and low-frequency, daily return based, covariance forecast, $\widehat{\Sigma}_{i,j,t}^{LF}$, are used,

$$\widehat{\Sigma}_{i,j,t}^{HF} = \frac{1}{k}\sum_{s=t-k+1}^{t} RC_{i,j,s}^{sub}$$

$$\widehat{\Sigma}_{i,j,t}^{LF} = \frac{1}{k}\sum_{s=t-k+1}^{t} r_{i,t}r_{j,t},$$

where k is the forecast horizon, except in the 1-day-ahead case for $\widehat{\Sigma}_{i,j,t}^{LF}$, where $k = 30$ to prevent the covariance matrix from being singular.

2.3 Forecast Evaluation

To compare the relative merits of the alternative covariance matrix forecasts, three methods of forecast evaluation are employed: Firstly, direct comparisons are made using the MSE loss function;

[12] The ATM implied volatility is used to standardise the computation of δ. In application, $\bar{\sigma}$ is set to the implied volatility of the option whose strike price is closest to being ATM.

secondly, the optimality of the forecasts are examined using Mincer-Zarnowitz regressions; and thirdly, an economic loss function is employed. By using three evaluation techniques, a broader understanding of the forecasting performance of the seven models is constructed. Before providing the detail pertaining to each of these evaluation techniques, the precise construction of the forecast target is outlined.

2.3.1 Forecast Target

For 1-day ahead forecast evaluation, the proxy used to represent the true covariance matrix is the realised covariance matrix (RCM), $\mathbf{\Sigma}_{t+1}$, where forecasts are made at time t and each element comprises the $RC^{\rceil}_{t+1}$ estimates. In the multiperiod (30-day, 90-day and 180-day) forecasting exercises the forecast target becomes,

$$\bar{\mathbf{\Sigma}}_{t+k} = \frac{\sum_{s=t+1}^{t+k} \mathbf{\Sigma}_{t+s}}{k}, \tag{15}$$

where k is the forecast horizon. Similarly, if $\mathbf{H}_{t+1}^{\mathcal{M}_u}$ is the 1-day ahead forecasted covariance matrix at date t produced by the model $\mathcal{M}_u$, where,

$$\begin{aligned} \mathcal{M}_u \quad &= \quad \{\text{DCC}, \text{Factor Double ARCH}, \text{CF-ARMA}, \text{Factor ARMA}, \\ &\quad\quad \text{option-factor-implied}, \widehat{\Sigma}_{i,j,t}^{LF}, \widehat{\Sigma}_{i,j,t}^{LF}\Big\} \end{aligned}$$

and $u = 1, \ldots, 5$, ($\widehat{\Sigma}_{i,j,t}^{LF}$ and $\widehat{\Sigma}_{i,j,t}^{HF}$ estimators are excluded), then the k-period-ahead covariance matrix forecast is given by,

$$\bar{\mathbf{H}}_{t+k}^{\mathcal{M}_u} = \frac{\sum_{s=t+1}^{t+k} \mathbf{H}_{t+s}^{\mathcal{M}_u}}{k}. \tag{16}$$

Thus, the forecasting exercises assess the ability of each model to accurately forecast the mean daily covariance matrix over each horizon.

2.3.2 Statistical Loss Functions

To conduct direct comparisons of the forecasting accuracy of the different models, the mean squared error (MSE) loss function is employed, chosen because of its robustness to the proxy of the true covariance matrix (Patton (2008) and Patton and Sheppard (2009)). That is, the ranking of alternative forecasts is identical whether the true covariance matrix or a noisy, unbiased, proxy is used as the forecast target. In the context of forecasting the covariance matrix, the multivariate MSE (MMSE) is the appropriate loss function, summarising forecasting accuracy in a single statistic. However, the univariate MSE (UMSE) is also applied to the individual elements of the covariance matrix to provide a contrasting analysis that is more intuitive. If a model is found to provide superior forecasts according to the MMSE then it seems natural that the model should provide superior forecasts for each element of the covariance matrix. The robust MMSE, $\bar{\mathcal{L}}_{\mathcal{M}_u}^{MMSE}$, defined in Patton and Sheppard (2009) is given by,

$$\begin{aligned} \mathcal{L}_{\mathcal{M}_u,t}^{MMSE} \quad &= \quad tr\left(\bar{\Sigma}_t^2 - (H_t^{\bar{\mathcal{M}}_u})^2\right) - tr\left(H_t^{\bar{\mathcal{M}}_u}\left(\bar{\Sigma}_t - H_t^{\bar{\mathcal{M}}_u}\right)\right), \\ \bar{\mathcal{L}}_{\mathcal{M}_u}^{MMSE} \quad &= \quad \frac{1}{T}\sum_{t=1}^{T} L_{\mathcal{M}_u,t}^{MMSE}, \end{aligned} \tag{17}$$

where $tr(\cdot)$ is the trace operator. It should be noted that the MMSE in (18) is normalised to take the value 0 when the true and forecasted covariance matrices coincide. The UMSE for an element of the covariance matrix takes the familiar form,

$$\begin{aligned} \mathcal{L}^{UMSE}_{\mathcal{M}_u,t} &= \left(RC^{\mathcal{P}_v}_{i,j,t} - h^{\mathcal{M}_u}_{i,j,t} \right), \\ \bar{\mathcal{L}}^{UMSE}_{\mathcal{M}_u} &= \frac{1}{T} \sum_{t=1}^{T} L^{UMSE}_{\mathcal{M}_u,t}, \end{aligned} \tag{18}$$

where $h^{\mathcal{M}_u}_{i,j,t}$ is the $(i,j)^{\text{th}}$ element of $\bar{\mathbf{H}}^{\mathcal{M}_u}_t$.

The QLIKE loss function is robust and has been shown by Patton and Sheppard (2009) to be more powerful than the MSE in simulations. However, its multivariate form relies on the covariance matrix being positive definite and its univariate form on the forecast and target being positive. Since the Factor ARMA and option-factor-implied forecasts are restricted to the set of positive semi-definite matrices and covariance terms are defined over $\mathbb{R}$ and not $\mathbb{R}^{++}$, the QLIKE loss function was excluded from the forecast evaluations.

The model confidence set (MCS), introduced by Hansen et al. (2009), is used to determine whether a significant difference exists between the performance of the forecasts produced by the alternative models. The principle behind MCS-test is to extract a subset of superior models, $\mathcal{M}^* = \{\mathcal{M}_1, \ldots, \mathcal{M}_{U^*}\}$, from a larger set of competing models, $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_U\}$, where $U^* \leq U$, by iterating through an equivalence test and elimination rule. At each iteration the equivalence test evaluates the null hypothesis $H_0 : \mathcal{M}_u \in \mathcal{M}^*$ for $u = 1, \ldots, U$ and, if rejected, a model, selected by the elimination rule, is removed from $\mathcal{M}^*$.

There are two advantages associated with this approach. Firstly, unlike the Diebold-Mariano-West test (Diebold and Mariano (1995) and West (1996)), it can be used to compare the performance of multiple models. Secondly, there is no requirement for a benchmark model to be preselected. Therefore, although the reality check and superior predictive ability (SPA) tests of White (2000) and Hansen (2005), respectively, are able to compare multiple models, they assess the performance of the models relative to a predetermined benchmark.

In evaluating the seven models of Section 2.2, t-tests of the differences in loss resulting from each model are used to form the basis of the equivalence test. The t-statistics are computed from[13],

$$\begin{aligned} d_{u,v,t} &= L_{\mathcal{M}_u,t} - L_{\mathcal{M}_v,t}, \quad \text{for } v \neq u, \\ \bar{d}_{u,v,t} &= \frac{1}{T} \sum_{t=1}^{T} d_{u,v,t}, \\ \bar{d}_{u,t} &= \frac{1}{U} \sum_{w} \bar{d}_{u,v,t}, \quad \text{for } v \neq u, \\ t_u &= \frac{\bar{d}_{u,t}}{\sqrt{\widehat{var}(\bar{d}_{u,t})}}, \quad \forall u. \end{aligned} \tag{19}$$

If the loss function is MMSE, the equivalence test is conducted on,

$$T_{\max} = \max_u |t_u|. \tag{20}$$

Absolute values are taken to account for the two-sided nature of the MMSE, which can, counter-intuitively, take both positive and negative values. For the UMSE loss function no absolute values

[13]Note, the superscript has been dropped from $L_{\mathcal{M}_u,t}$ to indicate that (20) can be computed using either $L^{MMSE}_{\mathcal{M}_u,t}$ or $L^{UMSE}_{\mathcal{M}_u,t}$.

are taken so that the equivalence test is based on,

$$T_{\max} = \max_u t_u. \tag{21}$$

There is no asymptotic sampling theory for (20) or (21). Consequently, the estimator $\sqrt{\widehat{var}(\bar{d}_{u,t})}$ and p-values for (20) and (21) are computed via the block bootstrap, the algorithm for which can be found in the appendix of Hansen et al. (2009). Block sizes of 10 were used for 1-day and 30-day ahead forecasts whilst 90-day and 180-day ahead forecasts were evaluated with block sizes of 15 and 20, respectively.

2.3.3 Optimality Tests

It is common in the volatility forecasting literature to test the performance of a forecasts by Mincer-Zarnowitz regressions (Diebold and Lopez (1996) and Patton and Sheppard (2009)). These involve regressing a proxy for the true volatiliy against a constant and the volatility forecast. A forecast is then said to be unbiased if the constant is estimated to be zero and the slope coefficient is one. The R^2 values resulting from these regressions are also commonly used to compare the information content of the models, with higher R^2 values indicating a higher information content, and rank the models. However, as highlighted in Taylor (Chap. 15 2005), the ranking based on R^2 may differ to that based on a direct comparison of the forecasts, even if the MSE loss function is employed, which is the loss function implicitly assumed in the R^2. This is because the R^2 is comparing the performance of the forecasts after they have been projected onto the forecast target.

There is no direct generalisation of the univariate Mincer-Zarnowitz regressions to the multivariate framework. The approach taken here is to apply univariate Mincer-Zarnowitz regressions to the forecasts of each element of the covariance matrix,

$$Vech\left(\Delta\boldsymbol{\Sigma}_t\right)_i = a_i + b_i Vech\left(\Delta\bar{\mathbf{H}}_t^{\mathcal{M}_u}\right)_i + e_{i,t}, \tag{22}$$

where,

$$\Delta\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t - \bar{\mathbf{H}}_{t-1}^{\mathcal{M}_u}, \tag{23}$$

$$\Delta\bar{\mathbf{H}}_t^{\mathcal{M}_u} = \bar{\mathbf{H}}_t^{\mathcal{M}_u} - \bar{\mathbf{H}}_{t-1}^{\mathcal{M}_u}, \tag{24}$$

where $Vech(\cdot)_i$ is the i^{th} element of the vector resulting from the application of the $Vech(\cdot)$ operator[14]. Following Fleming (1998), who found it important when evaluating volatility forecasts, first differences of the covariance matrix time series are taken to remove the effects of spurious regression that arises from serial correlation induced by overlapping forecasts. Although, in the context of volatility forecasting, Christensen and Prabhala (1998) suggest using non-overlapping forecasts to erradicate the possibility of spurious regression, this would leave too few observations in our sample to make meaningful inferences.

The regressions in (22) are estimated using GMM which is capable of handling heteroscedasticity and autocorrelation in the errors. Formally, the moment conditions are defined by,

$$\mathbf{f}_{i,t}(\boldsymbol{\theta}) = \mathbf{h}_{i,t} \otimes e_{i,t}(\boldsymbol{\theta})$$

where $e_{i,t}$ is the error from the regression (22), $\mathbf{h}_{i,t}$ is a vector containing instruments and $\boldsymbol{\theta} = \{a_i, b_i\}$. To estimate the parameters in (22) by minimising the squared errors the instruments include,

$$\mathbf{h}_{i,t} = \left(1, \bar{\mathbf{H}}_t^{\mathcal{M}_u}\right)_i.$$

[14]An alternative to the generalisation chosen here is $Vech\left(\Delta\boldsymbol{\Sigma}_t\right)_i = a + bVech\left(\Delta\bar{\mathbf{H}}_t^{\mathcal{M}_u}\right)_i + e_{i,t}$, where a and b are scalars. However, the dimensions invovled (applying the $Vech(\cdot)$ operator to the 17×17 covariance matrix results in a vector with 153 elements) means parameter estimation becomes infeasible.

The parameters, $\boldsymbol{\theta}$, are then estimated by minimising the quadratic function,

$$\mathbf{Q}_{i,T}(\boldsymbol{\theta}) \quad = \quad \mathbf{g}_{i,T}(\boldsymbol{\theta})'\mathbf{W}_{i,T}\mathbf{g}_{i,T}(\boldsymbol{\theta}),$$

where,

$$\mathbf{g}_{i,T}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{i,t}.$$

A two-step estimator is emloyed, where $\mathbf{W}_{i,T}$ is set equal to the identity matrix in the first step and to the inverse of the estimated spectral density in the second step. A Newey-West estimator is used to estimate the spectral density in the second step to account for heteroscedasticity and autocorrelation in the erros. Finally, the joint null hypothesis, H_0: $a_i = 0 \cap b_i = 1$, is tested using Wald statistics[15].

Thus far, the methods outlined assess the ability of each model to accurately forecast the conditional covariance matrix. Depending upon the loss function of the users of the covariance forecasts, the conditional covariance matrix may not be the appropriate forecast target (Patton and Timmermann, 2007). An alternative criterion that may be used to test the optimality of forecasts is whether the information contained in one forecast encompasses the information in another, or, in other words, whether the information contained in one forecast is orthogonal to the information contained in the competing forecasts (Diebold and Lopez, 1996).

Again highlighted by Fleming (1998), the GMM estimation methodology enables the orthogonality of alternative forecasts to be assessed, accomplished by augmenting the set of instruments used to define $\mathbf{f}_{i,t}(\theta)$. To test the orthogonality of a covariance forecast generated by model $\mathcal{M}_v$ to the forecast generated by model $\mathcal{M}_v$, the set of instruments are augmented to include,

$$\mathbf{h}_{i,t} \quad = \quad \left\{\boldsymbol{\iota}, Vech\left(\Delta\bar{\mathbf{H}}_t^{\mathcal{M}_u}\right)_i, Vech\left(\Delta\bar{\mathbf{H}}_t^{\mathcal{M}_v}\right)_i\right\}. \tag{25}$$

Since the number of moment conditions now exceeds the number of parameters to be estimated, the model is overidentified and orthogonality can be assessed by testing the overidentifying restrictions using the J-statistic,

$$TJ_{i,T} \quad = \quad T\mathbf{Q}_{i,T}(\theta) \sim \chi_1^2. \tag{26}$$

The degrees of freedom in the χ^2 distribution are set equal to the number of moment conditions minus the number of parameters estimated. It should also be noted that to conduct these tests the regression in (22) was run without conducting the differencing in (23)-(24) due to the J-statistic being unaffected by the spurious regression problem.

2.3.4 Economic Loss Function

An alternative method to evaluate forecasts is to employ an economic loss function. In this respect, the procedure outlined in Engle and Colacito (2006) is adopted. They show that the variance of a mean-variance optimised portfolio is minimised only when the true covariance matrix is used to compute weights. The same criterion can be used to compare the accuracy of two covariance estimates: The estimate providing the lowest portfolio variance can be considered to be more accurate. To formally test for any significant difference in the portfolio variances, and, hence, the accuracy of the covariance estimates, DM tests may then be employed.

[15]Formulae for the Newey-West estimator and the standard errors of the estimated parameters are commonplace and can be found in Campbell et al. (1997) for example.

The optimal portfolios are computed following the classical mean-variance portfolio allocation problem,

$$\min_{\mathbf{w_t}} \mathbf{w_t'} \mathbf{H_t} \mathbf{w_t}$$
$$\text{s.t.} \mathbf{w_t} \boldsymbol{\mu} = \mu_0,$$

where $\mathbf{H_t}$ is a covariance matrix estimate, $\mathbf{w_t}$ are the weights assigned to the assests in the portfolio, $\boldsymbol{\mu}$ is a vector of expected excess returns and μ_0 is the required rate of return on the portfolio. The solution is well-known and given by,

$$\mathbf{w_t} \quad = \quad \frac{\mathbf{H_t^{-1}} \boldsymbol{\mu}}{\boldsymbol{\mu}' \mathbf{H_t^{-1}} \boldsymbol{\mu}} \mu_0.$$

The weights are not constrained to sum to 1 so that a weight of $1 - \sum_{i=1}^{N} w_i$ is assigned to the risk-free asset. This framework has been frequently used to assess alternative covariance estimates, but is usually restricted to examining the minimum variance portfolio case. By limiting the evaluation to the minimum variance portfolio, it is implicitly assumed that all assets have identical expected returns. In addition to this being an unsatisfactory assumption, it is also understood that correct expected return estimates are the most important component in forming optimal portfolios(Chopra and Ziemba, 1993). Engle and Colacito (2006) show that a more accurate covariance estimate leads to a lower portfolio variance for all possible vectors of expected returns and therefore that the variance of the optimal portfolio should be examined for a range of expected return vectors. In cases where a large number of assets are being used, the number of combinations of expected returns would become unmanageable. Therefore, in these circumstances, it is recommended that only hedging portfolios are considered and this approach is adopted here. These are formed by setting the value of the expected return vector corresponding to the hedged asset to 1 and the expected returns of all other assets to zero. Consequently, with N assets there will be N expected return vectors.

Although it is common to estimate the optimal portfolio variance by squaring the portfolio return, the long forecast horizons of 30-, 90- and 180-days would result in very noisy estimates. To minimise this noise, the RV of each hedging portfolio is computed. Comparisons between pairs of portfolio RVs computed using different covariance estimates are then made by forming a series of differences in the RVs for each hedging portfolio,

$$z_t^k \quad = \quad RV_{\mathcal{M}_u,n,t}^{hp} - RV_{\mathcal{M}_v,n,t}^{hp}, \quad \text{for } u \neq v,$$

where $RV_{\mathcal{M}_u,n,t}^{hp}$ is the RV of the n^{th} hedging portfolio obtained using model $\mathcal{M}_u$. Whether there is a significant difference between the RVs can be tested jointly by running the regression,

$$\mathbf{Z}_t \quad = \quad \alpha_{hp} \boldsymbol{\iota} + \boldsymbol{\epsilon}_{hp,t}, \tag{27}$$

where $\mathbf{Z}_t = (z_t^1, z_t^2, \ldots, z_t^k)$ and $\boldsymbol{\iota}$ is $N \times 1$ vector of ones. If α_{hp} is found to be negative (positive) and significantly different to 0, then it can be concluded that the covariance forecast of model $\mathcal{M}_u$ ($\mathcal{M}_v$) is superior. Due to the 30-, 90- and 180- day forecast horizons overlapping and the resulting autocorrelation induced in the errors, the regression in (27) is estimated by GMM using a Newey-West heteroscedasticity and autocorrelation consistent error covariance matrix estimate. In addition to pairwise comparisons, the MCS-test was also applied.

Note that in order to keep the analysis manageable, the economic loss function is only assessed for out-of-sample forecasts. This also comprises the economically relevant period of the data.

3 Data

A total of 17 stocks included in the Dow Jones Industrial Average (DJIA) over the sample period from January 2002 to September 2008 were selected for the covariance matrix forecast comparisons. Constituents of the DJIA were focused on to ensure sufficient liquidity existed in both the stock and option markets for RC and risk-neutral moments to be computed accurately. Nevertheless, a lack of liquidity and the constraint $N \leq M$, imposed by the sampling frequency of intraday returns, where the selection of M is discussed below, prevented all constituents of the DJIA being included in the sample. In order to estimate models based on a market model assumption, the S&P 500 index was also included in the sample as the proxy for the market portfolio. Intraday prices were obtained from the TAQ database and daily prices from CRSP.

