Evaluating Volatility and Correlation Forecasts[*]

Andrew J. Patton
Department of Economics
and
Oxford-Man Institute of Quantitative Finance
University of Oxford
`andrew.patton@economics.ox.ac.uk`

Kevin Sheppard
Department of Economics
and
Oxford-Man Institute of Quantitative Finance
University of Oxford
`kevin.sheppard@economics.ox.ac.uk`

October 29, 2007

1   Introduction

This chapter considers the problem of evaluation and comparison of univariate and multivariate volatility forecasts, with explicit attention paid to the fact that in such applications the object of interest is unobservable, even *ex post*. Thus the evaluation and comparison of volatility forecasts must rely on direct or indirect methods of overcoming this difficulty. Direct methods use a "volatility proxy", i.e. some observable variable that is related to the latent variable of interest. We will assume the existence of an unbiased volatility proxy, such as daily squared returns for the daily conditional variance of returns. Indirect methods of overcoming the latent nature of the variable of interest include comparing forecasts via mean-variance portfolio decisions or comparisons based on portfolio "tracking error".

A guiding theme of our analysis is the focus on tests that are "robust" in two ways: Firstly, we want tests that are robust to the presence of noise in the volatility proxy, if a proxy is used. The presence of noise in the proxy may affect – indeed, is likely to affect – the power of the tests, but it should not distort the asymptotic size of the test. Secondly, we desire tests that are robust to the conditional distribution of returns. Some existing volatility forecast evaluation and comparison techniques rely, in an important way, on assumptions about the conditional distribution of returns beyond the conditional second moment. While these techniques may have strong economic motivations, and thus be of interest in their own right,

we argue that they are not pure tests of *volatility* forecasts and that they should not be used without a convincing economic argument. Such arguments will generally be specific to each application.

The main results and recommendations of this chapter, based on theoretical considerations and on simulation studies, can be summarised as follows. Firstly, we suggest a minor modification of the widely-used Mincer-Zarnowitz regression for testing volatility forecast optimality which exploits the additional structure that holds under the null hypothesis. This "MZ-GLS" test has good size and much better finite sample power than other MZ tests. Secondly, we find that the use of loss functions that are "non-robust", in the sense of Patton [52], can yield perverse rankings of forecasts, even when accurate volatility proxies are employed. This emphasises the need to pay careful attention to the selection of the loss function in Diebold-Mariano [23] and West [57] tests when evaluating volatility forecasts. Amongst the class of robust loss functions for volatility forecast evaluation, and the multivariate generalisation of these loss functions provided in this chapter, our simulations suggest that the "QLIKE" loss function yields the greatest power. Finally, consistent with the large and growing literature on realised volatility, our simulations clearly demonstrate the value of higher-precision volatility proxies, such as realised variance or daily high-low range [5]. Even simple estimators based on 30-minute returns provide large gains in power and improvements in finite-sample size.

The problems which arise in evaluating a single volatility forecast, a set of volatility forecasts, or a complete covariance matrix are so similar that dividing the chapter into univariate and multivariate sections would be misleading. This chapter avoids the false dichotomy between univariate and multivariate volatility forecasting and evaluation as much as possible. As a results, this chapter is primarily organised along the method of evaluation, not the dimension of the problem studied. Sections 2 and 3 of this chapter focus on direct methods for forecast evaluation and comparison using a volatility proxy, while Sect. 4 discusses indirect, or economic, methods.