3.1 Stock Price Data

A list of the 17 companies and their associated tickers is provided in Table 1. Closing prices from CRSP, including the S&P index level, were used as daily price observations whereas the intraday record of transaction and quote prices comprised the high-frequency price observations. For a high-frequency proxy of the S&P 500 index an exchange traded fund (ETF), the Spider, which replicates the S&P 500 index, was selected from the TAQ database.

3.1.1 Data Filtering

Before using transaction or quote prices in the computation of returns, a number of filtering rules, similar to those in Barndorff-Nielsen et al. (2008b), were applied[16]. In all cases prices were sampled between 09:35:00 Eastern standard time (EST) and 16:00:00 EST. The first 5 minutes after the open are ignored due to the distortions introduced by the market opening procedure. Any prices (transaction or quote) taking a zero value were deleted and series of prices having an identical time stamp were replaced by the mean of the series. Lastly, the price filter suggested by Brownless and Gallo (2006) was implemented on transaction and mid-quote prices. Formally, this can be represented by,

$$|\Delta p_{t,j,q}| > 3s(\Delta p_{t,j,q}, n) + g \begin{cases} 1 & \text{Delete} \\ 0 & \text{Retain} \end{cases}, \tag{28}$$

where $|\Delta p_{t,j,q}| = |p_{t,j,q} - p_{t,j-1,q}|$, g is a granularity parameter and $s(\Delta p_t, n)$ is the standard deviation of price changes computed over sample spanning n observations centered on p_t. In application, n was set to 60 and g to 2. Therefore, a price was deleted if the absolute change in its value deviated by more than 3 standard deviations and 2 cents, computed from a sample, centered on the price being interrogated, of 60 price change observations[17].

For quote prices, additional filters included the removal of any observation for which the spread between the bid and ask prices was negative or was greater than 50 times the median spread for the associated day. Also, only quotes originating from the security's primary exchange, where the stock was listed, were retained (Table 1) to minimise the impact of different market structures. For

[16]The importance of filtering high-frequency data for aberrant prices is outlined in Dacarogna et al. (2001) and Falkenberry (2002)

[17]At the beginning and end of a sample for any given day, it is not possible to center the computation of the standard deviation around 60 price observations. Instead, at the beginning of a sample day, 60 lead prices were used and at the end of the sample 60 lag prices were used.

all but 2 of the sample stocks, the largest number of quotes were on the primary exchanges[18]. Supplementary filters for transaction prices included the deletion of any prices flagged to be incorrect by the CORR and COND variables, although trades from all exchanges were retained.

3.1.2 Preliminary Data Analysis

The effect of market microstructure frictions and the multiple approaches to computing RC means there are many alternative ways of applying the metric. Decisions that must be made relate to the choice between transaction and quote data, previous tick (PT) and linear interpolation (LI) imputation methods, the frequency at which to sample prices and the precise RC estimator to implement. Although Martens (2004) has examined some of these issues in both a simulation and empirical study, there is no set of universal guidelines and they are, therefore, examined here.

The RCs are computed using returns from prices imputed according to both PT and LI methods and for both transaction and quote data, where mid-quotes were used. To investigate the impact of the Epps effect (Epps, 1979), which potentially causes a severe downward bias, and other market microstructure frictions, the RC_t^{adj} and RC_t^{snd} estimators were applied using 7 sampling intervals: 30 seconds and 1, 5, 11, 17.5, 35 and 55 minutes, corresponding to values of M of 770, 385, 77, 35, 22, 11, and 7, respectively. The RC_t^{sub} estimator was applied at 5 of the sample intervals: the 5 minute grid was subsampled at 1 minute intervals, producing 5 subsamples; the 11 minute grid at 1 minute intervals, producing 11 subsamples; the 17.5 min grid at 2.5 min intervals, producing 8 subsamples, the 35 min grid at 5 minute intervals, producing 7 subsamples and the 55 minute grid at 5 minute intervals, producing 11 subsamples.

To make comparisons of the estimators across frequencies and datasets, signature plots, introduced by Andersen et al. (2000), are compared. The signature plots in Fig. 1 graph the mean RC at different sampling intervals, where, in order to keeep the computational burden reasonable, the mean RCs have been averaged across all stocks in the covariance matrix. These values along with their standard deviaitons are also summarised in Table 2.

Fig. 1 and Table 2 reveal that the different types of RC estimator are associated with Epps effects of differing severity. For RC_t^{snd} in Fig. 1a the Epps effect is clear and occurs for RCs measured below 5 mins, whilst for RC_t^{adj} it manifests at a sampling interval below 1 min. There is no visible Epps effect for RC_t^{sub} because this estimator was not examined for sampling intervals below 5 mins. It is also important to note that the RC_t^{sub} estimator is associated with less sampling variation (see Table 2 and note the different y-axis scales used in Fig. 1). The different data types appear to be associated with different levels of sampling variation. Table 2 shows that for each estimator the RCs based on transaction data have a lower standard deviation than their counterparts estimated from mid-quotes. However, the use of LI or PT imputation methods appears to have little influence on the estimators.

Based on the preceding analysis, RC will be measured with the RC_t^{sub} estimator using transaction data linearly interpolated onto a 17.5 min sampling grid in the empirical applications. Transaction data is selected for its reliability. Although not reported, there is a substantially larger number of days on which quote data is missing compared to transaction data. The 17.5 min grid is felt to provide an adequate trade-off between the efficiency of the RC estimate, theoretically achieved through sampling returns over the shortest possible time interval, and the introduciton of microstructure noise in form of bid-ask bounce and the Epps effect, which increase in severity as the sampling

[18]The primary exchange for the SPY is the American Stock Exchange and for IBM it is the New York Stock Exchange. However, the largest number of quotes were found on the NASDAQ for both securities.

interval decreases. The RC_t^{sub} estimator with LI imputation are selected to minimise sampling variation.

3.2 Option Data

Option data was obtained from the Optionmetrics IvyDB database. To select observations, several filters were applied. For an observation to be retained the option had to trade on the observation date, the open interest had to be greater than zero, the bid price had to exceed the ask price, both bid and ask prices had to be greater than zero and less than the underlying asset's price and it had to be OTM. A further criterion was that three strikes had to remain after applying the previous filters, including at least one call and one put.

Implied volatilities supplied by IvyDB were used in fitting the quadratic function in (14). Although options on individual stocks have American payoff functions, their implied volatilities have been demonstrated to lie close to those from equivalent European options, with differences often contained within the variation caused by evaluating implied volatilities at either bid or offer quotes (Bakshi et al. (2003); Jiang and Tian (2005)). When converting implied volatilities back to option prices, discounted dividends were first subtracted from the underlying price allowing the BS formula without dividends to be employed. Risk free interest rates were also obtained from the IvyDB.

To compute *Quad* and *Cubic* required numerical integration, conducted according to Simpson's rule with 2000 interpolation points [19]. Due to liquidity, averages of VAR and $SKEW$ were taken each week, where weeks were defined to start and end on Thursdays and Wednesdays, respectively, limiting the production of forecasts by the option-factor-implied model to a weekly frequency[20].

4 In-Sample Results

The in-sample period covers 02/01/2002 to 04/01/2006 and provides between 1000 and 1085 days for parameter estimation, depending on whether the data is low frequency daily or high-frequency intraday data. To ensure the dates on which forecasts are made are consistent, multiperiod forecasts are constructed on the same dates as the option-factor-implied model and therefore occur with a weekly frequency. In the following subsections the in-sample forecasting performance of the models is assessed, along with the impact of including or excluding overnight returns.

4.1 1-Day-Ahead Forecasts

4.1.1 Parameter Estimates

For the models based on daily data, Tables 3 and 4 contain the estimated parameters of the DCC and Factor Double ARCH models, with standard errors in parentheses. Most parameters in the univariate GJR-GARCH models of the DCC specification are significant and conform to the well-established findings for equity markets. The positive and significant values for γ_i across all stocks are consistent with the leverage effect being pervasive in individual equity prices and the level of

[19] This number of interpolation points is the same as in Chang et al. (2009)

[20] Weekly averages of VAR and $SKEW$ have been taken by Bakshi et al. (2003), Jiang and Tian (2005) and Taylor et al. (2009) to address liquidity.

persistence is very high across all stocks, with $\alpha_i + \frac{1}{2}\gamma_i + \beta_i$ taking an average value of 0.992. The correlation parameters α_q and β_q are both significant and indicate strong persistence in correlation, their sum being 0.999. Similar results are found for the GJR-GARCH parameters of the Factor Double ARCH model in Table 4 demonstrating the GJR-GARCH to be successful in modelling idiosyncratic return variances. Also note the estimated values of $\beta_{m,i}^{ARCH}$ are all close to the value 1, which is expected given that the stocks are for large companies in the DJIA whose systematic risk will mimic market risk.

Turning to the models based on high-frequency data, parameter estimates for the CF-ARMA and Factor ARMA models are summarised in Tables 5 and 6. The parameter estimates of the ARMA(2,1) model in the Factor ARMA framework are similar across all stocks. The parameter on the AR(1) term is large relative to that on the AR(2), often taking a value close to one. Although the parameters on the AR(2) term are small, they are significant in all cases except for HD and MSFT. The paramters on the MA(1) term show less variation, all being negative and taking an average value of -0.7496. It is noteworthy that the parameters pertaining to SPY, the market proxy, differ slightly to those estimated for individual stocks. The parameter on the AR(1) (AR(2)) term is lower (higher) in comparison to the estimates for the individual stocks and the parameter on the MA(1) term is higher[21]. The mean values of $\beta_{m,i,t}^{realised}$ are listed in the table for each stock and can be seen to be comparable to the values estimated under the Factor double ARCH model, where it should be noted that $B = 1$ in equation (12).

Only the mean, minimum and maximum values of the the ARMA(2,1) parameter estimates for the CF-ARMA model are shown in Table 5 due to the large number, 153, of Cholesky factors that have to be modelled. The mean values given in Table 5 show the parameters continue to conform to the regular pattern discussed for the Factor ARMA model; AR(1) parameters are large compared to AR(2) parameters and frequently take a value close to 1.

Due to the limitations of the data, the option-factor-implied forecasts are computed each week and remain fixed until the following week at which point the forecasts are updated. Also, as a substitute for 1-day-ahead forecasts, the 30-day-ahead forecasts will be used, which provide predictions concerning the mean level of variances and covariances over this horizon. Although these forecasts should not challenge to the competing models in terms of accuracy, they do provide a useful yardstick to highlight the benefits of modelling the dynamics of the variance-covariance matrix over a more *ad hoc* approach.

4.1.2 Statistical Loss Function Assessment

The results from analysing the MMSE for the 1-day ahead horizon are summarised in Panel A of Table 8. The p-values represent the probability of each model being included in the MCS. The option-factor-implied forecasts are clearly rejected, but there is no evidence that the other models should be excluded from the MCS. The elimination of the option-factor-implied forecasts from the MCS is not surprising given the *ad hoc* nature of their computation. The absence of a significant difference between the forecasting performances of the remaining models indicates the accuracy of their forecasts can be considered equivalent.

The nature of the MMSE means it is difficult to, informally, rank the models. An MMSE of 0 indicates the forecast perfectly represents the true (proxy) covariance matrix, but the difference between 0 and the MMSE does not help rank the models. For example, the difference between 0 and the mean MMSE for the option-factor-implied forecasts is not as great as the difference

[21]Although Pong et al. (2004) and Pong et al. (2008) estimate ARMA(2,1) models on currencies, the estimates found here for equities are similar.

for the CF-ARMA and Factor ARMA models, yet these models are contained within the MCS. An analogous situation holds for the maximum and minimum MMSE. Although the option-factor-implied forecasts do produce the largest maximum MMSE, they do not produce the most negative value. Thus, there is no intuitive interpreatation of the MMSE.

To provide a robustness check, the MCS is additionally applied to the UMSEs computed for the forecasts of each element of the covariance matrix. Panel A of Table 9 summarises the results, where three averaged MSE statistics are listed for each model. The first is an average across all covariance and variance series (MSE-VC), the second is an average across covariances (MSE-C) and the third is an average across variances (MSE-V). The MCS-test was also applied to the forecasts of each element of the covariance matrix. The last three columns provide the number of individual covariance matrix element forecasts that were eliminated from the MCSs at the 5%, 10% and 15% significance levels, respectively, for each model.

The results show a large majority of the option-factor-implied and $\widehat{\Sigma_t^{LF}}$ forecasts are excluded from the MCS at all significance levels. The next largest number of exclusions are from the CF-ARMA model followed by the DCC, Factor Double ARCH, Factor ARMA and $\widehat{\Sigma_t^{HF}}$, although the actual number of rejections are substantially lower in comparison to the option-factor-implied and $\widehat{\Sigma_t^{LF}}$ forecasts. These results suggest that both option-factor-implied and $\widehat{\Sigma_t^{LF}}$ forecasts are significantly poorer than those offered by the remaining models. The failure of the MCS-test applied to MMSE to exclude the $\widehat{\Sigma_t^{LF}}$ forecasts may be indicative of a lack of power.

Apart from offering a robustness check, the UMSEs are intuitive to interpret, allowing the models to be ranked informally. Based on the MSE-VC, the Factor ARMA is found to have the most accurate forecasts, followed by the $\widehat{\Sigma_t^{HF}}$, CF-ARMA, Factor Double ARCH, DCC, $\widehat{\Sigma_t^{LF}}$ and lastly the option-factor-implied forecasts. These rankings are consistent with the number of rejections found when applying the MCS-test to the UMSEs, but differ to ranks based on the difference between 0 and the MMSE. The MSE-C and MSE-V statistics separate the relative importance of accurately forecasting the variance and covariances in MSE-VC. The informal ranks based on MSE-C and MSE-V are almost identical so that, at least in this example, the models overall forecasting performance is reflected in both the covariances and variances[22].

4.1.3 Optimality Assessment

Table 11 Panel A summarises the results of running the optimality regression (22) for each forecasting model and for all elements of the covariance matrix. For each model the mean values of a_i and b_i, taken across $i = 1, \ldots, \frac{N(N+1)}{2}$, are given, and denoted $\bar{a}$ and $\bar{b}$, respectively, along with the number of covariance elements for which H_0:$a_i = 0$ and H_0:$b_i = 1$ are rejected according to t-tests at the 5% significance level. The last two columns provides the mean value of the adjusted coefficient of determination, $\bar{R}^2$, and the number of covariance matrix elements for which H_0:$a_i = 0 \cap b_i = 1$ is jointly rejected by Wald tests at the 5% significance level[23].

From Table 11 Panel A it can be seen that across all models the optimality condition is rejected for a large proportion of the covariance matrix elements. The Factor ARMA model results in the fewest joint rejections of the optimality condition, followed closely by the DCC, whilst the number of rejections for the remaining models are all in excess of 135. It should also be noted that

[22]The Factor Double ARCH model is ranked above the CF-ARMA model using MSE-C and the option-factor-implied forecasts are ranked above $\widehat{\Sigma_t^{LF}}$ using MSE-V

[23]From this point forward, the joint null hypothesis, H_0:$a_i = 0 \cap b_i = 1$, will be referred to as the optimality condition.

the DCC, Factor Double ARCH and CF-ARMA models all appear to substantially underestimate future changes in the covariance matrix elements, since all have mean $\bar{b} \geq 2.64$. However, the majority of the b_i values are not significantly greater than 1 on an individual basis.

Surprisingly, the highest $\bar{R}^2$ is obtained for $\widehat{\Sigma}^{LF}_{i,j,t}$, with $\widehat{\Sigma}^{HF}_{i,j,t}$ generating the second highest value. Continuing in descending order of $\bar{R}^2$, the DCC, Factor Double ARCH and option-factor-implied models are ranked joint third, followed by the Factor ARMA and CF-ARMA. These rankings are in marked contrast to those based on UMSEs. Section 2.3.3 remarked that models may be ranked differently according to the MSE and adjusted $\bar{R}^2$ criteria since the MSE evaluates the raw forecasts whilst the adjusted $\bar{R}^2$ assesses forecasts once they have been linearly projected onto the forecast target. Although the $\widehat{\Sigma}^{HF}_{i,j,t}$ estimator continues to perform well under this metric, the dynamic models based on high-frequency data are now the poorest performing models. Moreover, the improved performance of the option-factor-implied model under this criterion is the first indication that such an approach to measuring covariances may be beneficial.

Together, the number of rejections of the optimality condition and $\bar{R}^2$ suggest that all the models produce biased forecasts for the majority of the covariance matrix elements and that once this bias is corrected for, the low-frequency based models produce more accurate forecasts than the high-frequency based models, except for $\widehat{\Sigma}^{HF}_{i,j,t}$. Thus, it would appear the high-frequency models that performed well according to direct comparisons of the MSE loss function produced biased forecasts, where the bias was beneficial to the overall accuracy of the forecasts. In contrast, the option-factor-based forecasts, although expected to perform poorly due to the *ad hoc* nature in which they were constructed for the 1-day horizon, also appear to have been biased, but in this case the bias appears to have had a materially negative impact on their accuracy.

Table 12 Panel A summarises the results of the orthogonality tests by enumerating the elements of the covariance matrix for which the J-statistic was significant at the 5% level for each model. Each row contains the model for which the regression (22) was estimated and each column lists the instrument added in (25). Each cell of the table then contains the number of covariance elements for which the orthogonality condition was rejected for the model in a given row and an instrument in the corresponding column. A rejection is interpreted to mean the information contained in the covariance forecast for an element of the covariance matrix from the model listed in the row is not orthogonal to the information in the forecast produced by the model in the corresponding column. The results in Panel A do not provide trenchant conclusions as to whether the information in one model dominates all others. However, the relatively low number of rejections observed for the DCC and Factor Double ARCH are indicative of these models incorporating information excluded from the others. It is also interesting that $\widehat{\Sigma}^{HF}_{i,j,t}$ and $\widehat{\Sigma}^{LF}_{i,j,t}$ do not appear to contain unique information, as evidenced by the relatively high number of rejections in the rows corresponding to these two models.

4.1.4 The Effect of Overnight Returns

Thus far, the forecast target has been based on RC^{sub}_t, which is only measured over the period during which the stock market is open, namely 09:30-16:00 EST. The forecasting models based on daily data produce forecasts of the covariance for the full day, which includes the inactive, or overnight, period. In the context of volatility forecasting, Martens (2002) has shown using a target representing volatility for only the active period may result in an underestimation of forecast accuracy. To determine the effect of overnight returns on forecasting performance, the above analysis is repeated using forecast targets that incorporate information in the overnight period[24] Firstly, the RC metrics that incorporate overnight returns are briefly introduced.

To obtain an estimate of the integrated covariation over both the active and inactive periods, overnight returns must be incorporated into the computation of RC_t^{sub}. The first method to achieve this is,

$$RC_t^{total,sub} = RC_t^{sub} + RC_t^{ovn}, \tag{29}$$

and RC_t^{ovn} is given by,

$$RC_t^{ovn} = r_{i,0,t} r_{j,0,t}, \tag{30}$$

where $r_{i,0,t}$ is, for asset i, the return from the close to the open of the market. Although RC_t^{ovn} is an unbiased estimator of the integrated covariation during the close to open period, it is a substantially less efficient estimator than any of the RC estimators for the active period. The additional variance in $RC_t^{total,sub}$ resulting from RC_t^{ovn} may mean it is more effective to scale the RC estimators for the active period by a constant to produce an estimate for the entire 24 hour day.

Both approaches have been considered and compared for the estimation of RV over the 24 hour period in Hansen and Lunde (2005). In their study, the optimal combination of squared overnight returns and the RV, for the active period of the day, is also examined. It is found that the RV contains far more information for estimating the volatility over the entire day than the squared overnight returns, approximately 80% across a range of stocks, supporting the use of a constant to scale the RV to estimate the volatility over the full day. Martens (2002) has also demonstrated that scaling produces more accurate full day volatility estimates when high-frequency returns are unavailable for the inactive period. Although RV and RC measure different quantities, the results of Martens (2002) and Hansen and Lunde (2005) provide some justification for applying the scaling approach to RC. Hence, a second estimator for the full 24 hour day may be formed by,

$$RC_t^{scale,sub} = cRC_t^{sub}. \tag{31}$$

The constant, c, may be estimated by taking the mean of the ratio of $RC_t^{total,sub}$ and RC_t^{sub} across $t = 1, \ldots, T$. The reduction in sampling variation associated with $RC_t^{scale,sub}$ and the estimates of c are shown in Table 13[25].

Panel B of Tables 8 and 9 summarises the forecasting performances of each model according to MMSE and UMSE, respectively, where $RC_t^{total,sub}$ is used as the forecast target. The results in Panel B resemble those in Panel A of Table 8; the option-factor-implied forecasts are excluded from the MCS. However, the results in Panel B of Table 9 do differ noticeably to those of Panel A. The $\widehat{\Sigma}_{i,j,t}^{LF}$ estimator again results in the largest number of covariance matrix elements being excluded from the MCS, but now the number of exclusions is similar for the $\widehat{\Sigma}_{i,j,t}^{HF}$ and option-factor-implied estimators. The forecasts from the other models are always included in the MCS. The rankings of the models based on UMSE also differ in Panel B, with the Factor Double ARCH model being ranked first, followed by the Factor ARMA, DCC, CF-ARMA, option-factor-implied, $\widehat{\Sigma}_{i,j,t}^{HF}$ and $\widehat{\Sigma}_{i,j,t}^{HF}$ models. Therefore, the inclusion of the overnight period does appear to have improved the ranking of the low-frequency dynamic models as well as the option-factor-implied model. However, the values of MSE-VC, MSE-C and MSE-V are universally higher and is probably a consequence of the higher sampling variation observed in $RC_t^{total,sub}$. Both these changes represent the effect including the overnight period can have on the results.

[24] Utilising daily data, the forecasts produced by the DCC and Factor Double ARCH models naturally correspond to periods covering the entire day. On the other hand, the forecasts obtained from the CF-ARMA and Factor ARMA models depend on the RC metric used. Since the parameter estimates examined in Section 4.1 are based on RC_t^{sub} measured over the active period only, to obtain forecasts for the full day, these models were re-estimated using the RC metrics that incorporate the inactive, overnight period. The parameters are not reported to conserve space.

[25] Note, c was estimated from the median of the ratios $RC_t^{total,sub}/RC_t^{sub}$ over $t = 1, \ldots, T$.

Results from the optimality tests are contained in Panel B of Tables 11 and 12. The results are analogous to those of the RC_t^{sub} forecast target in that the optimality condition is rejected for a large proportion of the covariance matrix elements for most models. However, the number of rejections is noticeably lower for the DCC and Factor Double ARCH models. Therefore, the inclusion of the overnight period in the $RC_t^{total,sub}$ estimator has rendered these models less bias, again demonstrating the effect the overnight period can have on the results. The ranking of the models by $\bar{R}^2$ has not been changed unrecognisably in Panel B; $\widehat{\Sigma}_{i,j,t}^{LF}$ is now ranked first, ahead of $\widehat{\Sigma}_{i,j,t}^{HF}$, the Factor ARMA is ranked joint third with the DCC and Factor Double ARCH and the CF-ARMA is ranked joint last with the option-factor-implied model. It should be recognised that the values of $\bar{R}^2$ are lower for each model, also reflecting the higher sampling variation of the $RC_t^{total,sub}$ estimator. The interpretation of the orthogonality tests in Panel B of Table 12 is nearly identical to that in Panel A.

Panels C of Tables 8, 9, 11 and 12 repeat the results where the $RC_t^{scale,sub}$ estimator is used as the forecast target. For brevity, the results will not be discussed in detail and only important changes will be highlighted. Firstly, the direct comparisons based on MMSE and UMSE appear to reflect those of Panel A more closely. A large number of the covariance forecasts from both the option-factor-implied and $\widehat{\Sigma}_{i,j,t}^{LF}$ models are excluded from the MCS and the option-factor-implied forecasts are ranked last again, according to the UMSE. The DCC and Factor Double ARCH continue to have the fewest covariance forecasts reject the optimality condition and the model rankings according to $\bar{R}^2$ remain almost unchanged to those in Panel B whilst the magnitudes of the $\bar{R}^2$s return to a similar levels seen in Panel A. There is little change in the orthogonality results.

The changes to the results described above showed that including the overnight period in the forecast target can ellicit important changes, particularly for the DCC and Factor Double ARCH models. The $RC_t^{scale,sub}$ estimator is chosen over $RC_t^{total,sub}$ because of its lower sampling variation. Therefore, in the following analyses, $RC_t^{scale,sub}$ will be used as the forecast target.

4.1.5 Graphical Illustration

In order to provide a means to interpret the results visually, Fig. 2 plots the covariance series between Coca Cola and Citigroup and the 1-day-ahead forecasts produced by the seven models. Given the number of stocks, it is impossible to plot each series in the covariance matrix and, therefore, one series is selected as an example. The Coca-Cola-Citigroup series is chosen because its $RC_t^{scale,sub}$ has the largest sample variance and therefore should represent a challenging example for the models. If the models are able to capture the covariance dynamics adequately in this example, it is assumed their performance in other series would exceed that found here.

To ensure the covariance series and the seven forecasts are discernable, they have been divided into four subplots. Fig. 2a plots the forecasts from the low-frequency dynamic models, Fig. 2b plots the forecasts from the high-frequency dynamic models, Fig. 2c plots the option-factor-implied forecasts and Fig. 2d plots the historical average based models. The realised series is included in each subplot for comparison. From Fig. 2a it can be seen that the DCC forecasts trace the realised series very closely whilst the Factor Double ARCH forecasts show little variation and tend to lie above the realised covariances. There is a pronounced peak in the covariance series which occurs in July 2002. The DCC model fails to match the magnitude of this abrupt increase but does provide a much more marked response than the Factor Double ARCH model.

In Fig. 2b the CF-ARMA forecasts appear to track the realised covariances during the tranquil period, from 2003 onwards, with reasonable accuracy. However, during the more turbulent period prior to 2003, the forecasts behave erratically with a tendancy to overestimate the covariances.

In contrast, the Factor ARMA model forecasts appear to remain unbiased throughout the sample period; the forecasts trace a path that is fairly central to that taken by the realised covariances. Again, both models fail to adequately capture the abrupt increase in realised covariance during July 2002.

The forecasts produced by the option-factor-implied model can be seen to be highly inaccurate in Fig. 2c. The method used for their construction means that the forecasts are not expected to match the dynamics of the realised covariances with a great deal of accuracy. Assuming this example is representative for other elements of the covariance matrix, it is clear why this model is rejected by the MMSE and UMSE loss functions; the forecasts persistently remain above the realised series. This also explains the very high number of optimality condition rejections. In contrast, the $\widehat{\Sigma}_{i,j,t}^{LF}$ and $\widehat{\Sigma}_{i,j,t}^{HF}$ forecasts lie close to the realised series and are difficult to identify in Fig. 2d. One reason for this difficulty is that the y-axis scale has to be larger than in the previous figures because of the overreaction in the $\widehat{\Sigma}_{i,j,t}^{LF}$ forecasts. In contrast to the other models, the $\widehat{\Sigma}_{i,j,t}^{LF}$ produces excessively high covariance forecasts and overshoots the peak in covariance observed during July 2002.

Collectively, the plots suggest the rejection of the option-factor-implied forecasts according to MMSE and the large number of covariance forecasts excluded from the UMSE based MCSs are likely to be due to the very poor representation of the covariance dynamics and apparent bias in the forecasts. Similarly, the large number of $\widehat{\Sigma}_{i,j,t}^{LF}$ forecasts excluded from the UMSE based MCSs is probably due to their excessive variation. At least visually, the Factor ARMA and DCC models appear to track the dynamics of the covariance series with the greatest accuracy which is largely reflected in the above results.

4.2 Multiperiod Forecasts

In this section 30, 90 and 180 day forecasts are evaluated. In order to make the forecasts from historical time series comparable to those obtained through the option-factor-implied model, forecasts are made on dates that option-factor-implied forecasts are available. Given the discussion in Section 3.2, this means multiperiod forecasts are made every Wednesday on a weekly basis. For the in-sample period, there are 205, 198 and 186 weeks available for the 30, 90 and 180 day forecasts, respectively.

4.2.1 Statistical Loss Function Assessment

The MCS for the multiperiod forecasts based on the MMSE differs to that found for the 1-day ahead forecasts, as shown in Panels D-F of Table 8. Unexpectedly, the $\widehat{\Sigma}_{i,j,t}^{HF}$ estimator is excluded from the MCS of the 30-day forecasts, whilst the CF-ARMA is excluded for the 90- and 180-day forecasts. Once again, to provide a robustness check and allow a more intuitive discussion, Panels D-F of Table 9 summarise the results from applying the UMSE. Comparing Tables 8 and 9 a clear discrepancy emerges between the results. For each multiperiod forecast horizon, the largest number of covariance forecasts excluded from the MCS occurs for the option-factor-implied model.

The 30-day ahead model rankings, constructed from MSE-VC, do not differ substantially from the general pattern observed for the 1-day ahead scenario. The $\widehat{\Sigma}_{i,j,t}^{HF}$ forecasts perform well, being ranked first, whilst the $\widehat{\Sigma}_{i,j,t}^{LF}$ and option-factor-implied forecasts perform poorly, being ranked second from last and last, respectively. However, the ranks for the 90- and 180-day forecasts differentiate from this familiar pattern. The Factor Double ARCH, DCC and Factor ARMA models rank first, second and third in each case, whilst the $\widehat{\Sigma}_{i,j,t}^{LF}$ and option-factor-implied forecasts are ranked second from last and last, respectively, in both cases. Although the $\widehat{\Sigma}^{LF}_{i,j,t}$ and option-factor-implied forecasts continue to have the lowest ranks, it is illuminating that the the low-frequency dynamic models are ranked first and second in both cases. Since these forecast horizons represent time periods over which the realised covariance probably reverts towards its unconditional mean, the performance of the low-frequency dynamic models are likely to benefit from the variance targeting incorporated into the estimation procedure.

It is surprising the option-factor-implied forecasts perform poorly. Over longer forecast horizons it is expected that the information contained in this forward-looking measure would be advantageous. Secondly, although there is consistency between the ranks, using MSE-VC, for the 90- and 180-day forecasts and the 30-day ranks resemble those obtained for 1-day ahead forecasts, it should be remembered that these are informal. The differences observed between the MSE-VC values are very small and the number of forecasts excluded from the MCSs is low and decreases as the forecast horizon increases. This suggests the performance of the different models is very similar over these horizons.

4.2.2 Optimality Assessment

From Table 11 Panel D it can be seen that the $\widehat{\Sigma}^{HF}_{i,j,t}$ and $\widehat{\Sigma}^{LF}_{i,j,t}$ forecasts result in the fewest joint rejections of the optimality condition for the 30-day forecast horizon, followed by the DCC and Factor Double ARCH models. If $\bar{R}^2$ values are used to rank the models then the Factor Double ARCH provides superior forecasts, followed by the Factor ARMA and option-factor-implied models. Thus, like the 1-day ahead case this suggests the option-factor-implied approach may be used to form good covariance forecasts. Contrary to the results based on MSEs, the CF-ARMA is ranked second from last, just outperforming $\widehat{\Sigma}^{LF}_{i,j,t}$.

A similar scenario arises for the 90- and 180-day forecast horizons. In both cases the DCC and Factor Double ARCH result in the fewest joint rejections of the optimality condition. In terms of ranking models according to $\bar{R}^2$, the option-factor-implied covariances continue to perform well, being ranked third and second in the cases of 90- and 180-day forecast horizons, respectively. The $\widehat{\Sigma}^{HF}_{i,j,t}$ forecasts are ranked first in both cases and the DCC and Factor Double ARCH models are persistently ranked above the CF-ARMA and Factor ARMA models.

The orthogonality results for the 30-, 90- and 180-day forecast horizons are presented in Panels D-F, respectively, of Table 12. As before, the relatively low number of orthogonality rejections for the DCC and Factor Double ARCH models suggests they incorporate information excluded from the alternative forecasts. It should be noted that the $\widehat{\Sigma}^{HF}_{i,j,t}$ estimator now also appears to contain information orthogonal to the alternative estimators at the 30- and 90-day horizons, as suggested by the low number of rejections. Finally, the consistently high number of rejections for the option-factor-implied covariances when the Factor Double ARCH model is used as an instrument is indicative that the two estimators incorporate similar information and is likely to arise from the common assumption of a market-model. A similar, albeit weaker, relationship can be observed between the Factor ARMA and Factor Double ARCH models, but no such relationship is discernable between the Factor ARMA and option-factor-implied forecasts.

4.2.3 Graphical Illustration

In analogy to Section 4.1.5, Fig. 3 plots the in-sample 30-day Coca-Cola-Citigroup covariance forecasts generated by each model against the forecast target. Time series plots for 90- and 180-

day forecasts are excluded to conserve space, although the broad conclusions from the following discussion also apply in those cases.

In comparison to the 1-day ahead case, both the DCC and Factor Double ARCH forecasts appear to trace the forecast target with considerable accuracy, but, in similarity to the 1-day ahead case, both also fail to fully replicate the pronounced peak during July 2002. Fig. 3b shows the CF-ARMA forecasts are now much less erratic and tend to underestimate the covariance whilst the Factor ARMA model produces more volatile forecasts that frequently overestimate the covariance. The tendancy to under- and overestimate the covariance is particularly marked around the peak in covariance during July 2002.

Although still poor, the option-factor-implied forecasts now track the significant changes in the covariance series, albeit with a conspicuous positive bias. Another feature is the highly volatile nature of the forecasts relative to those of the other models and may in part be due to estimation error, where it should be remembered that as few as three option prices are used to estimate the implied betas and implied variances. Lastly, the forecasts produced by $\widehat{\Sigma}^{LF}_{i,j,t}$ and $\widehat{\Sigma}^{HF}_{i,j,t}$ in Fig. 3d provide, as with the 1-day ahead covariances, a good representation of the realised series, the main drawback being the accentuated lag between large changes in the realised series and the forecasts.

4.2.4 Discussion of In-Sample Results

There are several dimensions on which the performance of the models may be demarcated. Firstly, the models may be assessed individually. Secondly, data type may be used, being either high-frequency, low-frequency or option based. Thirdly, the model type can be used to categorise the performances, with the models being either dynamic, simple historical or option-implied. Lastly, the models can be divided into those that assume an underlying market model and those that do not. Collectively, the results do not suggest there is a consistent difference in the performance of the models based on data type, model type or the use of a market model assumption. However, a few noteworthy differences do appear to exist on an individual basis. In particular, the option-factor-implied model consistently performs very poorly when assessed by UMSE. It is ranked last by MSE-VC across all forecast horizons, has the largest number of individual covariance forecasts eliminated from the MCS based on UMSE and is excluded from the MCS based on MMSE at the 1-day horizon. The poor performance over the multi-day horizons is unexpected, since it seems reasonable for the forward-looking information incorporated in this model to be most beneficial over these horizons. The $\widehat{\Sigma}^{LF}_{i,j,t}$ estimator is also consistently ranked second from last according to MSE-VC and has the second largest number of individual covariance forecasts eliminated from the MCS based on UMSE.

However, the optimality tests provide some contrasting results. For example, the number of optimality condition rejections for the DCC model is low relative to the competing models at all horizons, suggesting this model produces forecasts that are less biased. When $\bar{R}^2$ is used to rank models, the performance of the option-factor-implied model improves considerably for multi-day horizons. This substantive improvement allied with the universally large number of optimality condition rejections suggests correcting for bias using a linear regression may improve the forecasting performance of the model. Such an adjustment is made by Pong et al. (2004) to their implied volatility forecasts and is supported by the, very informal, graphical evidence for Coca-Cola-Citigroup which shows a persistent positive bias in the forecasts. A positive bias is also consistent with the finding of Driessen et al. (2009) of a negative correlation risk premium. Although this result demonstrates that the forward-looking information contained in the forecasts may be beneficial over multi-day horizons, there is no clear advantage to using options as found in the volatility literature. Lastly, the orthogonality tests are more ambiguous, with the relatively low number of rejections for the

DCC and Factor Double ARCH indicating these models may incorporate information excluded from the remaining models.

As a final remark, it is disappointing that none of the models can be identified as clearly superior to all others, either through statistical loss functions or Mincer-Zarnowitz regressions. This contrasts with the volatility forecasting literature, where the merits and demerits of alternative models are better understood. The absence of such clean results in the multivariate case is indicative of the additional difficulty involved in forecasting covariance matrices.

5 Out-of-Sample Results

Results for the economically more relevant out-of-sample period, 05/01/2006 to 25/09/2008, are presented below and contrasted to those found for the in-sample period. Model parameters (unreported) are re-estimated on a rolling window basis, where the data over the prior 1000 days is used.

5.1 Statistical Loss Function Assessment

The MMSE results summarised in Table 14 differ considerably to those in Table 8. Now the DCC is consistently excluded from the MCS for all horizons. However, as for the in-sample case, the results in Table 15 based on the UMSE do not agree. The largest number of covariance forecasts excluded from the MCS are for the option-factor-implied model, when forecasting over 1- and 30-days, whereas similar numbers are excluded for the option-factor-implied and Factor Double ARCH modes over the 90- and 180-day horizons. Furthermore, the DCC is ranked second based on MSE-VC for the 1-day horizon, although it does rank fifth for the 30- and 180-day horizons and sixth for the 90-day horizon.

Moreover, the option-factor-implied model is no longer consistently ranked last in the out-of-sample period. For the 1-day and 30-day horizons it is ranked second from last, but for the 90- and 180-day horizons it is ranked a respectful fourth. Along with the MCS-test on the MMSE, this is a clear departure from the in-sample results and suggests that in the more relevant out-of-sample forecasting scenario, option data is useful. Comparing the in-sample and out-of-sample results, it also appears the information incorporated in the different data types is driving the forecasting performances to a greater extent than the precise specification assumed for the covariance dynamics. In particular, the Factor ARMA and Factor Double ARCH models both substantially outperformed the option-factor-implied forecasts in-sample, but out-of-sample the dynamic models are producing inferior forecasts, at least at the 90- and 180-day horizons, despite all three models deriving their forecasts from the market-model.

Of course, the usual caveat of the rankings being informal and based on the MSE-VC applies. The differences in the values of the MSE-VCs are extremely small and the largest number of covariance forecasts excluded by the formal MCS-test come from the option-factor-implied and Factor Double ARCH models.

5.2 Optimality Assessment

The Wald tests of the optimality condition in Table 15 mimic those of the in-sample period in so far as the number of rejections for the DCC model is consistently low relative to the majority of alternative models. However, a large number of rejections is now associated with the Factor Double ARCH. The $\widehat{\Sigma}^{HF}_{i,j,t}$ estimator generates the fewest rejections at the 30- and 90-day forecast horizons, whilst the $\widehat{\Sigma}^{LF}_{i,j,t}$ estimator generates a low number of rejections at the 90- and 180-day forecast horizons.

Turning to the ranking of the models by $\bar{R}^2$, $\widehat{\Sigma}^{HF}_{i,j,t}$ clearly has the highest value for 1-day-ahead forecasts, more than double the value of the Factor ARMA model which is ranked second. The Factor Double ARCH is ranked third, one place higher than the CF-ARMA. At the 30-, 90- and 180-day forecasts the option-factor-implied forecasts prevail. Therefore, once the advantage of forecasting in-sample is removed from the DCC, Factor Double ARCH, CF-ARMA and Factor ARMA models, the benefits of the forward-looking information contained in option-factor-implied forecasts can be observed. These benefits are not highlighted when testing forecasting accuracy through MSE, probably because of the positive bias in the option-factor-implied forecasts, evidenced again by the values of $\bar{a}$ and $\bar{b}$. It should also be noted that out-of-sample forecasts result in a considerably lower $\bar{R}^2$ for most models, except the option-factor-implied forecasts.