1.1   Notation

We will present the notation for the general multivariate case first and specialise the notation for the univariate case where needed. Let $\mathbf{r}_t \equiv [r_{1t},\ r_{2t}, \ldots , r_{Kt}]'$ be the $K \times 1$ vector valued variable whose conditional covariance is of interest. The information set used in defining the conditional variance is denoted $\mathcal{F}_{t-1}$, and is assumed to contain the history of past returns, but may also include other variables and/or variables measured at a higher frequency than $\mathbf{r}_t$, such as intra-daily returns. Denote $V\left[\mathbf{r}_t|\mathcal{F}_{t-1}\right] \equiv V_{t-1}\left[\mathbf{r}_t\right] \equiv \mathbf{\Sigma}_t$, a $K \times K$ symmetric, positive definite matrix composed of elements $\sigma_{ij,t}$ where $\sigma_{ii,t} \equiv \sigma_{i,t}^2$ denotes the conditional variance of the ith return and $\sigma_{ij,t}$ denotes the conditional covariance between the ith and j^{th} series. We will assume throughout that $E\left[\mathbf{r}_t|\mathcal{F}_{t-1}\right] \equiv E_{t-1}\left[\mathbf{r}_t\right] = \mathbf{0}$, and thus $\mathbf{\Sigma}_t = E_{t-1}\left[\mathbf{r}_t\mathbf{r}_t'\right]$. Let $\varepsilon_t \equiv \mathbf{\Sigma}_t^{-1/2}\mathbf{r}_t$ denote the "standardised vector of returns", where $\mathbf{\Sigma}_t^{1/2}$ is a matrix that satisfies $\mathbf{\Sigma}_t^{1/2'}\mathbf{\Sigma}_t^{1/2} = \mathbf{\Sigma}_t$.[1] We assume that

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathbf{F}_t\left(\mathbf{0}, \mathbf{\Sigma}_t\right) \tag{1}$$

where $\mathbf{F}_t$ is some distribution with zero mean and finite covariance $\mathbf{\Sigma}_t$. In some applications we will use a stronger condition that $\mathbf{r}_t|\mathcal{F}_{t-1}$ has a constant conditional distribution and hence constance higher order moments, i.e.

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathbf{F}\left(\mathbf{0}, \mathbf{\Sigma}_t\right) . \tag{2}$$

---

[1]For example, the "square root" matrix, $\mathbf{\Sigma}_t^{1/2}$, can be based on the Cholesky or the spectral decomposition of $\mathbf{\Sigma}_t$. The Cholesky square-root is not invariant to the order of the variables in $\mathbf{r}_t$, and so any subsequent computations may change following a simple re-ordering of these variables. For this reason we recommend the use of the square-root based on the spectral decomposition.

Let a forecast of the conditional covariance of $\mathbf{r}_t$ be denoted $\mathbf{H}_t$, or $\mathbf{H}_t^A, \mathbf{H}_t^B, \mathbf{H}_t^C, \ldots$ if there multiple forecasts are under analysis. The loss function in the multivariate case is $L : \mathbb{M}_+^K \times \mathcal{H}^K \to \mathbb{R}_+$, where the first argument of L is $\mathbf{\Sigma}_t$ or some proxy for $\mathbf{\Sigma}_t$, denoted $\hat{\mathbf{\Sigma}}_t$, and the second is $\mathbf{H}_t$. $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the non-negative and positive real line, $\mathcal{H}^K$ is a compact subset of $\mathbb{M}_{++}^K$, and $\mathbb{M}_+^K$ and $\mathbb{M}_{++}^K$ denote the positive semi-definite and positive definite subspaces of the set of all real symmetric $K \times K$ matrices. Note that $\mathbb{M}_+^1 = \mathbb{R}_+$ and $\mathbb{M}_{++}^1 = \mathbb{R}_{++}$.

Commonly used univariate volatility proxies are the squared return, r_t^2, realised volatility based on m intra-daily observations ("m-sample RV"), $RV_t^{(m)}$, and the range, RG_t, while commonly used covariance proxies are the outer product of returns, $\mathbf{r}_t \mathbf{r}_t'$ and realised covariance, $RC_t^{(m)}$. In this chapter we will treat the forecasts as "primitive", and make no attempt to study them via the models, if any, that are used to generate the forecasts. Specification, estimation and selection of univariate and multivariate volatility models is considered in [63, 54, 42].

We define a conditionally unbiased volatility proxy, denoted $\hat{\sigma}_t^2$, and a conditionally unbiased covariance proxy, denoted $\hat{\mathbf{\Sigma}}_t = [\hat{\sigma}_{ij}]_t$ as variables that satisfy:

$$E\left[\hat{\sigma}_t^2 | \mathcal{F}_{t-1}\right] = \sigma_t^2 \ \ a.s., \ t = 1, 2, .. \tag{3}$$

$$E\left[\hat{\mathbf{\Sigma}}_t | \mathcal{F}_{t-1}\right] = \mathbf{\Sigma}_t \ \ a.s., \ t = 1, 2, .. \tag{4}$$

We will assume that at least one such proxy is available in all cases, though we make no further assumptions about the accuracy or consistency of the proxy.

2 Direct Evaluation of Volatility Forecasts

In this section we review tests for the evaluation of volatility forecasts using a volatility proxy. Recalling that $\sigma_t^2 \equiv V\left[r_t | \mathcal{F}_{t-1}\right]$ and drawing on White [61, Definition 5.2], we define an optimal univariate volatility forecast as one that satisfies the following null hypothesis:

$$H_0^* \ : \ h_t = \sigma_t^2 \ \ a.s., \ t = 1, 2, \ldots \tag{5}$$
$$\text{vs. } H_1^* \ : \ h_t \neq \sigma_t^2 \text{ for some } t$$

The corresponding null for multivariate volatility forecasts is:

$$H_0^* \ : \ \mathbf{H}_t = \mathbf{\Sigma}_t \ \ a.s., \ t = 1, 2, \ldots \tag{6}$$
$$\text{vs. } H_1^* \ : \ \mathbf{H}_t \neq \mathbf{\Sigma}_t \text{ for some } t$$

The above null hypotheses are the ones that would ideally be tested in forecast evaluation tests. Instead, simple implications of these hypotheses are usually tested; we review the most common tests below.

2.1 Forecast Optimality Tests for Univariate Volatility Forecasts

One common method of evaluating forecasts is the Mincer-Zarnowitz, or MZ, regression [48], which involves regressing the realisation of a variable on its forecast (see also [55]). However, unlike standard forecast evaluation problems, the conditional variance is never observed, and the usual MZ regression is infeasible for volatility forecast evaluation. Using a conditionally unbiased estimator of the conditional variance, the feasible MZ regression:

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + e_t \tag{7}$$

yields unbiased estimates of α and β. The usual MZ test can then be conducted:

$$H_0 \quad : \quad \alpha = 0 \cap \beta = 1 \tag{8}$$
$$\text{vs. } H_1 \quad : \quad \alpha \neq 0 \cup \beta \neq 1$$

The OLS parameter estimates will be less accurately estimated the larger the variance of $\left(\sigma_t^2 - \hat{\sigma}_t^2\right)$, which suggests the use of high frequency data to construct more accurate volatility proxies [3]. While using less accurate estimates of $\hat{\sigma}_t^2$ affects the precision of α and β, and thus the power of the test to detect deviations from forecast optimality, it does not affect the validity of the test.[2]

The standard MZ regression can detect certain deviations from H_0^* but is not consistent against all possible deviations. While it is possible to construct consistent MZ test (see [10], [22] [11]), so-called "augmented MZ regressions", constructed using additional $\mathcal{F}_{t-1}$-measurable instruments to increase the power of the test in certain directions, are more common. Standard instruments include the lagged volatility proxy, $\hat{\sigma}_{t-1}^2$, the lagged standardised volatility proxy, $\hat{\sigma}_{t-1}^2/h_{t-1}$, sign based indicators, $I_{[r_{t-1}<0]}$, or combinations of these. These instruments are motivated by a desire to detect neglected nonlinearities or persistence in the volatility forecast. Grouping the set of $\mathcal{F}_{t-1}$-measurable instruments into a vector $\mathbf{z}_{t-1}$, the augmented MZ regression, null and alternative hypotheses are

$$\hat{\sigma}_t^2 \quad = \quad \alpha + \beta h_t + \boldsymbol{\gamma}'\mathbf{z}_{t-1} + e_t \tag{9}$$
$$H_0 \quad : \quad \alpha = 0 \cap \beta = 1 \cap \boldsymbol{\gamma} = \mathbf{0}$$
$$\text{vs. } H_1 \quad : \quad \alpha \neq 0 \cup \beta \neq 1 \cup \boldsymbol{\gamma} \neq \mathbf{0} .$$