In accordance with the in-sample period, the results of the orthogonality tests in Table 16 have no distinctive characteristics. Broadly, the DCC and Factor Double ARCH models do appear to contain information orthogonal to the alternatives, extending the finding from the in-sample period. Although these two models have the fewest orthogonality rejections associated with them in the majority of cases, this assertion is much more tentative here since the difference between the number of rejections is not as substantial as was found for the in-sample period. As an exception, the Factor ARMA model appears to subsume information contained in the alternative models in the majority of cases for the 1-day horizon. When used as an instrument it causes the largest number of rejections in all the models and has the fewest rejections when the CF-ARMA, option-factor-implied and $\widehat{\Sigma}^{HF}_{i,j,t}$ forecasts are used as instruments.

5.3 Economic Loss Function

Tables 17, 18, 19 and 20 summarise the RVs from the hedging portfolios, p-values from the application of the MCS-test and t-tests that compare pairwise relative magnitudes for horizons of 1-, 30-, 90- and 180-days, respectively. Panel A of each table provides scaled versions of the mean hedging portfolio RVs; each mean RV is scaled by the mean RV of the model that on average generated the lowest portfolio RV and is then multiplied by 100. Each row contains the RVs from a portfolio that hedged the stock listed for that row, with p-values from MCS-tests in parentheses under each value. Panel B provides t-statistics for the null hypothesis that α_u in (27) is zero for each pairwise combination of the portfolio RVs. A significant negative t-statistic indicates the model listed for that column produces a superior covariance forecast relative to the model listed in the corresponding row.

However, firstly, it should be noted the covariance forecasts from the CF-ARMA and Factor ARMA models are excluded and that only the 1-day-ahead option-factor-implied covariance forecasts are included. This is due to the covariance forecasts from these models producing erratic weight vectors which in turn resulted in exceptionally high portfolio RVs. Due to the absence of a shortselling constraint, it was common for these covariance forecasts to produce weight vectors that placed a very large levered position in one stock, funded by a large short position in another[26]. Any material difference in the returns from the levered and short position resulted in substantial changes in the portfolio value, significantly increasing its variance. As a result, these results were excluded as they were not deemed sensible.

[26]Experimentation with mean-variance efficient portfolios other than those corresponding to the hedging scenario produced similar results.

Turning to the results in Table 17 it can be seen that the DCC model is found to generate hedging portfolios with the lowest RV on average over the 1-day horizon. More importantly, for the majority of hedging portfolios, all models are excluded from the MCS except the DCC at the 5% significance level. The t-tests in panel B confirm the hedging portfolio RVs resulting from the DCC model to be significantly lower than all the alternative models. Hence, the DCC model can be considered to outperform all it competitors. Based on the t-tests in Panel B, the Factor Double ARCH model is ranked second, generating hedging portfolios with RVs significantly lower than all other models except the DCC, and the high-frequency historical covariance estimates are ranked third. Despite the large mean RVs resulting from the option-factor-implied covariances, no significant difference was found between the portfolio RVs generated by this model and the low-frequency historical covaraince estimates[27].

The results over the 30-day horizon in Table 18, where the option-factor-implied covariance forecasts are now excluded, show that it is the high-frequency historical covariance forecasts that are now superior. For the majority of hedging portfolios all other models are excluded from the MCS at the 5% significance level and the pairwise comparisons in Panel B reveal the RVs to be significantly lower than the alternative models. The DCC is ranked second, with the RVs of the associated hedging portfolios being significantly lower than those of the Factor Double ARCH and low-frequency historical covariance forecasts. The results for the 90- and 180-day-ahead cases are similar. From tables 19 and 20, it can be seen that again the high-frequency historical covariance forecasts appear to produce superior covariance forecasts relative to all alternative models. In both cases the DCC model would appear to be ranked second according to the mean RVs of its associated hedging portfolios, but the t-tests show that the Factor Double ARCH model generates hedging portfolios with lower RVs on average. The low-frequency covariance forecasts remain the poorest according to the pairwise comparisons.

5.4 Graphical Illustration

To complete the empirical results, time series plots of the forecasts against the forecast targets are provided. For consistency, the Coca-Cola-Citigroup covariance is selected as the example and only the 1- and 30-day forecasts are examined, in Figs. 4 and 5, respectively. Briefly, the main features of the 1-day ahead time series in Fig. 4 appear to be consistent with those in Fig. 2. The DCC forecasts trace the realised covariance with reasonable accuracy, unlike the Factor Double ARCH forecasts which remain largely unresponsive. The Factor ARMA forecasts appear to provide a more accurate representation of the realised covariances compared to those of the CF-ARMA. Unsurprisingly, the option-factor-implied forecasts do not provide accurate forecasts whilst the $\widehat{\Sigma}^{LF}_{i,j,t}$ and $\widehat{\Sigma}^{HF}_{i,j,t}$ forecasts can be seen to be quite accurate. None of the forecasts fully predict the sporadic sharp increases in covariance observed from August 2007 onwards, although $\widehat{\Sigma}^{LF}_{i,j,t}$ again frequently overestimates these peaks.

Near identical obsevations can be made from Fig. 5, except the lag between the realised covariance and the $\widehat{\Sigma}^{LF}_{i,j,t}$ and $\widehat{\Sigma}^{HF}_{i,j,t}$ estimators means these forecasts now appear very poor. Consequently, there are departures from the conclusions made in Section 4.2.3. For example, the Factor Double ARCH forecasts remain unresponsive in the 30-day out-of-sample forecast scenario and the option-factor-implied forecasts do not appear to mimic the changes in realised covariance as accurately.

[27] The summarised portfolio mean RVs are a little misleading. The large relative values reported in table 17 result from occasional days when the hedging portfolio RV from the option-factor-implied covariances is very large. On the majority of days in the out-of-sample period, the RVs are similar to those obtained from the low-frequency historical covariances.

5.5 Discussion of Out-of-Sample Results

In similarity to the in-sample results, the out-of-sample results also do not present any clear differences between models based on data type, model type or inclusion of a factor structure. However, there are some important differences. In similarity to the in-sample results, the $\widehat{\Sigma}_{i,j,t}^{LF}$ estimator ranks poorly according to MSE-VC and the largest number of individual covariance forecasts excluded from the MCS are for the option-factor-implied model based on UMSE across all forecast horizons. However, the option-factor-implied model is no longer ranked so poorly according to MSE-VC. Although ranked second from last based on MSE-VC at the 1- and 30-day horizons, it ranks fourth for the 90- and 180-day horizons, which suggests this forecast may have more value out-of-sample.

The optimality results also show a further improvement in the performance of the option-factor-implied model, where it ranks first for all horizons except 1-day based on $\bar{R}^2$. Such a contrast between the in-sample and out-of-sample performances may be partly a consequence of the other models performing more poorly due to the advantage of parameter estimation and forecasting occurring in-sample being removed. Once again, the large number of optimality condition rejections and the informal evidence gleened from the plot of the Coca-Cola-Citigroup example shows that linearly projecting the option-factor-implied forecasts onto the forecast target probably benefits their performance by removing positive bias. Otherwise the out-of-sample optimality results are in agreement with those from the in-sample period: The DCC model continues to produce forecasts that are consistently less biased relative to the majority of other models across all forecast horizons and the orthogonality results remain difficult to interpret.

The results obtained from the economic loss function provide additional insights into the relative performance of the models in an economically relevant situation. Firstly, it is interesting that the covariance forecasts from the CF-ARMA and Factor ARMA models produce hedging portfolio weights that behave erratically. Therefore, although the covariance forecasts produced by these models are not deemed to be particularly poor when evaluated using statistical loss functions, they lose this quality in an economically relevant setting. Secondly, the superior performance of the DCC model for the 1-day-ahead forecast horizon supports its claims of providing an accurate representation of multivariate volatility dynamics. Lastly, over longer horizons, the results support the simple high-frequency historical covariance estimate as the best forecast amongst the alternatives considered here.

It should also be noted that the results do not agree with those of the two most closely related studies of Voev (2007) and Chiriac and Voev (2008). In Voev (2007) both the individual elements of the covariance matrix and its Cholesky factors are represented by ARMA(1,1) processes and compared to historical covariance estimates. They show the dynamic models to be superior forecasters relative to the historical covariance estimates based on the UMSE obtained for each covariance matrix element[28]. In an economically more relevant comparison, Chiriac and Voev (2008) find forecasts from a VARFIMA model applied to Cholesky factors to result in mean-variance optimised portfolios that second order stochastically dominate those resulting from DCC and BEKK forecasts. However, these studies also differ in that Voev (2007) only considers low-frequency daily data and Chiriac and Voev (2008) includes only 6 stocks in the covariance matrix. Neither study examines option data or factor models. However, the prior evidence on the performance of implied covariances computed from equity options is more consistent with the results here. The implied-correlation index of Skintzi and Refenes (2005) is found to provide upwardly bias forecasts whilst the R^2 from Mincer-Zarnowitz regressions is higher for the HEITIC model in Buss and Vilkov (2008), which combines both option data and historical data, than historical covariance estimates.

[28] They also show shrinking the covariance matrices, as proposed by Ledoit and Wolf (2003, 2004), to be beneficial.

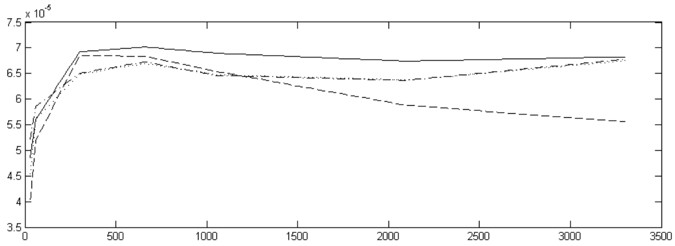
6 Conclusion

It is difficult to identify a single superior model and make statements on the merits of using different data types, model types or on the effect of assuming a market model structure. However, the performance of some individual models has drawn attention. For example, the contrast in the performance of the option-factor-implied model from the in-sample period to the out-of-sample period was particularly pronounced, migrating from being the poorest performing model to potentially one of the best, especially for multi-day forecast horizons. The Mincer-Zarnowitz regressions provided the strongest evidence in support of the option-factor-implied model's potential, which in turn suggeted the forecasts may be further improved by performing a linear bias correction.

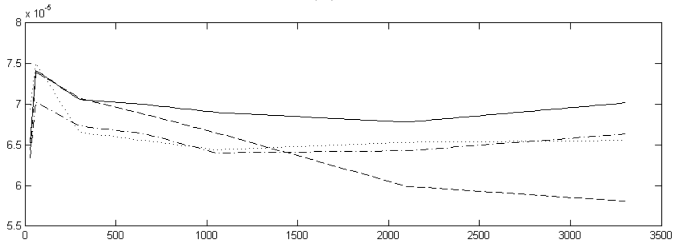
The differences in performance under alternative evaluation methods also provided for interesting interpretations of the results. Most notable were the inability of the CF-ARMA, Factor ARMA and option-factor-implied models to produce reasonable hedging portfolio weight vectors and the domination of the $\widehat{\Sigma}_t^{HF}$ over the 30-, 90- and 180-day horizons when employing the economic loss function. Therefore, although it is difficult to make general statements concerning the performance of different data types, the near universally superior performance of $\widehat{\Sigma}_t^{HF}$ relative to $\widehat{\Sigma}_t^{LF}$ tentatively favours the use of high-frequency data. Indeed, it is difficult to avoid the pervasiveness of $\widehat{\Sigma}_{i,j,t}^{HF}$ as one of the better forecasts according to all evaluation criteria, reinforced by its performance under the economic loss function, and makes it a strong contender to be the method recommended in forecasting covariance matrices. At the very least it brings the advantages of dynamic models into question.

In addition to forecasting performance, there are practical considerations when implementing these models. In particular, there needs to be more observations than assets to compute a covariance matrix based on RC. Here we found 17.5 min returns to be the appropriate frequency at which to sample returns in order to trade-off sampling variation against microstructure noise, limiting the number of assets that can be included in the covariance matrix to 22. This does not present a problem until vast dimensional matrices are required, which are arguably more relevant to practitioners, and imposes an important constraint on the application of $\widehat{\Sigma}_t^{HF}$. Thus, research on methods to mitigate the effect of microstructure noise are important to increase the number of assets that can be included in covariance matrices constructed from RC. In conjunction with the absence of any significant differences between the performance of dynamic models based on high-frequency and low-frequency data, this constraint also lends support to the use of the low-frequency based models such as DCC and the Factor Double ARCH, especially since estimation techniques have been developed for vast dimensional cases (Engle et al. (2008) and Engle (2009b)). Taking the argument one step further, the low-frequency dynamic models still require a sufficient number of historical observations to reliably estimate their parameters. For instance, in the DCC model the constant terms in the GARCH-like specification governing the dynamics of the correlations were estimated using the variance-targeting technique, which requires there to be more historical observations than assets. In contrast to all these approaches, the option-factor-implied forecasts can be constructed from very few observations obtained on a single day. These practical considerations combined with the forecasting results suggest each of the forecasting approaches examined may be the preferred method depending on the application.

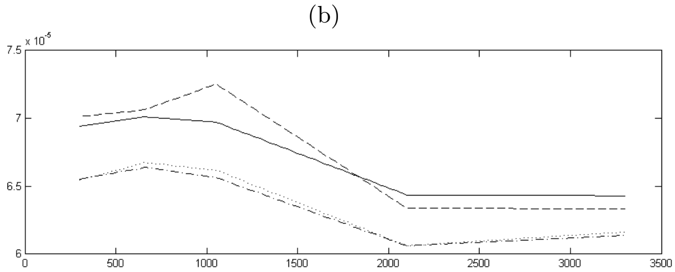
Lastly, several extensions to this research are envisaged. Firstly, the effect of bias correcting the option-factor-implied forecasts should be examined. Secondly, the range of models considered is naturally limited and could be increased. Finally, a larger set of assets could be selected so that the forecasting experiment conducted here could be repeated on multiple covariance matrices composed of different assets or the issue of forecasting vast dimensional covariance matrices could be examined.



(a)

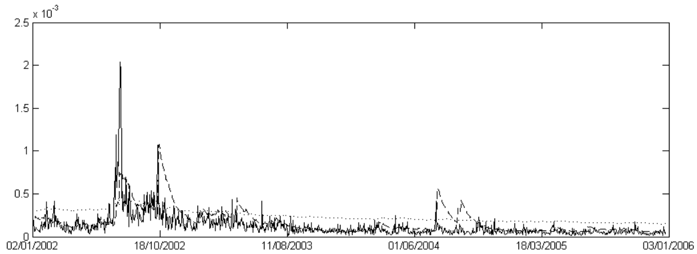


(b)

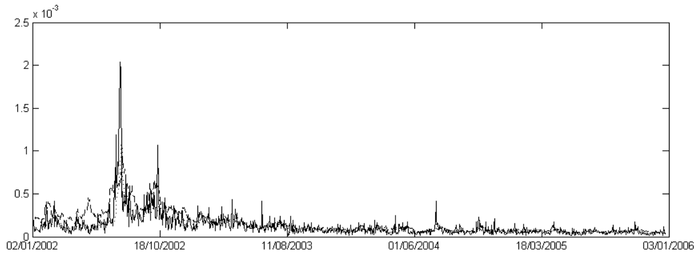


(c)

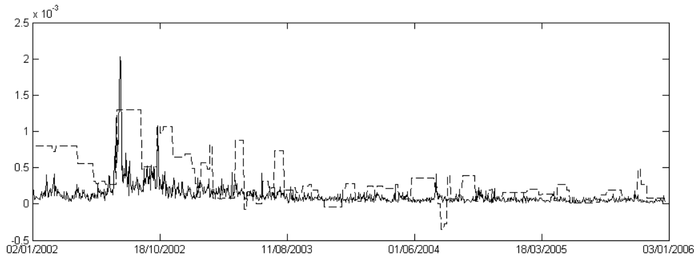
Figure 1: Signature plots of RC_t^{snd}, RC_t^{ll} and RC_t^{sub}, where returns have been computed using both transaction prices and mid-quotes in conjunction with either a previous tick (PT) or linear interpolation (LI) imputation method. To compute RC_t^{snd} and RC_t^{ll} 7 intraday time intervals were used, 30 seconds and 1, 5, 11, 17.5, 35 and 55 minutes, whereas 5 time intervals were used in computing RC_t^{sub}, 5, 11, 17.5, 35 and 55 minutes. The mean of each metric is taken across all stocks and scaled by 10,000 for presentation. The x-axis measures the interval over which intraday returns were computed in seconds. Fig. 1a contains signature plots of RC_t^{snd}, Fig. **??** of RC_t^{ll} and Fig. **??** of RC_t^{sub}. Each subfigure contains the respective metric measured using transaction prices using either LI (—) or PT (– – –) imputation methods and mid-quotes, also using LI (– - –) or PT (- - -) imputation methods. Note, in 1c the shortest interval over which returns were measured is 5 minutes (300 seconds), whereas the shortest interval is 30 seconds in all other plots.



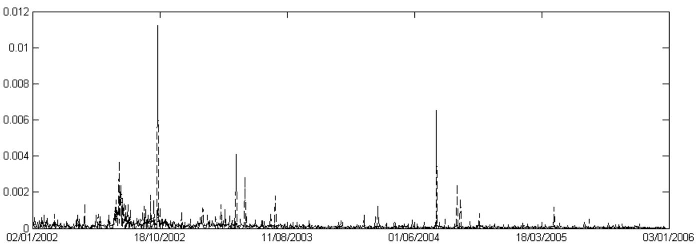
(a)



(b)

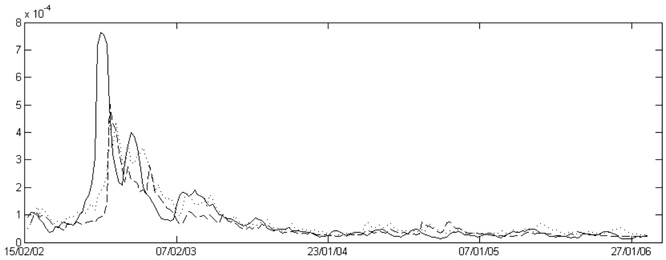


(c)

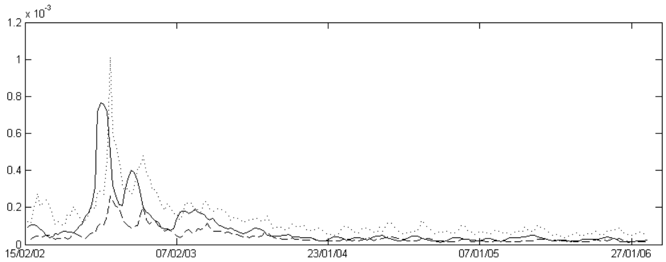


(d)

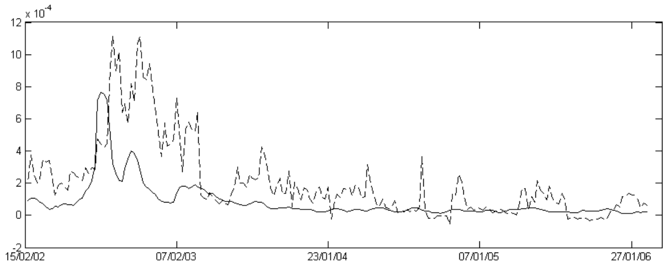
Figure 2: Plots of 1-day ahead in-sample forecasts against the RC_t^{sub} forecast target. Fig. 2a plots the 1-day ahead forecasts from the DCC (– – –) and Factor Double ARCH (- - -) models against the forecast target (—). Fig. 2b plots the 1-day ahead forecasts from the CF-ARMA (– – –) and Factor ARMA (- - -) models against the forecast target (—). Fig. 2c plots the 1-day ahead Option-Factor-Implied forecasts (– – –) against the forecast target (—). Fig. 2d plots the 1-day ahead Σ_t^{LF} forecasts (– – –) and Σ_t^{HF} forecasts (- - -) against the forecast target (—). Note, the scale in Fig. 2d differs to Figs. 2a- 2c.