Note that the residual of the above regression will generally be heteroskedastic, even under H_0^*, and so robust standard errors [60] are required. This prompted some authors to consider a different approach: testing for serial correlation in the standardised volatility proxy:

$$\frac{\hat{\sigma}_t^2}{h_t} \quad = \quad \delta + \vartheta\frac{\hat{\sigma}_{t-1}^2}{h_{t-1}} + u_t \tag{10}$$
$$H_0 \quad : \quad \delta = 1 \cap \vartheta = 0$$
$$\text{vs. } H_1 \quad : \quad \delta \neq 1 \cup \vartheta \neq 0$$

This approach generates residuals that are homoskedastic under H_0^* if the noise in the proxy $\left(\eta_t \equiv \hat{\sigma}_t^2/\sigma_t^2\right)$ has constant conditional variance[3], and so if this assumption holds then robust standard errors are not required.

2.2 MZ Regressions on Transformations of $\hat{\sigma}_t^2$

The use of squared returns in MZ regressions has caused some researchers concern, as statistical inference relies on fourth powers of the returns, and thus returns that are large in magnitude have a large impact on test results. One frequently proposed alternative is to use transformations of the volatility proxy and forecast to reduce the impact of large returns (see [41, 13]). Two such examples are:

$$|r_t| \quad = \quad \alpha + \beta\sqrt{h_t} + e_t, \text{ and} \tag{11}$$
$$\log\left(r_t^2\right) \quad = \quad \alpha + \beta\log(h_t) + e_t \tag{12}$$

---

[2]Chernov [15] provides a detailed analysis of the implications of measurement error in the *regressors*, such as the $\mathbf{z}_{t-1}$ variables in (9).

[3]This corresponds to conditional homokurticity, discussed below, of the returns if the volatility proxy used is a squared return.

Using these regressions can result in size distortions, even asymptotically, due to noise in the volatility proxy, $\hat{\sigma}_t^2$. Take the regression in (11) as an example: under H_0^* the population values of the OLS parameter estimates are easily shown to be:

$$\alpha = 0$$

$$\beta = \begin{cases} E\left[|\varepsilon_t|\right], & \text{if } E_{t-1}\left[|\varepsilon_t|\right] \text{ is constant} \\ \sqrt{\frac{v-2}{\pi}}\Gamma\left(\frac{v-1}{2}\right)/\Gamma\left(\frac{v}{2}\right), & \text{if } r_t|\mathcal{F}_{t-1} \sim \text{Student's } t\left(0, \sigma_t^2, v\right), v > 2 \\ \sqrt{2/\pi} \approx 0.80, & \text{if } r_t|\mathcal{F}_{t-1} \sim N\left(0, \sigma_t^2\right) \end{cases}$$

where $\text{Student's } t\left(0, \sigma_t^2, v\right)$ is a Student's t distribution with mean zero, variance σ_t^2 and v degrees of freedom. When returns have the Student's t distribution, the population value for β decreases towards zero as $v \downarrow 2$, indicating that excess kurtosis in returns increases the distortion in this parameter. In the regression in (12) the population OLS parameters under H_0^* are:

$$\alpha = \begin{cases} E\left[\log \varepsilon_t^2\right], \\ \log\left(v - 2\right) + \Psi\left(\frac{1}{2}\right) - \Psi\left(\frac{v}{2}\right), & \text{if } r_t|\mathcal{F}_{t-1} \sim \text{Student's } t\left(0, \sigma_t^2, v\right), v > 2 \\ -\log\left(2\right) - \gamma_E \approx -1.27, & \text{if } r_t|\mathcal{F}_{t-1} \sim N\left(0, \sigma_t^2\right) \end{cases}$$

$$\beta = 1$$

where Ψ is the digamma function and $\gamma_E = -\Psi\left(1\right) \approx 0.58$ is Euler's constant [39]. Under the Student's t distribution, the above expression shows $\alpha \rightarrow -\infty$ as $v \downarrow 2$. Thus while both of these alternative MZ regressions may initially appear reasonable, without some modification they lead to the undesirable outcome that the perfect volatility forecast, $h_t = \sigma_t^2 \text{ } a.s.$, will be rejected with probability approaching one as the sample size increases. In both cases the perverse outcomes are the result of the imperfect nature of any volatility proxy; if volatility was observable, regressions on the transformation would lead to the correct result.

A second alternative is to adjust the volatility proxy, either exactly or approximately, so as to make it unbiased for the quantity of interest, σ_t or $\log \sigma_t$, and thus avoid any asymptotic size distortions [13, 18, 6]. However, these adjustments require further assumptions about the distribution of the noise in the volatility proxy, and test statistics may be misleading if the assumptions are violated.

2.3 Forecast Optimality Tests for Multivariate Volatility Forecasts

The results for testing optimality of conditional volatility can be directly extended to conditional covariance forecasts. The simplest optimality test examines the unique elements of the forecast covariance $\mathbf{H}_t$ separately using feasible MZ regressions

$$\hat{\sigma}_{ij,t} = \alpha_{ij} + \beta_{ij} h_{ij,t} + e_{ij,t} \tag{13}$$

or augmented MZ regressions

$$\hat{\sigma}_{ij,t} = \alpha_{ij} + \beta_{ij} h_{ij,t} + \boldsymbol{\gamma}'_{ij}\mathbf{z}_{ij,t} + e_{ij,t} . \tag{14}$$

resulting in $K(K + 1)/2$ regressions and test statistics. This may be problematic, particularly when K is large as some rejections are expected even if the conditional covariance is correct. Alternatively, a joint test,

that all of the coefficients are simultaneously zero can be tested the by forming a vector process using the half-vector operator[4] (vech),

$$\text{vech}(\hat{\Sigma}_t) = \boldsymbol{\alpha} + \text{diag}(\boldsymbol{\beta})\mathbf{H}_t + \boldsymbol{\varepsilon}_t \tag{15}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $K(K+1)/2$ parameter vectors and diag is the diagonal operator[5]. Explicit expressions for a heteroskedasticity consistent covariance estimator can be found in many panel data texts (see, e.g. [7]). Despite availability of a joint test, the finite sample properties may be adversely affected by the large dimension of the parameter covariance matrix. A simple specification could be constructed using only a common parameter across all pairs,

$$\text{vech}(\hat{\Sigma}_t) = \alpha + \beta\,\text{vech}\,(\mathbf{H}_t) + \boldsymbol{\varepsilon}_t \tag{16}$$

and testing whether α and β are 0 and 1, respectively[6].

2.4 Improved MZ Regressions Using Generalised Least Squares

The residuals from the feasible MZ and augmented MZ regressions above will generally be heteroskedastic, and the size and power properties of these tests can be improved using generalised least squares. Consider a decomposition of the volatility proxy into the true variance and a multiplicative error term: $\hat{\sigma}_t^2 \equiv \sigma_t^2 \eta_t$, where $E_{t-1}\left[\eta_t\right] = 1$, and the feasible univariate MZ regression from above:

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + e_t$$

Under H_0^*, residuals from this regression will be

$$
\begin{aligned}
e_t &= \hat{\sigma}_t^2 - h_t = \sigma_t^2\left(\eta_t - 1\right) \\
\text{so } E_{t-1}\left[e_t\right] &= 0 \\
\text{but } V_{t-1}\left[e_t\right] &= \sigma_t^4 V_{t-1}\left[\eta_t\right] \equiv \varsigma_t^2
\end{aligned}
$$

If ς_t^2 was known for all t, then GLS estimation of the feasible MZ regression would simply be

$$\frac{\hat{\sigma}_t^2}{\varsigma_t} = \alpha\frac{1}{\varsigma_t} + \beta\frac{h_t}{\varsigma_t} + \tilde{e}_t \; . \tag{17}$$

In the special case where the volatility proxy is a squared return, $\hat{\sigma}_t^2 = r_t^2$ and where the standardised returns are *conditionally homokurtic*[7], i.e. $E_{t-1}\left[r_t^4\right]/\sigma_t^4 = \kappa \,\forall\, t$, the MZ-GLS takes a very simple form:

$$\frac{\hat{\sigma}_t^2}{h_t} = \alpha\frac{1}{h_t} + \beta + \tilde{e}_t \; . \tag{18}$$

since $V_{t-1}[\hat{\sigma}_t^2] = \sigma_t^4\kappa = h_t^2\kappa \propto h_t^2$ under H_0^*.