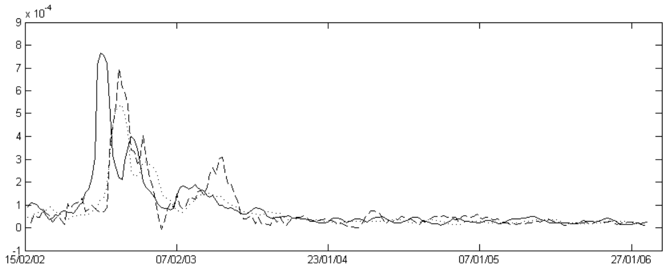
(a)



(b)

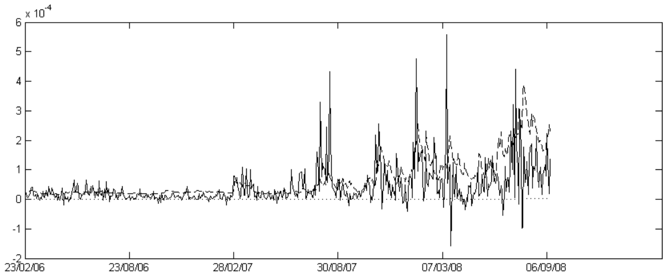


(c)

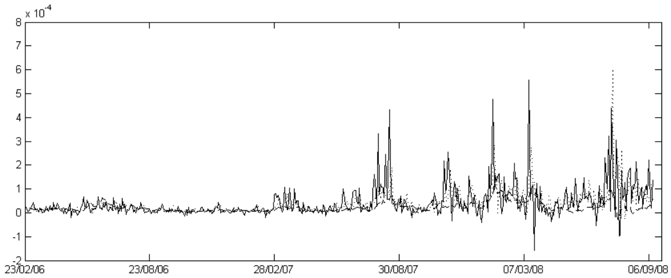


(d)

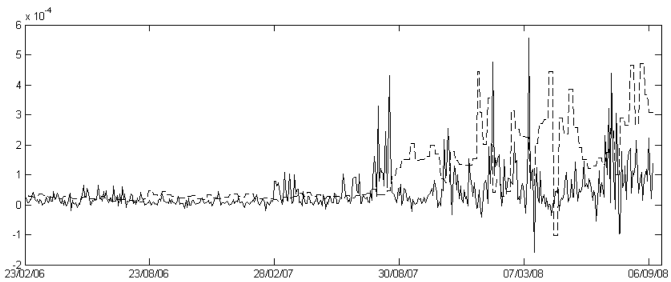
Figure 3: Plots of 30-day ahead in-sample forecasts against the RC_t^{sub} forecast target. Fig. 3a plots the 30-day ahead forecasts from the DCC (– – –) and Factor Double ARCH (- - -) models against the forecast target (—). Fig. 3b plots the 30-day ahead forecasts from the CF-ARMA (– – –) and Factor ARMA (- - -) models against the forecast target (—). Fig. 3c plots the 30-day ahead Option-Factor-Implied forecasts (– – –) against the forecast target (—). Fig. 3d plots the 30-day ahead Σ_t^{LF} forecasts (– – –) and Σ_t^{HF} forecasts (- - -) against the forecast target (—).



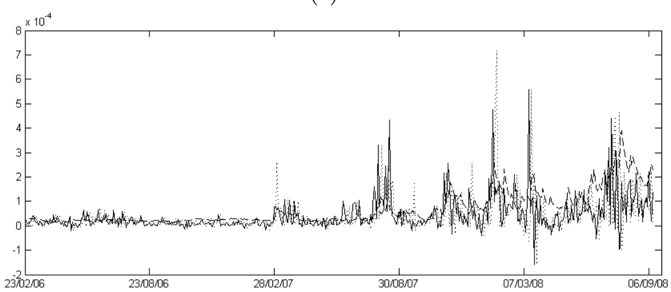
(a)



(b)

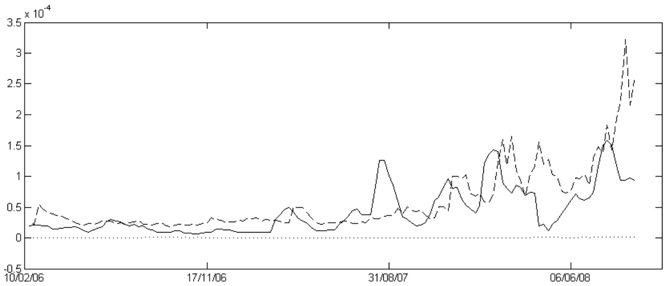


(c)

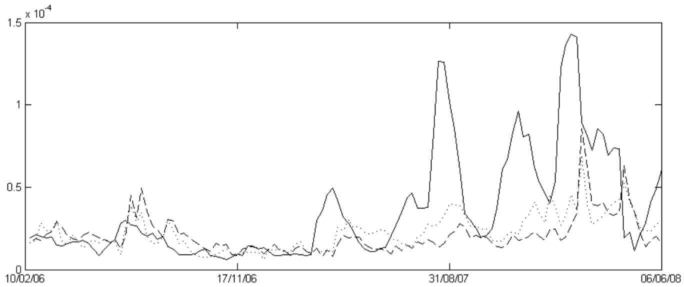


(d)

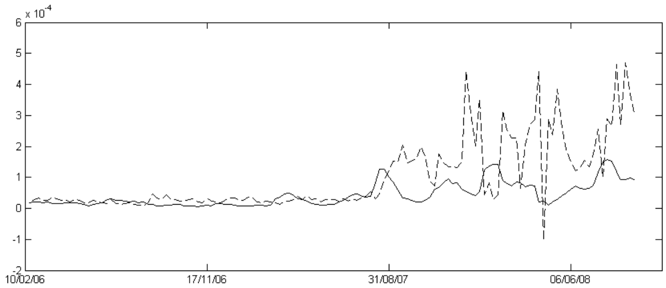
Figure 4: Plots of 1-day ahead out-of-sample forecasts against the RC_t^{sub} forecast target. Fig. 4a plots the 1-day ahead forecasts from the DCC (– – –) and Factor Double ARCH (- - -) models against the forecast target (—). Fig. 4b plots the 1-day ahead forecasts from the CF-ARMA (– – –) and Factor ARMA (- - -) models against the forecast target (—). Fig. 4c plots the 1-day ahead Option-Factor-Implied forecasts (– – –) against the forecast target (—). Fig. 4d plots the 1-day ahead Σ_t^{LF} forecasts (– – –) and Σ_t^{HF} forecasts (- - -) against the forecast target (—).



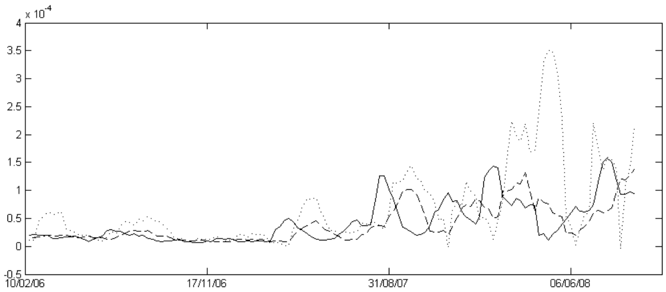
(a)



(b)



(c)



(d)

Figure 5: Plots of 30-day ahead out-of-sample forecasts against the RC_t^{sub} forecast target. Fig. 5a plots the 30-day ahead forecasts from the DCC (– – –) and Factor Double ARCH (- - -) models against the forecast target (—). Fig. 5b plots the 30-day ahead forecasts from the CF-ARMA (– – –) and Factor ARMA (- - -) models against the forecast target (—). Fig. 5c plots the 30-day ahead Option-Factor-Implied forecasts (– – –) against the forecast target (—). Fig. 5d plots the 30-day ahead Σ_t^{LF} forecasts (– – –) and Σ_t^{HF} forecasts (- - -) against the forecast target (—).

Table 1: Summary of companies. The name (Company Name), ticker symbol (Ticker), primary exchange (Ex.) and the number of days of missing data for transaction (T Days) and quote (Q Days) price data are provided. Note, N denotes the New York Stock Exchange and T the NASDAQ.

Company Name	Ticker	Ex.	T Days	Q Days
ALCOA INC	AA	N	0	104
BOEING CO	BA	N	0	3
CITIGROUP INC	C	N	0	103
GENERAL ELECTRIC CO	GE	N	0	86
HOME DEPOT INC	HD	N	0	1
HONEYWELL INTERNATIONAL INC	HON	N	1	3
INTERNATIONAL BUSINESS MACHS COR	IBM	T	0	186
INTEL CORP	INTC	T	0	0
J P MORGAN CHASE & CO	JPM	N	0	1
COCA COLA CO	KO	N	0	1
MCDONALDS CORP	MCD	N	0	1
PHILIP MORRIS COS INC	MO	N	1	2
MERCK & CO INC	MRK	N	0	1
MICROSOFT CORP	MSFT	T	0	0
PROCTER & GAMBLE CO	PG	N	0	12
WAL MART STORES INC	WMT	N	0	2
EXXON MOBIL CORP	XOM	N	0	133

Table 2: Summary of mean RC across all stocks for RC_t^{snd}, RC_t^{ll} and RC_t^{sub} estimators using transaction data with either LI imputation (Int-trans) or PT imputation (Prev-trans) and quote data with either LI imputation (Int-quote) or PT imputation (Prev-quote). The mean RCs are scaled by 10,000 and listed for sampling intervals of 30, 60, 300, 660, 1050, 2100 and 3300 seconds. Standard deviations are in parentheses and also scaled by 10,000.

Estimator	Data Type	Sampling Intervals						
		30	60	300	660	1050	2100	3300
RC_t^{snd}	Int-trans	0.486	0.56	0.693	0.702	0.689	0.674	0.682
		(1.819)	(1.375)	(1.429)	(1.536)	(1.489)	(1.641)	(1.764)
	Prev-trans	0.399	0.52	0.685	0.684	0.654	0.588	0.556
		(0.79)	(0.95)	(1.379)	(1.486)	(1.388)	(1.452)	(1.406)
	Int-quote	0.522	0.586	0.65	0.673	0.646	0.637	0.677
		(3.8)	(3.906)	(4.49)	(4.515)	(2.677)	(2.76)	(3.003)
	Prev-quote	0.454	0.561	0.649	0.669	0.648	0.637	0.675
		(4.633)	(3.849)	(3.219)	(3.41)	(2.911)	(2.833)	(3.071)
RC_t^{ll}	Int-trans	0.653	0.741	0.706	0.7	0.69	0.678	0.702
		(1.585)	(1.618)	(1.551)	(1.64)	(1.807)	(2.081)	(2.462)
	Prev-trans	0.64	0.74	0.707	0.688	0.665	0.599	0.581
		(1.128)	(1.381)	(1.534)	(1.616)	(1.707)	(1.77)	(1.992)
	Int-quote	0.694	0.75	0.665	0.655	0.644	0.653	0.656
		(3.316)	(3.625)	(3.411)	(3.47)	(2.702)	(3.295)	(3.374)
	Prev-quote	0.634	0.703	0.673	0.664	0.64	0.643	0.663
		(3.164)	(3.256)	(3.047)	(3.487)	(2.937)	(3.302)	(3.354)
RC_t^{sub}	Int-trans			0.694	0.701	0.697	0.644	0.643
				(1.354)	(1.438)	(1.459)	(1.459)	(1.624)
	Prev-trans			0.701	0.706	0.725	0.634	0.633
				(1.361)	(1.44)	(1.512)	(1.437)	(1.6)
	Int-quote			0.655	0.664	0.656	0.606	0.614
				(1.893)	(1.66)	(1.797)	(1.724)	(1.857)
	Prev-quote			0.654	0.667	0.662	0.606	0.616
				(1.918)	(1.666)	(1.802)	(1.719)	(1.847)

Table 3: Summary of the paramter estimates from the DCC model (See Section 2.2.1). The row labelled DCC contains the α_{qc} and β_{qc} parameter estimates, in the columns labelled α and β, respectively, of the process governing the quasi-correlations.

	μ	θ	ω	α	γ	β
AA	-1.83E-04	-6.39E-03	1.60E-06	1.15E-02	2.70E-02	9.70E-01
	(4.01E-07)	(1.31E-03)	(3.36E-12)	(5.51E-05)	(3.05E-04)	(8.88E-05)
BA	6.17E-04	-1.12E-01	1.31E-06	2.43E-07	4.04E-02	9.75E-01
	(2.46E-07)	(1.25E-03)	(1.59E-12)	(9.04E-04)	(3.04E-04)	(1.37E-03)
C	-6.06E-05	-1.84E-02	1.38E-06	1.65E-02	8.19E-02	9.36E-01
	(3.24E-07)	(6.65E-03)	(1.04E-12)	(2.60E-04)	(1.94E-03)	(1.06E-03)
GE	-1.51E-04	-4.19E-03	2.67E-07	3.01E-07	3.86E-02	9.79E-01
	(2.79E-07)	(3.83E-03)	(5.71E-14)	(2.80E-04)	(2.17E-04)	(1.52E-04)
HD	-2.03E-04	-5.70E-02	1.47E-06	1.30E-07	5.30E-02	9.69E-01
	(3.42E-07)	(2.08E-03)	(8.45E-13)	(1.20E-04)	(2.53E-04)	(7.07E-05)
HON	9.68E-05	-5.91E-02	1.78E-06	1.32E-02	3.61E-02	9.63E-01
	(3.65E-07)	(1.36E-03)	(1.41E-12)	(9.37E-05)	(4.55E-04)	(9.47E-05)
IBM	-3.84E-04	7.18E-03	3.74E-06	2.62E-02	1.61E-01	8.87E-01
	(2.88E-07)	(1.98E-03)	(2.55E-11)	(2.17E-03)	(2.34E-02)	(1.33E-02)
INTC	-2.41E-04	-6.03E-02	1.10E-06	1.24E-02	2.04E-02	9.74E-01
	(5.91E-07)	(2.43E-03)	(1.01E-12)	(1.92E-04)	(2.49E-04)	(7.54E-05)
JPM	8.43E-05	-5.49E-02	6.02E-07	2.07E-02	4.96E-02	9.52E-01
	(4.06E-07)	(5.30E-03)	(3.49E-13)	(3.11E-04)	(7.24E-04)	(6.34E-04)
KO	-1.55E-04	-2.06E-02	1.58E-06	2.04E-02	5.24E-02	9.44E-01
	(1.55E-07)	(1.29E-03)	(2.01E-12)	(3.08E-04)	(2.44E-03)	(5.50E-04)
MCD	2.39E-04	-1.99E-02	2.05E-06	2.30E-02	3.75E-02	9.53E-01
	(3.07E-07)	(2.51E-03)	(2.62E-12)	(2.42E-04)	(6.31E-04)	(3.10E-04)
MO	4.70E-04	2.41E-02	4.45E-06	5.28E-09	7.25E-02	9.49E-01
	(3.28E-07)	(2.37E-03)	(1.59E-11)	(5.99E-04)	(1.76E-03)	(2.62E-03)
MRK	-2.52E-04	4.29E-02	5.87E-06	3.45E-07	6.01E-02	9.49E-01
	(3.16E-07)	(1.76E-03)	(4.05E-11)	(2.72E-03)	(1.60E-03)	(3.17E-03)
MSFT	-2.52E-04	-2.4E-07	-2.44E-07	-2.44E-07	-2.44E-07	-2.44E-07
	-2.4E-07	(2.53E-03)	(4.14E-12)	(2.66E-03)	(2.70E-03)	(2.00E-03)
PG	3.74E-04	-1.17E-01	1.92E-06	8.59E-03	7.72E-02	9.36E-01
	(8.95E-08)	(1.98E-03)	(1.26E-12)	(2.30E-04)	(1.03E-03)	(5.25E-04)
WMT	-2.09E-04	-7.54E-02	7.46E-07	1.66E-02	2.00E-02	9.69E-01
	(1.69E-07)	(1.73E-03)	(5.41E-13)	(2.31E-04)	(2.72E-04)	(1.88E-04)
XOM	3.71E-04	-7.19E-02	3.96E-06	3.56E-02	4.32E-02	9.21E-01
	(1.80E-07)	(2.21E-03)	(5.68E-12)	(2.39E-04)	(8.87E-04)	(6.74E-04)
DCC				8.971E-09		9.998E-01
				(1.693E-09)		(7.670E-10)

Table 4: Factor Double ARCH parameter estimates (See Section 2.2.2). Each row contains parameter estimates for individual stocks. Standard errors are given in parentheses.

	μ	β^{ARCH}	ω	α	γ	β
AA	-3.08E-04	1.27E+00	2.37E-21	1.00E-05	1.00E-03	9.99E-01
	(1.47E-05)	(1.67E-03)	(4.56E-10)	(1.31E-04)	(2.37E-04)	(3.50E-05)
BA	5.03E-04	9.50E-01	6.62E-06	3.90E-02	5.90E-02	9.01E-01
	(1.45E-05)	(1.70E-03)	(2.21E-07)	(1.00E-03)	(1.48E-03)	(2.38E-03)
C	-1.96E-04	1.35E+00	2.23E-06	1.18E-01	-5.10E-02	8.88E-01
	(1.11E-05)	(1.88E-03)	(4.44E-08)	(2.01E-03)	(1.42E-03)	(1.66E-03)
GE	-2.82E-04	1.22E+00	2.53E-07	2.00E-03	2.30E-02	9.82E-01
	(1.03E-05)	(1.25E-03)	(3.63E-09)	(2.83E-04)	(3.57E-04)	(2.43E-04)
HD	-3.44E-04	1.11E+00	4.03E-07	8.00E-03	2.50E-02	9.76E-01
	(1.55E-05)	(1.75E-03)	(1.36E-08)	(5.02E-04)	(7.58E-04)	(2.51E-04)
HON	-2.44E-05	1.21E+00	2.91E-06	2.40E-02	4.00E-02	9.43E-01
	(1.55E-05)	(1.72E-03)	(6.31E-08)	(8.09E-04)	(1.44E-03)	(1.09E-03)
IBM	-5.15E-04	1.07E+00	8.65E-06	6.70E-02	2.10E-01	7.91E-01
	(1.22E-05)	(1.64E-03)	(1.80E-07)	(1.68E-03)	(4.31E-03)	(3.19E-03)
INTC	-4.03E-04	1.64E+00	8.76E-13	1.00E-05	3.00E-03	9.96E-01
	(1.89E-05)	(2.20E-03)	(6.26E-09)	(7.24E-04)	(7.93E-04)	(3.09E-04)
JPM	-7.06E-05	1.53E+00	8.46E-07	3.70E-02	3.50E-02	9.42E-01
	(1.35E-05)	(2.22E-03)	(2.17E-08)	(5.60E-04)	(1.35E-03)	(8.13E-04)
KO	-2.40E-04	5.72E-01	2.33E-06	4.90E-02	3.60E-02	9.17E-01
	(1.11E-05)	(1.49E-03)	(1.87E-07)	(1.83E-03)	(4.00E-03)	(4.58E-03)
MCD	1.45E-04	6.70E-01	1.20E-06	2.20E-02	1.80E-02	9.65E-01
	(1.63E-05)	(1.86E-03)	(3.34E-08)	(5.36E-04)	(7.26E-04)	(4.25E-04)
MO	3.95E-04	4.64E-01	3.40E-06	1.00E-05	7.00E-02	9.53E-01
	(1.69E-05)	(2.13E-03)	(1.69E-07)	(1.39E-03)	(2.25E-03)	(2.82E-03)
MRK	-3.84E-04	8.32E-01	3.08E-06	1.00E-05	3.80E-02	9.66E-01
	(1.44E-05)	(1.42E-03)	(1.74E-07)	(2.52E-03)	(1.28E-03)	(2.47E-03)
MSFT	-3.82E-04	1.20E+00	1.32E-07	2.00E-02	-2.00E-02	9.88E-01
	(1.21E-05)	(1.36E-03)	(2.51E-07)	(1.98E-02)	(6.50E-03)	(1.67E-02)
PG	2.92E-04	4.77E-01	1.21E-06	1.00E-02	7.40E-02	9.39E-01
	(9.46E-06)	(1.25E-03)	(2.65E-08)	(4.85E-04)	(8.69E-04)	(7.19E-04)
WMT	-3.33E-04	8.49E-01	8.74E-07	2.20E-02	-2.00E-03	9.70E-01
	(1.06E-05)	(1.22E-03)	(1.72E-08)	(3.43E-04)	(4.60E-04)	(3.51E-04)
XOM	2.75E-04	9.09E-01	1.07E-06	4.90E-02	-2.90E-02	9.57E-01
	(1.06E-05)	(1.42E-03)	(3.25E-08)	(4.20E-04)	(5.69E-04)	(6.22E-04)
DCC	6.83E-05	-5.20E-02	4.69E-07	1.00E-05	9.10E-02	9.50E-01
	(1.02E-05)	(1.42E-03)	(1.03E-08)	(4.15E-04)	(7.00E-04)	(6.26E-04)

Table 5: Summary of CF-ARMA parameters (See Section 2.2.3). The dimensions of the covariance matrix (17×17) requires 153 Cholesky factors to be modelled. To conserve space, only the mean, minimum and maximum values of each parameter across the 153 models are given. The last column contains counts of the number of significant parameter values across the models.