---

[4]This operator stacks the columns of the lower triangle of a square matrix [45].

[5]The diagonal operator transforms a $K \times 1$ vector into a $K \times K$ matrix with the vector along the diagonal. That is, $\text{diag}(\boldsymbol{\beta}) = \left(\boldsymbol{\beta}\boldsymbol{\iota}'\right) \odot \mathbf{I}_K$, where $\boldsymbol{\iota}$ is a $K \times 1$ vector of ones, $\mathbf{I}_K$ is the K-dimensional identity matrix, and $\odot$ represents the Hadamard product (for element-by-element multiplication).

[6]While the parameters of this regression can be estimated using OLS by stacking the elements of $\text{vech}(\hat{\Sigma}_t)$, the errors will generally be cross-sectionally correlated and White standard errors will not be consistent. Instead, a pooled-panel covariance estimator appropriate for cross-correlated heteroskedastic data should be used.

[7]This holds, for example, if returns are conditionally Normally distributed, or conditionally Student's t distributed with constant degrees of freedom greater than 4.

This simple standardisation is not guaranteed to produce efficient estimates for arbitrary volatility proxies, for example realised variance or range. However, it is generally the case that the volatility of the proxy is increasing in the level of the proxy. Thus using the specification in (18) may result in improved finite sample performance, even when not fully efficient. A formal proof of this conjecture is left to future research[8].

The application of MZ-GLS to covariance forecast evaluation is similar although there are choices which may affect the finite sample properties of the parameter estimators. The direct extension of the univariate framework specifies the MZ-GLS as

$$\frac{\hat{\sigma}_{ij,t}}{\varsigma_{ij,t}} = \alpha_{ij}\frac{1}{\varsigma_{ij,t}} + \beta_{ij}\frac{h_{ij,t}}{\varsigma_{ij,t}} + \tilde{e}_{ij,t} \ . \tag{19}$$

where $\varsigma_{ij,t} \equiv V_{t-1}[\hat{\sigma}_{ij,t}]$, and again $\alpha_{ij} = 0 \cap \beta_{ij} = 1$ under H_0^*. If the volatility proxies are squares or cross-products of conditionally homokurtic returns, the MZ-GLS can again be specified using the forecasts as weights

$$\frac{\hat{\sigma}_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t} + h_{ij,t}^2}} = \alpha_{ij}\frac{1}{\sqrt{h_{ii,t}h_{jj,t} + h_{ij,t}^2}} + \beta_{ij}\frac{h_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t} + h_{ij,t}^2}} + \tilde{e}_{ij,t} \tag{20}$$

The denominator of the left-hand-side of the above equation can be written as $\sqrt{h_{ii,t}h_{jj,t}}\sqrt{1 + \varrho_{ij,t}^2}$. It may be noted that the contribution to the heteroskedasticity of $\hat{\sigma}_{ij,t}$ from the $\sqrt{1 + \varrho_{ij,t}^2}$ term is generally small since this is bounded between 1 and $\sqrt{2}$. A slightly simpler specification for evaluating conditional covariance models that accounts for the largest contributors to the heteroskedasticity of $\hat{\sigma}_{ij,t}$ can be specified as

$$\frac{\hat{\sigma}_{ij,t}}{\sqrt{h_{ii,t}h_{jj,t}}} = \alpha_{ij}\frac{1}{\sqrt{h_{ii,t}h_{jj,t}}} + \beta_{ij}\varrho_{ij,t} + \tilde{e}_{ij,t} \ . \tag{21}$$

and can be interpreted as a regression of the correlation proxy on the forecast correlation[9] [56]. When $i = j$ all of these specifications reduce to the volatility MZ-GLS in (18).