	Mean	Min	Max	Sig. Count
μ	2.72E-04	-4.28E-06	1.73E-03	151
ϕ_1	6.25E-01	-7.33E-01	1.29E+00	139
ϕ_2	8.85E-03	-3.09E-01	1.62E-01	100
ψ	-5.38E-01	-9.81E-01	7.93E-01	27

Table 6: Parameter estimates for the Factor ARMA model (See Section 2.2.4). Rows contain parameter estimates for each stock and the SPY. Standard errors are given in parentheses.

	μ	λ_1	λ_2	η
AA	-3.25E-01	9.10E-01	5.16E-02	-7.02E-01
	(4.01E-03)	(1.80E-03)	(1.46E-03)	(1.87E-03)
BA	-1.15E-01	9.76E-01	1.10E-02	-7.54E-01
	(2.73E-03)	(2.73E-03)	(2.50E-03)	(3.10E-03)
C	-8.69E-02	1.08E+00	-9.37E-02	-7.70E-01
	(1.35E-03)	(1.65E-03)	(1.58E-03)	(1.41E-03)
GE	-8.52E-02	9.36E-01	5.47E-02	-7.45E-01
	(1.48E-03)	(1.41E-03)	(1.35E-03)	(1.31E-03)
HD	-1.42E-01	9.84E-01	1.00E-03	-7.81E-01
	(2.36E-03)	(1.65E-03)	(1.48E-03)	(1.67E-03)
HON	-1.72E-01	9.68E-01	1.20E-02	-7.53E-01
	(2.32E-03)	(1.46E-03)	(1.34E-03)	(1.37E-03)
IBM	-1.54E-01	9.56E-01	2.73E-02	-7.14E-01
	(1.95E-03)	(1.44E-03)	(1.36E-03)	(1.37E-03)
INTC	-1.12E-01	9.97E-01	-9.60E-03	-7.49E-01
	(2.00E-03)	(2.03E-03)	(1.86E-03)	(2.13E-03)
JPM	-1.07E-01	1.03E+00	-4.56E-02	-7.66E-01
	(1.74E-03)	(1.72E-03)	(1.59E-03)	(1.69E-03)
KO	-1.64E-01	9.54E-01	2.84E-02	-7.30E-01
	(2.29E-03)	(1.62E-03)	(1.49E-03)	(1.59E-03)
MCD	-1.11E-01	1.05E+00	-6.21E-02	-8.49E-01
	(2.04E-03)	(1.33E-03)	(1.20E-03)	(1.10E-03)
MO	-4.91E-01	1.02E+00	-7.02E-02	-7.08E-01
	(2.18E-02)	(7.61E-03)	(5.29E-03)	(8.16E-03)
MRK	-3.03E-01	1.05E+00	-8.31E-02	-7.92E-01
	(5.15E-03)	(1.97E-03)	(1.50E-03)	(1.84E-03)
MSFT	-1.16E-01	9.86E-01	1.70E-03	-7.44E-01
	(1.70E-03)	(1.66E-03)	(1.57E-03)	(1.44E-03)
PG	-3.64E-01	1.05E+00	-8.71E-02	-7.36E-01
	(4.48E-03)	(2.10E-03)	(1.76E-03)	(1.87E-03)
WMT	-6.64E-02	1.07E+00	-7.42E-02	-8.32E-01
	(1.24E-03)	(1.29E-03)	(1.26E-03)	(8.92E-04)
XOM	-2.30E-01	1.00E+00	-2.64E-02	-7.18E-01
	(2.54E-03)	(1.63E-03)	(1.49E-03)	(1.41E-03)
SPY	-1.70E-01	8.59E-01	1.24E-01	-6.51E-01
	(2.38E-03)	(1.81E-03)	(1.72E-03)	(1.89E-03)

Table 7: Summary of implied betas and risk neutral variances (VAR) and skewnesses ($SKEW$) for the 30-, 90- and 180-day forecast horizons. The mean of each statistic is taken across the in-sample period.

	30 Day			90 Day			180 Day		
	β^{imp}	VAR	$SKEW$	β^{imp}	VAR	$SKEW$	β^{imp}	VAR	$SKEW$
AA	1.25	0.0172	-0.5245	1.19	0.0316	-0.5869	1.17	0.0510	-0.2240
BA	1.15	0.0155	-0.2791	1.13	0.0274	-0.6321	1.13	0.0499	-0.3981
C	1.15	0.0167	-0.3824	1.08	0.0250	-0.8334	1.23	0.0544	-0.6889
GE	1.16	0.0157	-0.4786	1.04	0.0241	-0.7535	1.16	0.0514	-0.5780
HD	1.29	0.0159	-0.4271	1.23	0.0299	-0.7819	1.32	0.0591	-0.5729
HON	1.13	0.0171	-0.3763	1.06	0.0292	-0.5702	1.01	0.0506	-0.3318
IBM	1.12	0.0136	-0.3768	1.04	0.0242	-0.7150	1.18	0.0484	-0.5723
INTC	1.71	0.0294	-0.3295	1.48	0.0507	-0.6554	1.69	0.1012	-0.5071
JPM	1.38	0.0261	-0.4000	1.33	0.0400	-0.8885	1.47	0.0810	-0.6511
KO	0.85	0.0099	-0.2254	0.73	0.0135	-0.6204	0.78	0.0268	-0.4853
MCD	1.05	0.0192	-0.3899	0.93	0.0257	-0.4867	0.83	0.0373	-0.4163
MO	1.14	0.0150	-0.3544	1.13	0.0255	-0.7180	1.27	0.0551	-0.6164
MRK	1.10	0.0125	-0.4155	1.05	0.0217	-0.6689	1.11	0.0422	-0.4554
MSFT	1.22	0.0180	-0.3684	1.12	0.0300	-0.7303	1.29	0.0631	-0.5709
PG	0.79	0.0079	-0.0128	0.77	0.0129	-0.7422	0.86	0.0266	-0.4530
WMT	0.88	0.0103	-0.3078	0.91	0.0183	-0.6842	1.04	0.0405	-0.5573
XOM	0.85	0.0088	-0.4230	0.81	0.0159	-0.6533	0.89	0.0296	-0.4484
S&P 500	-	0.0063	-1.1310	-	0.0244	-1.2196	-	0.0282	-1.2946

Table 8: Summary of in-sample multivariate MSEs for each model across all forecast horizons. The p-values are associated with tests of whether each model belongs to the model confidence set (MCS) and are derived from the maximum of the absolute value of the t-statistic. Panels A-C provide the results for the 1-day ahead horizon using the RC_t^{sub}, $RC_t^{sub,total}$ and $RC_t^{sub,scale}$ targets, respectively. Panels D-F provide the results for the 30-day, 90-day and 180-day horizons, respectively.

	Mean	Max	Min	p-value		Mean	Max	Min	p-value
Panel A					*Panel D*				
DCC	3.66E-06	1.21E-03	-2.2E-05	1		2.37E-06	1.57E-04	-2.2E-05	0.661
Factor Double ARCH	3.15E-06	1.11E-03	-2.9E-05	0.872		1.73E-06	1.45E-04	-3.3E-05	1
CF-ARMA	7.62E-06	1.29E-03	-2E-05	0.660		7.44E-06	1.84E-04	-7.4E-08	0.661
Factor ARMA	5.83E-06	1.12E-03	-4.8E-05	0.660		1.89E-06	1.28E-04	-3.8E-05	0.992
Option	-4.9E-06	9.44E-04	-1.11E-04	<0.001		-7.5E-06	8.81E-05	-1.26E-04	0.661
LF-Historical	-3.6E-06	1.57E-04	-1.37E-03	0.660		5.17E-07	1.59E-04	-4.97E-05	0.706
HF-Historical	2.82E-06	5.65E-04	-1.70E-04	0.660		4.39E-06	1.66E-04	-2.3E-05	0.025
Panel B					*Panel E*				
DCC	5.65E-05	8.99E-03	-1.4E-05	1		1.69E-06	5.20E-05	-1.3E-05	0.925
Factor Double ARCH	5.6E-05	8.99E-03	-2.4E-05	0.986		1.48E-06	5.31E-05	-2.3E-05	0.906
CF-ARMA	6.2E-05	9.00E-03	-1.4E-07	0.939		6.71E-06	7.43E-05	-7.2E-08	0.028
Factor ARMA	6.15E-05	9.00E-03	-5.1E-06	0.939		2.67E-06	5.33E-05	-6.7E-06	0.906
Option	4.13E-05	8.98E-03	-8.77E-05	0.001		-3E-06	4.45E-05	-5.31E-05	0.805
LF-Historical	4.67E-05	9.00E-03	-1.19E-03	0.939		7.76E-07	5.46E-05	-2.74E-05	0.906
HF-Historical	4.7E-05	9.00E-03	-3.36E-04	0.939		2.18E-06	6.28E-05	-1.7E-05	1
Panel C					*Panel F*				
DCC	8.61E-06	1.96E-03	-1.9E-05	0.758		1.82E-06	3.47E-05	-1.2E-05	0.944
Factor Double ARCH	8.41E-06	2.04E-03	-2.3E-05	1		1.97E-06	3.58E-05	-1.8E-05	0.989
CF-ARMA	1.22E-05	2.13E-03	-4.7E-05	0.758		6.54E-06	4.81E-05	-2.3E-08	0.003
Factor ARMA	1.21E-05	2.13E-03	-4.7E-05	0.758		3.19E-06	3.76E-05	-2E-06	0.944
Option	-1.9E-06	1.83E-03	-1.13E-04	0.001		-1.9E-06	2.62E-05	-3.79E-05	0.821
LF-Historical	-4.3E-07	2.72E-04	-1.14E-03	0.758		2.36E-06	4.34E-05	-1.35E-05	0.944
HF-Historical	-4.3E-07	2.72E-04	-1.14E-03	0.758		2.24E-06	3.85E-05	-8.5E-06	1

Table 9: Summary of in-sample UMSEs across all elements of the covariance matrix for forecast horizons of 1-, 30-, 90- and 180-days. Column 1 lists the models used to generate the forecasts. Columns 2-4 give the MSE of the forecasts averaged across the elements of the covariance matrix. Note, MSE-VC is the mean MSE taken across all elements of the covariance matrix, MSE-C is the mean MSE taken across only those elements representing covariances and MSE-V is the mean MSE taken across those elements representing variances. Columns 5-7 summarise the proportion (in percent) of covariance elements exluded from the model confidence set (MCS) at the 5%, 10% and 15% significance levels, respectively. Panels A-C provide the results for the 1-day ahead forecast horizon where RC_t^{sub}, $RC_t^{sub,total}$ and $RC_t^{sub,scale}$ are used as the forecast targets, respectively. Panels D-F provide the results for the 30-, 90- and 180-day forecast horizons, where $RC_t^{sub,scale}$ is used as the forecast target.

	MSE-VC	MSE-C	MSE-V	5%	10%	15%
Panel A						
DCC	2.66E-08	3.67E-08	1.48E-07	11	21	30
Factor Double ARCH	2.28E-08	3.29E-08	1.11E-07	8	11	16
CF-ARMA	2.24E-08	3.31E-08	1.01E-07	15	26	33
Factor ARMA	1.7E-08	1.32E-08	7.53E-08	2	2	3
Option-Factor-Implied	4.3E-07	4.28E-07	5.27E-07	110	123	131
LF Historical	2.65E-07	2.6E-07	9.68E-07	143	146	149
HF Historical	1.92E-08	2.55E-08	9.57E-08	0	2	2
Panel B						
DCC	1.85E-07	1.66E-07	1.37E-06	0	0	0
Factor Double ARCH	1.81E-07	1.64E-07	1.39E-06	0	0	0
CF-ARMA	1.9E-07	1.76E-07	1.36E-06	0	0	0
Factor ARMA	1.83E-07	1.1E-07	1.34E-06	0	0	0
Option-Factor-Implied	2.52E-07	2.41E-07	1.72E-06	15	15	16
LF Historical	3.67E-07	3.33E-07	2.09E-06	110	132	138
HF Historical	2.71E-07	2.03E-07	2.27E-06	10	14	17
Panel C						
DCC	3.69E-08	5.56E-08	2.05E-07	2	3	7
Factor Double ARCH	3.29E-08	5.26E-08	1.91E-07	2	3	3
CF-ARMA	3.48E-08	5E-08	1.83E-07	9	14	18
Factor ARMA	3.13E-08	2.27E-08	1.65E-07	0	1	2
Option-Factor-Implied	1.27E-07	1.67E-07	5.48E-07	75	83	88
LF Historical	2.6E-07	2.5E-07	9.52E-07	136	144	145
HF Historical	2.97E-08	4.03E-08	1.72E-07	0	0	1
Panel D						
DCC	1.57E-08	2.1E-08	8.36E-08	0	0	0
Factor Double ARCH	1.2E-08	1.46E-08	5.23E-08	0	0	0
CF-ARMA	1.16E-08	1.45E-08	4.25E-08	0	0	1
Factor ARMA	1.41E-08	1.66E-08	2.71E-08	2	2	3
Option-Factor-Implied	1E-07	1.22E-07	3.94E-07	58	115	136
LF Historical	3.12E-08	3.67E-08	1.16E-07	0	1	4
HF Historical	9.23E-09	1.29E-08	3.65E-08	0	0	0
Panel E						
DCC	1.26E-08	1.66E-08	6.79E-08	1	1	1
Factor Double ARCH	1.08E-08	1.37E-08	4.91E-08	0	0	0
CF-ARMA	1.9E-08	2.29E-08	8.97E-08	0	1	2
Factor ARMA	1.32E-08	1.67E-08	7.37E-08	0	0	0
Option-Factor-Implied	4.8E-08	5.42E-08	6.79E-08	9	26	45
LF Historical	2.46E-08	2.82E-08	9.29E-08	0	0	3
HF Historical	1.42E-08	1.83E-08	5.85E-08	0	0	0
Panel F						
DCC	1.24E-08	1.6E-08	6.76E-08	1	1	1
Factor Double ARCH	1.13E-08	1.38E-08	5.28E-08	0	0	0
CF-ARMA	1.86E-08	2.21E-08	8.91E-08	0	1	2
Factor ARMA	1.32E-08	1.64E-08	7.8E-08	0	0	0
Option-Factor-Implied	4.37E-08	4.07E-08	8.5E-08	13	16	20
LF Historical	3.2E-08	3.62E-08	1.33E-07	1	4	18
HF Historical	1.59E-08	1.9E-08	6.53E-08	0	0	0

Table 10: Summary of out-of-sample MSEs across all elements of the covariance matrix for forecast horizons of 1-, 30-, 90- and 180-days. Column 1 lists the models used to generate the forecasts. Columns 2-4 give the MSE of the forecasts averaged across the elements of the covariance matrix. Note, MSE-VC is the mean MSE taken across all elements of the covariance matrix, MSE-C is the mean MSE taken across only those elements representing covariances and MSE-V is the mean MSE taken across those elements representing variances. Columns 5-7 summarise the proportion (in percent) of covariance elements exluded from the model confidence set (MCS) at the 5%, 10% and 15% significance levels, respectively. Panels A-D provide the results for the 1-, 30-, 90- and 180-day forecast horizons, where $RC_t^{sub,scale}$ is used as the forecast target.

	MSE-VC	MSE-C	MSE-V	5%	10%	15%
Panel A						
DCC	2.52E-08	2.96E-08	1.62E-07	3	4	4
Factor Double ARCH	3.31E-08	3.75E-08	1.62E-07	6	11	12
CF-ARMA	2.89E-08	3.61E-08	1.67E-07	2	3	4
Factor ARMA	2.61E-08	2.22E-08	8.8E-08	0	0	0
Option-Factor-Implied	1.98E-07	1.94E-07	2.35E-07	95	109	117
LF Historical	1.09E-05	2.85E-07	0.000181	1	2	2
HF Historical	1.95E-08	2.56E-08	9.4E-08	0	1	2
Panel B						
DCC	1.16E-08	6.87E-09	1.29E-07	4	8	14
Factor Double ARCH	1.05E-08	1.09E-08	3.55E-08	0	3	7
CF-ARMA	9.26E-09	1.09E-08	5.45E-08	1	1	3
Factor ARMA	8.87E-09	7.53E-09	3.04E-08	0	0	0
Option-Factor-Implied	1.98E-07	1.94E-07	2.35E-07	64	97	107
LF Historical	3.84E-07	2.77E-08	6.15E-06	9	16	21
HF Historical	4.59E-09	5.91E-09	2.45E-08	0	0	0
Panel C						
DCC	1.3E-08	4.47E-09	1.71E-07	4	9	12
Factor Double ARCH	8.2E-09	8.18E-09	2.75E-08	21	48	59
CF-ARMA	6.9E-09	9.02E-09	3.91E-08	3	8	8
Factor ARMA	7.14E-09	5.8E-09	2.88E-08	1	6	10
Option-Factor-Implied	7.3E-09	8.04E-09	1.85E-08	33	43	55
LF Historical	9.86E-08	7.23E-09	1.57E-06	4	5	11
HF Historical	2.36E-09	3.05E-09	1.2E-08	0	0	0
Panel D						
DCC	6.25E-09	4.2E-09	5.98E-08	5	9	12
Factor Double ARCH	7E-09	7.13E-09	2.73E-08	16	42	64
CF-ARMA	5.44E-09	6.86E-09	3.1E-08	2	6	9
Factor ARMA	6.55E-09	5.51E-09	2.32E-08	2	6	11
Option-Factor-Implied	5.63E-09	5.92E-09	1.14E-08	32	41	47
LF Historical	2.02E-09	2.55E-09	1.29E-08	1	1	1
HF Historical	2.54E-09	3.42E-09	1.37E-08	0	0	0

Table 11: Summary of optimality regression results for the in-sample period. Column 1 lists the models, column 2 provides an average of the a_i parameters across all elements of the covariance matrix, $\bar{a}$, column 3 gives the total number of rejections of H_0:$a_i = 0$ according to t-tests conducted at the 5% level, column 4 provides an average of the b_i parameters across all elements of the covariance matrix, $\bar{b}$, column 6 gives the total number of rejections of H_0:$b_i = 0$ according to t-tests conducted at the 5% level and column 7 provides the total number of joint rejections of H_0:$b_i = 0 \cap a_i = 0$ according to Wald tests conducted at the 5% level. Panels A-C summarise these results for the 1-day forecast horizon, where the covariance target corresponds to RC_t^{sub}, $RC_t^{total,sub}$ and $RC_t^{scaled,sub}$, respectively. Panels D-F summarise the results for the 30-day, 90-day and 180-day forecast horizons, respectively.

	$\bar{a}$		**$\bar{a}$**		**R**	**Wald**
Panel A						
DCC	-3.53E-05	111	3.37	41	0.10	121
Factor Double ARCH	-4.60E-05	131	3.18	4	0.10	142
CF-ARMA	2.70E-05	134	2.64	15	0.04	135
Factor ARMA	5.63E-06	22	0.77	105	0.13	118
Option	-1.99E-04	153	0.50	153	0.10	153
LF-Hist	-4.82E-05	103	0.51	153	0.51	153
HF-Hist	2.08E-07	0	0.63	147	0.32	144
Panel B						
DCC	3.85E-05	98	4.12	39	0.02	95
Factor Double ARCH	2.78E-05	65	3.29	26	0.02	59
CF-ARMA	7.25E-05	130	2.40	10	0.01	129
Factor ARMA	7.51E-05	151	0.71	47	0.02	149
Option	-1.19E-04	142	0.49	114	0.01	151
LF-Hist	2.55E-05	79	0.51	153	0.37	153
HF-Hist	-1.6E-07	0	0.49	153	0.24	153
Panel C						
DCC	-8.7E-06	79	3.94	48	0.09	112
Factor Double ARCH	-1.9E-05	92	3.87	4	0.10	116
CF-ARMA	2.76E-05	131	2.67	14	0.04	135
Factor ARMA	3.05E-05	151	0.85	94	0.10	153
Option	-1.66E-04	153	0.50	142	0.02	153
LF-Hist	-2.2E-05	53	0.52	153	0.51	153
HF-Hist	-4E-08	0	0.64	141	0.32	133
Panel D						
DCC	-8.36E-06	43	1.70	25	0.10	59
Factor Double ARCH	-1.50E-05	58	1.27	2	0.16	78
CF-ARMA	5.79E-05	122	4.82	30	0.09	123
Factor ARMA	4.47E-05	149	1.40	7	0.15	112
Option	-1.58E-04	138	0.58	147	0.13	149
LF-Hist	-2.21E-05	17	0.71	45	0.06	56
HF-Hist	2.51E-05	46	1.78	0	0.12	22
Panel E						
DCC	-1E-05	36	1.36	11	0.06	45
Factor Double ARCH	-1.2E-05	40	1.07	1	0.11	29
CF-ARMA	6.13E-05	116	4.84	3	0.03	102
Factor ARMA	5.51E-05	144	1.42	0	0.04	80
Option	-1.03E-04	135	0.60	140	0.09	149
LF-Hist	-2.2E-05	3	1.69	6	0.06	13
HF-Hist	-4.4E-06	0	3.71	98	0.12	105
Panel F						
DCC	-9.6E-06	37	1.31	16	0.04	37
Factor Double ARCH	-7.1E-06	27	0.98	1	0.06	17
CF-ARMA	6.47E-05	115	5.96	2	0.01	102
Factor ARMA	6.31E-05	152	1.82	0	0.01	107
Option	-9.9E-05	136	0.54	153	0.10	153
LF-Hist	-2.4E-05	0	4.16	74	0.08	105
HF-Hist	-1.1E-05	1	5.82	149	0.11	144

Table 12: Summary of orthogonality results for the in-sample period. The model listed in each row corresponds to the model for which the J-statistic in equation (26) is computed. The models listed in the columns correspond to the additional instrument in equation (25). The numbers in each cell count the rejections of H_0:$J_{i,T} = 0$ at the 5% level across all elements of the covariance matrix. A rejection indicates that the forecast produced by the model listed for that row is not orthogonal to the information contained in the model listed in the corresponding column. Panels A-C summarise these results for the 1-day forecast horizon, where the covariance target corresponds to RC_t^{sub}, $RC_t^{total,sub}$ and $RC_t^{scaled,sub}$, respectively. Panels D-F summarise the results for the 30-day, 90-day and 180-day forecast horizons, respectively.

	DCC	Factor Double ARCH	CF-ARMA	Factor ARMA	Option	LF-Hist	HF-Hist
Panel A							
DCC	-	0	22	95	22	13	17
Factor Double ARCH	10	-	22	94	22	13	17
CF-ARMA	73	96	-	101	5	8	36
Factor ARMA	113	127	101	-	35	10	12
Option	135	150	130	107	-	51	67
LF-Hist	153	153	153	148	110	-	55
HF-Hist	151	153	152	141	64	14	-
Panel B							
DCC	-	0	11	71	14	17	18
Factor Double ARCH	12	-	11	71	14	17	18
CF-ARMA	80	112	-	96	19	29	37
Factor ARMA	81	107	70	-	40	11	11
Option	129	150	136	117	-	39	60
LF-Hist	150	152	150	147	138	-	76
HF-Hist	148	152	150	148	138	11	-
Panel C							
DCC	-	0	25	93	13	24	24
Factor Double ARCH	12	-	25	93	13	24	24
CF-ARMA	67	93	-	96	6	18	18
Factor ARMA	124	128	103	-	54	20	20
Option	127	150	129	107	-	51	51
LF-Hist	152	153	153	149	152	-	0
HF-Hist	151	153	153	148	152	0	-
Panel D							
DCC	-	0	2	7	3	0	1
Factor Double ARCH	2	-	2	7	3	0	1
CF-ARMA	22	41	-	48	30	11	12
Factor ARMA	5	46	25	-	23	0	0
Option	100	133	90	79	-	69	114
LF-Hist	87	95	47	33	77	-	67
HF-Hist	17	49	19	4	12	2	-
Panel E							
DCC	-	0	2	7	3	0	1
Factor Double ARCH	4	-	12	5	7	0	3
CF-ARMA	26	66	-	30	17	1	2
Factor ARMA	42	90	60	-	32	7	7
Option	87	138	70	38	-	22	25
LF-Hist	89	83	37	16	69	-	50
HF-Hist	43	52	5	3	38	14	-
Panel F							
DCC	-	0	55	38	24	9	5
Factor Double ARCH	23	-	55	38	24	9	5
CF-ARMA	31	55	-	56	27	3	1
Factor ARMA	81	87	95	-	72	6	16
Option	114	139	115	58	-	33	9
LF-Hist	145	153	151	88	147	-	129
HF-Hist	130	147	124	76	129	120	-

Table 13: Summary of the sampling variation of $RC_t^{total,sub}$ and $RC_t^{scale,sub}$ and the estimated scale factor c used to compute $RC_t^{scale,sub}$. Each cell refers to the RC between the stocks listed in each row and corresponding column and contains three estimates. The first is the sample standard deviation of $RC_t^{total,sub}$ taken across the sample period; the second is the sample standard deviation of $RC_t^{scale,sub}$ taken across the sample period; the last is an estimate of the scale factor c computed from $T^{-1}\sum_{t=1}^{T}\frac{RC_t^{total,sub}}{RC_t^{sub}}$.

	AA	BA	C	GE	HD	HON	IBM	INTC	JPM	KO	MCD	MO	MRK	MSFT	PG	WMT	XOM
AA	8.32E-04																
	4.89E-04																
	1.46																
BA	3.70E-04	5.67E-04															
	1.73E-04	3.53E-04															
	1.25	1.50															
C	5.63E-04	4.60E-04	1.99E-03														
	3.64E-04	2.35E-04	1.10E-03														
	1.29	1.26	1.32														
GE	3.99E-04	3.31E-04	8.40E-04	8.33E-04													
	2.35E-04	1.94E-04	4.29E-04	5.36E-04													
	1.31	1.33	1.34	1.40													
HD	3.85E-04	3.21E-04	6.02E-04	3.92E-04	9.74E-04												
	2.14E-04	1.74E-04	3.77E-04	2.17E-04	4.43E-04												
	1.23	1.28	1.29	1.25	1.38												
HON	4.31E-04	3.54E-04	5.21E-04	4.19E-04	3.74E-04	1.01E-03											
	2.00E-04	1.78E-04	2.42E-04	2.02E-04	1.90E-04	4.37E-04											
	1.36	1.38	1.24	1.35	1.27	1.40											
IBM	3.63E-04	2.53E-04	4.59E-04	3.76E-04	3.22E-04	3.49E-04	6.30E-04										
	1.50E-04	1.30E-04	2.26E-04	1.77E-04	1.62E-04	1.38E-04	2.38E-04										
	1.22	1.27	1.25	1.31	1.25	1.28	1.40										
INTC	4.29E-04	3.79E-04	5.36E-04	4.17E-04	4.50E-04	4.30E-04	4.55E-04	1.58E-03									
	1.99E-04	1.64E-04	3.13E-04	2.30E-04	2.21E-04	1.86E-04	2.08E-04	4.91E-04									
	1.24	1.23	1.24	1.26	1.24	1.28	1.31	1.41									
JPM	5.34E-04	4.30E-04	1.39E-03	6.48E-04	5.64E-04	5.26E-04	4.70E-04	5.66E-04	1.61E-03								
	3.24E-04	2.31E-04	7.99E-04	4.14E-04	3.41E-04	2.60E-04	2.28E-04	3.23E-04	1.14E-03								
	1.27	1.22	1.36	1.35	1.31	1.30	1.24	1.30	1.37								
KO	2.36E-04	1.97E-04	2.84E-04	2.13E-04	2.39E-04	2.15E-04	2.22E-04	4.71E-04	3.23E-04	3.80E-04							
	1.16E-04	1.07E-04	1.56E-04	1.31E-04	1.22E-04	1.08E-04	8.91E-05	1.19E-04	1.87E-04	1.75E-04							
	1.20	1.18	1.20	1.24	1.20	1.18	1.20	1.22	1.18	1.35							
MCD	3.14E-04	2.64E-04	3.29E-04	2.70E-04	3.13E-04	2.79E-04	2.22E-04	3.27E-04	3.61E-04	1.77E-04	7.24E-04						
	1.36E-04	1.11E-04	1.89E-04	1.36E-04	1.47E-04	1.22E-04	9.46E-05	1.27E-04	1.89E-04	8.06E-05	3.57E-04						
	1.28	1.21	1.23	1.24	1.23	1.22	1.20	1.18	1.26	1.19	1.42						
MO	3.93E-04	3.88E-04	5.75E-04	4.29E-04	4.50E-04	3.68E-04	2.52E-04	5.22E-04	6.24E-04	1.90E-04	2.99E-04	2.69E-02					
	1.03E-04	1.03E-04	1.34E-04	1.07E-04	1.14E-04	1.03E-04	8.02E-05	1.11E-04	1.42E-04	7.40E-05	8.36E-05	4.64E-04					
	1.19	1.27	1.14	1.19	1.23	1.26	1.18	1.21	1.24	1.14	1.19	1.37					
MRK	3.26E-04	2.78E-04	3.85E-04	2.82E-04	3.14E-04	2.98E-04	2.52E-04	3.06E-04	4.16E-04	2.11E-04	2.15E-04	4.25E-03	2.52E-03				
	1.54E-04	1.49E-04	2.12E-04	1.59E-04	1.63E-04	1.34E-04	1.26E-04	1.49E-04	2.48E-04	1.24E-04	1.03E-04	9.17E-05	4.30E-04				
	1.24	1.29	1.22	1.23	1.26	1.22	1.25	1.15	1.26	1.26	1.26	1.23	1.44				
MSFT	3.41E-04	2.69E-04	4.86E-04	3.26E-04	3.29E-04	3.08E-04	3.10E-04	4.55E-04	4.57E-04	2.03E-04	2.38E-04	4.59E-04	2.48E-04	6.68E-04			
	1.56E-04	1.43E-04	2.45E-04	1.90E-04	1.66E-04	1.52E-04	1.64E-04	2.47E-04	2.56E-04	1.06E-04	1.08E-04	9.50E-05	1.37E-04	2.85E-04			
	1.19	1.26	1.26	1.26	1.16	1.24	1.33	1.28	1.25	1.20	1.23	1.19	1.22	1.37			
PG	1.84E-04	1.68E-04	2.10E-04	1.56E-04	1.85E-04	1.72E-04	1.32E-04	1.68E-04	2.24E-04	1.41E-04	1.40E-04	1.60E-04	1.69E-04	1.45E-04	2.15E-04		
	8.37E-05	9.35E-05	1.21E-04	1.05E-04	9.28E-05	8.42E-05	7.43E-05	9.50E-05	1.28E-04	7.58E-05	6.40E-05	6.11E-05	8.76E-05	8.35E-05	1.18E-04		
	1.14	1.28	1.17	1.30	1.16	1.20	1.19	1.16	1.23	1.25	1.24	1.19	1.25	1.19	1.32		
WMT	2.63E-04	2.42E-04	3.44E-04	2.52E-04	3.18E-04	2.51E-04	2.11E-04	3.08E-04	3.44E-04	1.58E-04	2.00E-04	3.19E-04	2.21E-04	2.47E-04	1.37E-04	3.37E-04	
	1.50E-04	1.37E-04	2.32E-04	1.72E-04	2.02E-04	1.44E-04	1.30E-04	1.62E-04	2.31E-04	1.04E-04	1.13E-04	8.92E-05	1.39E-04	1.42E-04	8.46E-05	2.44E-04	
	1.19	1.22	1.23	1.22	1.27	1.22	1.24	1.16	1.24	1.16	1.23	1.19	1.24	1.16	1.17	1.36	
XOM	3.10E-04	2.45E-04	4.58E-04	2.90E-04	2.78E-04	2.65E-04	2.26E-04	2.94E-04	4.08E-04	1.64E-04	2.07E-04	2.20E-04	2.38E-04	2.43E-04	1.55E-04	2.04E-04	3.63E-04
	1.97E-04	1.36E-04	2.41E-04	1.76E-04	1.46E-04	1.41E-04	1.15E-04	1.53E-04	2.36E-04	9.83E-05	1.03E-04	8.93E-05	1.29E-04	1.39E-04	7.65E-05	1.23E-04	2.38E-04
	1.36	1.25	1.16	1.21	1.08	1.25	1.17	1.17	1.17	1.17	1.15	1.14	1.18	1.22	1.18	1.12	1.34

Table 14: Summary of out-of-sample multivariate MSEs for each model across all forecast horizons. The p-values are associated with tests of whether each model belongs to the model confidence set (MCS) and are derived from the maximum of the absolute value of the t-statistic. Panels A-D provide the results for the 1-day, 30-day, 90-day and 180-day horizons, respectively.

	Mean	Max	Min	p-value
Panel A				
DCC	1.94E-06	1.69E-04	-6.9E-06	<0.001
Factor Double ARCH	3.98E-06	1.95E-04	-7.1E-07	0.684
CF-ARMA	4.01E-06	1.90E-04	-1.9E-07	0.684
Factor ARMA	2.49E-06	1.62E-04	-7E-06	0.684
Option	5.83E-07	1.97E-04	-2.37E-05	0.684
LF-Historical	1.94E-06	1.69E-04	-6.93E-06	1
HF-Historical	1.95E-06	1.41E-04	-3.90E-05	0.999
Panel B				
DCC	2.44E-06	2.36E-05	-1.2E-07	0.005
Factor Double ARCH	2.44E-06	2.36E-05	-1.2E-07	1
CF-ARMA	2.41E-06	2.55E-05	-9.3E-08	0.518
Factor ARMA	1.97E-06	2.10E-05	-3.2E-07	0.634
Option	-1.8E-06	1.48E-05	-2.64E-05	0.518
LF-Historical	-7.2E-07	1.27E-05	-1.34E-05	0.518
HF-Historical	8.96E-07	1.88E-05	-4.88E-06	0.518
Panel C				
DCC	5.89E-07	1.34E-05	-4.3E-06	0.001
Factor Double ARCH	2.32E-06	2.26E-05	5.73E-08	0.644
CF-ARMA	2.18E-06	2.38E-05	-1E-07	0.644
Factor ARMA	1.91E-06	2.27E-05	-1.1E-06	1
Option	-4E-07	5.31E-06	-5.80E-06	0.644
LF-Historical	-4.9E-07	4.98E-06	-6.65E-06	0.644
HF-Historical	9.4E-07	1.67E-05	-2.02E-06	0.644
Panel D				
DCC	5.51E-07	4.14E-06	-9.5E-07	<0.001
Factor Double ARCH	1.92E-06	6.28E-06	7.34E-08	0.570
CF-ARMA	1.66E-06	5.88E-06	-1.1E-07	0.570
Factor ARMA	1.47E-06	5.58E-06	-5.6E-07	1
Option	-5.2E-07	2.1E-06	-5.8E-06	0.570
LF-Historical	5.73E-07	3.09E-06	-1E-06	0.570
HF-Historical	1.15E-06	4.56E-06	-1E-07	0.608

Table 15: Summary of optimality regression results for the out-of-sample period. Column 1 lists the models, column 2 provides an average of the a_i parameters across all elements of the covariance matrix, $\bar{a}$, column 3 gives the total number of rejections of H_0:$a_i = 0$ according to t-tests conducted at the 5% level, column 4 provides an average of the b_i parameters across all elements of the covariance matrix, $\bar{b}$, column 6 gives the total number of rejections of H_0:$b_i = 0$ according to t-tests conducted at the 5% level and column 7 provides the total number of joint rejections of H_0:$b_i = 0 \cap a_i = 0$ according to Wald tests conducted at the 5% level. Panels A-D summarise the results for the 1-day, 30-day, 90-day and 180-day forecast horizons, respectively.

	$\bar{\mathbf{a}}$		$\bar{\mathbf{b}}$		$\mathbf{R^2}$	**Wald**
Panel A						
DCC	-2.2E-05	74	1.40	21	0.02	92
Factor Double ARCH	4.28E-05	145	84.48	124	0.09	148
CF-ARMA	2.9E-05	139	3.17	47	0.06	138
Factor ARMA	1.38E-05	120	0.90	81	0.14	139
Option	-7.3E-05	131	0.37	141	0.02	150
LF-Historical	-3.81E-03	144	0.30	104	0.01	148
HF-Historical	2.26E-06	8	0.58	152	0.31	153
Panel B						
DCC	-2.3E-05	59	0.70	47	0.02	70
Factor Double ARCH	5.2E-05	141	10.43	39	<0.01	151
CF-ARMA	3.15E-05	121	1.68	45	0.03	125
Factor ARMA	2.75E-05	129	1.09	11	0.05	125
Option	-8E-05	140	0.51	153	0.22	153
LF-Historical	-4.5E-05	85	0.52	104	0.05	130
HF-Historical	4.71E-06	0	0.98	1	0.05	1
Panel C						
DCC	-1.8E-05	48	0.79	35	0.02	62
Factor Double ARCH	5.52E-05	139	14.54	19	<0.01	146
CF-ARMA	3.36E-05	117	1.46	37	0.02	125
Factor ARMA	2.4E-05	105	0.66	76	0.03	144
Option	-4.1E-05	132	0.57	139	0.09	147
LF-Historical	-2.6E-05	51	0.49	60	<0.01	88
HF-Historical	1.22E-05	21	0.54	18	<0.01	21
Panel D						
DCC	-1.6E-05	51	0.86	12	0.01	40
Factor Double ARCH	5.52E-05	119	12.71	25	<0.01	139
CF-ARMA	3.17E-05	114	1.37	44	0.02	119
Factor ARMA	2.42E-05	104	0.73	33	0.01	110
Option	-3.7E-05	134	0.47	149	0.12	153
LF-Historical	5.46E-06	29	0.67	10	0.01	30
HF-Historical	1.94E-05	104	4.83	67	0.09	108

Table 16: Summary of orthogonality results for the out-of-sample period. The model listed in each row corresponds to the model for which the J-statistic in equation (26) is computed. The models listed in the columns correspond to the additional instrument in equation (25). The numbers in each cell count the rejections of H_0:$J_{i,T} = 0$ at the 5% level across all elements of the covariance matrix. A rejection indicates that the forecast produced by the model listed for that row is not orthogonal to the information contained in the model listed in the corresponding column. Panels A-D summarise these results for the 1-day, 30-day, 90-day and 180-day forecast horizons, respectively.

	DCC	**Factor Double ARCH**	**CF-ARMA**	**Factor ARMA**	**Option**	**LF-Hist**	**HF-Hist**
Panel A							
DCC	-	0	95	126	9	15	49
Factor Double ARCH	56	-	95	126	9	15	49
CF-ARMA	62	56	-	123	15	11	27
Factor ARMA	72	49	87	-	7	35	12
Option	120	123	144	151	-	88	151
LF-Hist	85	51	95	125	19	-	92
HF-Hist	114	83	117	129	19	39	-
Panel B							
DCC	-	0	24	53	27	19	26
Factor Double ARCH	33	-	24	53	27	19	26
CF-ARMA	47	70	-	57	44	48	74
Factor ARMA	56	60	36	-	41	52	93
Option	57	55	42	86	-	54	69
LF-Hist	48	50	35	52	51	-	55
HF-Hist	35	39	9	30	29	6	-
Panel C							
DCC	-	0	14	6	37	50	44
Factor Double ARCH	22	-	14	6	37	50	44
CF-ARMA	56	87	-	19	62	57	92
Factor ARMA	61	98	57	-	49	65	101
Option	57	55	42	86	-	54	69
LF-Hist	38	50	32	6	34	-	46
HF-Hist	34	46	59	4	34	56	-
Panel D							
DCC	-	0	26	18	55	53	47
Factor Double ARCH	30	-	26	18	55	53	47
CF-ARMA	63	88	-	24	76	65	89
Factor ARMA	80	87	74	-	75	104	100
Option	88	81	99	34	-	76	34
LF-Hist	57	65	58	30	69	-	50
HF-Hist	83	76	120	49	103	107	-

Table 17: Panel A shows the RVs of the hedging portfolios obtained from the DCC, Factor Double ARCH, Option-factor-implied, low-frequency (LF) historical and high-frequency (HF) historical covariance forecasts for the 1-day forecast horizon. Each RV is scaled by the RV of the DCC model, which on average generates the lowest RV values, and multiplied by a factor of 100. Each row provides the scaled RV for the portfolio hedging the stock listed in column 1 and under each value the MCS p-value is given in parentheses. Panel B shows the t-statistics from tests of H_0: $RV_1 - RV_2 = 0$, where RV_1 is the portfolio RV obtained from the model listed across the columns and RV_2 is the portfolio RV obtained from the model listed in the corresponding row.

	DCC	Factor Double ARCH	Option	LF Historical	HF Historical
Panel A					
AA	100	135.70	3431.82	222.80	163.10
	(1)	(0.047)	(0.047)	(0.047)	(0.047)
BA	100	123.90	6119.46	259.52	161.90
	(1)	(0.075)	(0.075)	(0.075)	(0.075)
C	100	205.34	3379.94	237.19	166.92
	(1)	(0.048)	(0.048)	(0.048)	(0.048)
GE	100	150.51	539.50	261.58	181.82
	(1)	(0.013)	(0.013)	(0.013)	(0.013)
HD	100	159.54	16086.62	229.75	162.21
	(1)	(0.086)	(0.086)	(0.086)	(0.086)
HON	100	152.59	635.70	228.22	171.32
	(1)	(0.001)	(0.001)	(0.001)	(0.001)
IBM	100	134.23	546.69	221.20	164.83
	(1)	(0.035)	(0.035)	(0.035)	(0.035)
INTC	100	131.20	13588.79	240.70	144.44
	(1)	(0.104)	(0.104)	(0.104)	(0.104)
JPM	100	201.02	57409.55	241.46	172.17
	(1)	(0.041)	(0.041)	(0.041)	(0.041)
KO	100	112.46	3359.19	222.24	161.81
	(1)	(0.083)	(0.083)	(0.083)	(0.083)
MCD	100	118.87	13538.17	241.61	160.70
	(1)	(0.096)	(0.096)	(0.096)	(0.096)
MO	100	90.60	1316.28	1056.19	133.17
	(0.081)	(1)	(0.078)	(0.078)	(0.078)
MRK	100	97.28	10654.43	234.11	155.01
	(0.354)	(1)	(0.102)	(0.102)	(0.102)
MSFT	100	118.14	2553.29	246.78	146.48
	(1)	(0.105)	(0.105)	(0.105)	(0.105)
PG	100	75.99	1774.53	146.58	109.52
	(0.108)	(1)	(0.108)	(0.108)	(0.108)
WMT	100	137.21	33987.87	239.56	167.96
	(1)	(0.12)	(0.12)	(0.12)	(0.12)
XOM	100	126.28	52118.31	204.11	166.56
	(1)	(0.11)	(0.11)	(0.11)	(0.11)
Panel B					
Factor Double ARCH	-7.48				
Option	-3.09	-1.80			
LF Historical	-4.70	-6.32	-1.07		
HF Historical	-12.69	-8.48	0.92	2.23	

Table 18: Panel A shows the RVs of the hedging portfolios obtained from the DCC, Factor Double ARCH, low-frequency (LF) historical and high-frequency (HF) historical covariance forecasts for the 30-day forecast horizon. Each RV is scaled by the RV of the HF historical covariance, which on average generates the lowest RV values, and multiplied by a factor of 100. Each row provides the scaled RV for the portfolio hedging the stock listed in column 1 and under each value the MCS p-value is given in parentheses. Panel B shows the t-statistics from tests of H_0: RV_1 - $RV_2 = 0$, where RV_1 is the portfolio RV obtained from the model listed across the columns and RV_2 is the portfolio RV obtained from the model listed in the corresponding row.

	DCC	Factor Double ARCH	LF Historical	HF Historical
Panel A				
AA	106.58	143.99	255.67	100
	(0.017)	(<0.001)	(<0.001)	(1)
BA	108.47	130.81	358.55	100
	(0.002)	(0.002)	(0.002)	(1)
C	112.96	219.60	470.38	100
	(0.225)	(0.225)	(0.225)	(1)
GE	103.04	141.94	356.61	100
	(0.096)	(0.075)	(0.075)	(1)
HD	106.75	164.01	275.13	100
	(0.001)	(0.001)	(<0.001)	(1)
HON	101.44	153.99	301.14	100
	(0.379)	(0.009)	(0.009)	(1)
IBM	95.63	124.60	357.43	100
	(1)	(0.053)	(0.053)	(0.564)
INTC	110.33	143.47	310.90	100
	(<0.001)	(<0.001)	(<0.001)	(1)
JPM	110.12	209.40	395.98	100
	(0.183)	(0.183)	(0.183)	(1)
KO	100.19	106.32	228.56	100
	(0.984)	(0.564)	(0.001)	(1)
MCD	110.54	125.74	462.38	100
	(0.066)	(0.066)	(0.066)	(1)
MO	101.80	102.00	184.29	100
	(0.697)	(0.697)	(0.075)	(1)
MRK	111.94	110.87	312.49	100
	(0.029)	(0.036)	(0.013)	(1)
MSFT	107.18	117.13	310.66	100
	(0.078)	(0.016)	(0.016)	(1)
PG	142.93	109.99	218.38	100
	(0.003)	(0.003)	(<0.001)	(1)
WMT	105.68	138.18	288.54	100
	(0.004)	(0.003)	(0.003)	(1)
XOM	105.18	123.90	225.99	100
	(0.002)	(0.002)	(<0.001)	(1)
Panel B				
Factor Double ARCH	-20.20			
LF Historical	-8.76	-21.32		
HF Historical	19.57	14.57	13.41	

Table 19: Panel A shows the RVs of the hedging portfolios obtained from the DCC, Factor Double ARCH, low-frequency (LF) historical and high-frequency (HF) historical covariance forecasts for the 90-day forecast horizon. Each RV is scaled by the RV of the HF historical covariance, which on average generates the lowest RV values, and multiplied by a factor of 100. Each row provides the scaled RV for the portfolio hedging the stock listed in column 1 and under each value the MCS p-value is given in parentheses. Panel B shows the t-statistics from tests of H_0: RV_1 - $RV_2 = 0$, where RV_1 is the portfolio RV obtained from the model listed across the columns and RV_2 is the portfolio RV obtained from the model listed in the corresponding row.

	DCC	**Factor Double ARCH**	**LF Historical**	**HF Historical**
Panel A				
AA	110.40	148.35	132.84	100
	(0.005)	(<0.001)	(<0.001)	(1)
BA	107.99	128.04	140.75	100
	(0.003)	(0.003)	(0.003)	(1)
C	110.07	206.61	134.44	100
	(0.038)	(0.001)	(0.02)	(1)
GE	104.77	135.88	125.63	100
	(0.044)	(<0.001)	(<0.001)	(1)
HD	104.72	153.00	124.97	100
	(0.031)	(0.002)	(0.002)	(1)
HON	105.47	158.72	136.43	100
	(<0.001)	(<0.001)	(<0.001)	(1)
IBM	99.52	128.04	159.19	100
	(1)	(0.045)	(0.045)	(0.883)
INTC	111.95	141.82	161.25	100
	(0.009)	(0.009)	(0.009)	(1)
JPM	109.70	203.83	141.73	100
	(0.013)	(0.001)	(0.013)	(1)
KO	97.23	103.91	126.43	100
	(1)	(0.649)	(0.053)	(0.767)
MCD	109.41	122.22	146.12	100
	(0.001)	(0.001)	(0.001)	(1)
MO	100.61	100.62	130.99	100
	(0.792)	(0.792)	(0.233)	(1)
MRK	113.18	111.65	128.75	100
	(0.008)	(0.008)	(0.001)	(1)
MSFT	115.57	118.52	163.67	100
	(0.087)	(0.087)	(0.087)	(1)
PG	140.59	113.76	135.98	100
	(0.026)	(0.026)	(0.026)	(1)
WMT	105.40	135.43	132.12	100
	(0.001)	(0.001)	(0.001)	(1)
XOM	99.97	113.75	123.66	100
	(1)	(0.037)	(<0.001)	(0.995)
Panel B				
Factor Double ARCH	19.57			
LF Historical	-8.76	-21.32		
HF Historical	19.57	14.57	13.41	

Table 20: Panel A shows the RVs of the hedging portfolios obtained from the DCC, Factor Double ARCH, low-frequency (LF) historical and high-frequency (HF) historical covariance forecasts for the 180-day forecast horizon. Each RV is scaled by the RV of the HF historical covariance, which on average generates the lowest RV values, and multiplied by a factor of 100. Each row provides the RV for the portfolio hedging the stock listed in column 1 and under each value the MCS p-value is given in parentheses. Panel B shows the t-statistics from tests of H_0: RV_1 - $RV_2 = 0$, where RV_1 is the portfolio RV obtained from the model listed across the columns and RV_2 is the portfolio RV obtained from the model listed in the corresponding row.

	DCC	**Factor Double ARCH**	**LF Historical**	**HF Historical**
Panel A				
AA	109.34	150.79	114.47	100
	(0.001)	(<0.001)	(<0.001)	(1)
BA	109.73	126.50	118.00	100
	(0.001)	(<0.001)	(<0.001)	(1)
C	103.45	194.20	117.44	100
	(0.096)	(<0.001)	(0.013)	(1)
GE	102.69	133.07	115.02	100
	(<0.001)	(<0.001)	(<0.001)	(1)
HD	103.81	146.01	110.56	100
	(0.047)	(<0.001)	(0.013)	(1)
HON	105.07	156.83	123.95	100
	(<0.001)	(<0.001)	(<0.001)	(1)
IBM	104.56	132.67	146.02	100
	(0.019)	(0.007)	(0.019)	(1)
INTC	111.32	140.47	138.72	100
	(0.001)	(<0.001)	(0.001)	(1)
JPM	106.51	198.73	114.94	100
	(0.022)	(<0.001)	(0.01)	(1)
KO	101.41	109.24	104.70	100
	(0.733)	(0.477)	(0.477)	(1)
MCD	107.56	122.01	123.77	100
	(0.001)	(<0.001)	(0.001)	(1)
MO	99.09	101.28	104.00	100
	(1)	(0.458)	(0.458)	(0.458)
MRK	114.87	113.92	124.05	100
	(0.056)	(0.056)	(0.056)	(1)
MSFT	125.93	120.35	134.44	100
	(0.128)	(0.128)	(0.128)	(1)
PG	146.96	120.40	123.56	100
	(0.007)	(0.007)	(0.007)	(1)
WMT	104.41	134.11	117.68	100
	(0.004)	(<0.001)	(0.004)	(1)
XOM	100.79	114.40	120.15	100
	(0.834)	(0.03)	(0.03)	(1)
Panel B				
Factor Double ARCH	70.68			
LF Historical	-2.44	-32.08		
HF Historical	109.82	73.45	47.40	

References

C. Alexander and A. Barbosa. Hedging index exchange traded funds. *Journal of Banking & Finance*, 32:326–337, 2008.

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

T. Andersen and T. Bollerslev. Answering the skeptics: Yes standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905, 1998.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Great realisations. *Risk*, 18:105–108, 2000.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76, 2001a.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *The Journal of the American Statistical Association*, 96:42–55, 2001b.

T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1):101–143, 2003.

F. Bandi and J. R. Russell. Realized covariation, realized beta, and microstructure noise. *working paper, University of Chicargo*, 2005.

O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High-frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72 (3):885–925, 2004.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Working paper*, 2008a.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Realised kernels in practice: Trades and quotes. *Working paper*, 2008b.

B. J. Blair, S-H. Poon, and S. J. Taylor. Forecasting s&p 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, 102: 5–26, 2001.

T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72:498–505, 1990.

T. Bollerslev, R. F. Engle, and J. M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131, 1988.

M. Britten-Jones and A. Neuberger. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866, 2000.

C. Brownless and G. Gallo. Financial econometric analysis at ultra-high frequency: data handling concerns. *Computational Statistics & Data Analysis*, 51:22322245, 2006.

A. Buss and G. Vilkov. Option-implied correlation and factor betas revisited. *Working Paper*, 2008.

H. N. E. Byström. Using simulated currency rainbow options to evaluate covariance matrix forecasts. *Journal of International Markets, Institutions and Money*, 12:216–230, 2002.

J. Y. Campbell, A. W. Lo, and A. C. MacKinlay. *The econometrics of financial markets*. Princeton University Press, New Jersey, 1997.

B-Y. Chang, P. F. Christoffersen, K. Jacobs, and G. Vainberg. Option-implied measures of risk. *Working Paper*, 2009.

R. Chiriac and V. Voev. Long memory modelling of realized covariance matrices. *Working Paper, University of Konstanz*, 2007.

R. Chiriac and V. Voev. Modelling and forecasting multivariate realized volatility. *Working Paper, University of Konstanz*, 2008.

V. Chopra and W. Ziemba. The effect of error in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.

B. J. Christensen and N. R. Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50:125–150, 1998.

G. A. Christodoulakis and S. E. Satchell. Correlated ARCH (corrarch: Modelling the time-varying conditional correlation between financial asset returns). *European Journal of Operational Research*, 139:351–370, 2002.

K. J. Cohen, G. A. Hawawini, R. A. Schwartz, and D. K. Whitcomb. Friction in the trading process and the estimation of systematic risk. *Journal of Financial Economics*, 12:263–278, 1983.

P. Collin-Dufresne. A short introduction to correlation markets. *Journal of Financial Econometrics*, 7(1):12–29, 2009.

M. M. Dacarogna, R. Gencay, U. A. Muller, R. B. Olsen, and O. V. Pictet. *An introduction to high frequency finance*. Academic Press, Toronto, 2001.

S. R. Das. Basel ii: Correlation related issues. *Journal of Financial Services Research*, 32:17–38, 2007.

M de Pooter, M. Martens, and D. van Dijk. Predicting the daily covariance matrix for s&p 100 stocks using intraday data - but which frequency to use? *Econometric Reviews*, 27(1-3):199–229, 2008.

F. X. Diebold and J. A. Lopez. Forecast evaluation and combination. In C. S. Maddala and C. R. Rao, editors, *Handbook of Statistics*. 1996.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

J. Driessen, P. Maenhout, and G. Vilkov. Option-implied correlations and the price of correlation risk. *The Journal of Finance, forthcoming*, 2009.

P. Embrechts. Linear correlation and EVT:Properties and caveats. *Journal of Financial Econometrics*, 7(1):30–39, 2009.

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

R. F. Engle. *Anticipating correlations a new paradigm for risk management*. Princeton University Press, Princeton, New Jersey, 2009a.

R. F. Engle. *High dimensional dynamic correlations*. Oxford University Press, 2009b.

R. F. Engle and R. Colacito. Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2):238–253, 2006.

R. F. Engle and K. F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.

R. F. Engle and J. Mezrich. GARCH for groups. *Risk*, 8:36–40, 1996.

R. F. Engle and K. Sheppard. Theoretcial and empirical properties of dynamic conditional correlation multivariate GARCH. *Working Paper*, 2001.

R. F. Engle, N. Shephard, and K. Sheppard. Fitting and testing vast dimensional time-varying covariance models. *Working Paper*, 2008.

T. W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298, 1979.

T. N. Falkenberry. High frequency data filtering. *Technical Report, Tick Data*, 2002.

J. Fleming. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance*, 5:317–345, 1998.

J. Fleming, C. Kirby, and B. Ostdiek. The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67:473–509, 2003.

A. R. Gallant, C-T. Hsu, and G. Tauchen. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *The Review of Economics and Statistics*, 81(4): 617–631, 1999.

M. S. Gibson and B. Boyer. Evaluating forecasts of correlation. *The Journal of Derivatives*, Winter: 18–38, 1998.

L. R. Glosten, R. Jagannathan, and D. Runkle. Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48:1779–1801, 1993.

P. R. Hansen. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4):365–380, 2005.

P. R. Hansen and A. Lunde. A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554, 2005.

P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Working Paper*, 2009.

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

G. J. Jiang and Y. S. Tian. The model-free implied volatility and its information content. *The Review of Financial Studies*, 18(4):1305–1342, 2005.

P. Jorion. Predicting volatility in the foreign exchange market. *Journal of Finance*, 50(2):507–528, 1995.

S. J. Koopman, B. Jungbacker, and E. Hol. Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12: 445–475, 2005.

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

D. Lien, Y. K. Tse, and A. K. C. Tsui. Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Econometrics*, 12:791–798, 2002.

J. M. Mahoney. Risk management of correlation products. *European Financial Management*, 3(2): 155–174, 1997.

A. M. Malz. Estimating the probability distribution of the future exchange rate from option prices. *Journal of Derivatives*, 5:18–36, 1997a.

A. M. Malz. Option-implied probability distributions and currency excess returns. *Staff Report, Federal Reserve Bank of New York*, 1997b.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

M. Martens. Measuring and forecasting s&p 500 index-futures volatility using high-frequency data. *The Journal of Futures Markets*, 22(6):497–518, 2002.

M. Martens. Estimating unbiased and precise realized covariances. *Working Paper*, 2004.

M. Martens and J. Zein. Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, 24(11):1005–1028, 2004.

A. J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, forthcoming*, 2008.

A. J. Patton and K. Sheppard. Evaluating volatility and correlation forecasts. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. 2009.

A. J. Patton and A. Timmermann. Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480):1172–1184, 2007.

S. Pong, M. B. Shackleton, S. J. Taylor, and X. Xu. Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking & Finance*, 28:2541–2563, 2004.

S. Pong, M. B. Shackleton, and S. J. Taylor. Distinguishing short and long memory volatility specifications. *Econometrics Journal*, 11:617–637, 2008.

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

S-H. Poon and C. W. J. Granger. Practical issues in forecasting volatility. *Financial Analysts Journal*, 61(1):45–56, 2005.

M. Scholes and J. Williams. Estimating beta from non-synchronous data. *Journal of Financial Economics*, 5:309–327, 1977.

K. Sheppard. Realized covariance and scrambling. *Working paper*, 2006.

V. D. Skintzi and A-P. N. Refenes. Implied correlation index: A new measure of diversification. *The Journal of Futures Markets*, 25(2):171–197, 2005.

V. D. Skintzi and S. Xanthopoulos-Sisinis. Evaluation of correlation forecasting models for risk management. *Journal of Forecasting*, 26:497–526, 2007.

S. J. Taylor. *Asset price dynamics, volatility, and prediction*. Princeton University Press, New Jersey, 2005.

S. J. Taylor, P. K. Yadav, and Y. Zhang. Cross-sectional analysis of risk-neutral skewness. *The Journal of Derivatives*, Summer:38–52, 2009.

V. Voev. Dynamic modelling of large dimensional covariance matrices. *Working Paper, University of Konstanz*, 2007.

K. D. West. Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084, 1996.

H. White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000.

L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Working paper, University of Illinois*, 2006.

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