2.5 Simulation Study

To assess the size and power properties of the MZ and MZ-GLS tests, we compare variations on two methods for evaluating the performance of volatility and covariance forecasts. The first method (MZ1) involves regressing a volatility proxy (either squared daily returns, or a realised volatility proxy constructed to resemble a realised variance based on 30-minute or 5-minute returns) on a constant and the forecast:

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + e_t$$

---

[8]In the presence of intra-daily heteroskedasticity, the variance of the proxy will not generally be proportional to the conditional variance and so a direct estimate will be required. If the proxy is realised variance, such an estimator will require an estimate of the "integrated quarticity", see [8]. For application of GLS, the estimator needs to be consistent in T, thus requiring that $m = O\left(T^{\delta}\right)$ for $\delta \geq 1$, with the specific rate δ depending on assumptions about the underlying diffusion. In finite samples, integrated quarticity is often estimated with substantial error, and so scaling by a consistent estimator of the proxy variance, while accurate asymptotically, may perform worse than simply scaling by the forecast.

[9]Strictly speaking, $\frac{\hat{\sigma}_{ij,t}}{\sqrt{\hat{\sigma}_{i,t}^2\hat{\sigma}_{j,t}^2}}$ would be the correlation proxy, however this proxy is not generally unbiased since it is a non-linear function of unbiased estimators. The modification in (21) is a compromise that makes use of the volatility forecasts, $h_{ii,t}$ and $h_{jj,t}$, which are unbiased and error free under H_0^*.

and then testing

$$H_0 \quad : \quad \alpha = 0 \cap \beta = 1$$
$$\text{vs } H_1 \quad : \quad \alpha \neq 0 \cup \beta \neq 1$$

This regression will have heteroskedastic errors under H_0^* and the covariance of the OLS parameters must be estimated using a heteroskedasticity consistent estimator, such as White's robust covariance estimator [60]. Alternatively, as discussed above, the volatility forecasts themselves can be used to obtain homoskedastic regression residuals via GLS. In this case we use OLS to estimate

$$\frac{\hat{\sigma}_t^2}{h_t} = \alpha \frac{1}{h_t} + \beta + \tilde{e}_t$$

We denote these two approaches as MZ1 (White) and MZ1-GLS. The proxy used, either daily returns squared or realised variance is denoted $RV^{(m)}$, where m is the number of intra-daily returns used to construct the realised variance. The second approach (MZ2) uses the standardised proxy in the regression, and has residuals that are homoskedastic under H_0^*, as long as the noise in the proxy $\left(\eta_t \equiv \hat{\sigma}_t^2 / \sigma_t^2\right)$ has constant conditional variance[10]:

$$\frac{\hat{\sigma}_t^2}{h_t} = \delta + \vartheta \frac{\hat{\sigma}_{t-1}^2}{h_{t-1}} + u_t \tag{22}$$

and then tests

$$H_0 \quad : \quad \delta = 1 \cap \vartheta = 0$$
$$\text{vs } H_1 \quad : \quad \delta \neq 1 \cup \vartheta \neq 0$$

While the MZ2 regression has an intuitive appeal, we are not aware of any studies of its performance relative to the MZ1 regressions in a realistic scenario.

The data generating process used in the simulation was specified as a standard GARCH(1,1)

$$DGP \quad : \quad r_t = \sigma_t \varepsilon_t, \, t = 1, 2, \ldots, T \tag{23}$$
$$\text{where } E_{t-1}[\varepsilon_t] \quad = \quad 0 \text{ and } E_{t-1}\left[\varepsilon_t^2\right] = 1$$
$$\sigma_t^2 \quad = \quad 0.05 + 0.85\sigma_{t-1}^2 + 0.10r_{t-1}^2 \tag{24}$$
$$T \quad = \quad \{100, 250, 500, 1000\}$$

We specify the distribution of the standardised innovations as:

$$\varepsilon_t \quad = \quad \sum_{m=1}^{78} \xi_{mt} \tag{25}$$
$$\xi_{mt} \quad \overset{\text{IID}}{\sim} \quad N(0, 1/78)$$

The distribution of the standardised innovations is designed to allow for DGP-consistent high-frequency estimators to be computed while preserving the GARCH(1,1) specification for the variance. The volatility proxies are computed from the ξ_{mt} as

$$RV_t^{(m)} = \sigma_t^2 \sum_{i=1}^{m} \left(\sum_{j=\lambda(i-1)+1}^{\lambda i} \xi_{it} \right)^2 \tag{26}$$

---

[10]This corresponds to conditional homokurticity of the returns if the volatility proxy used is a squared return, and if the return has conditional mean zero.
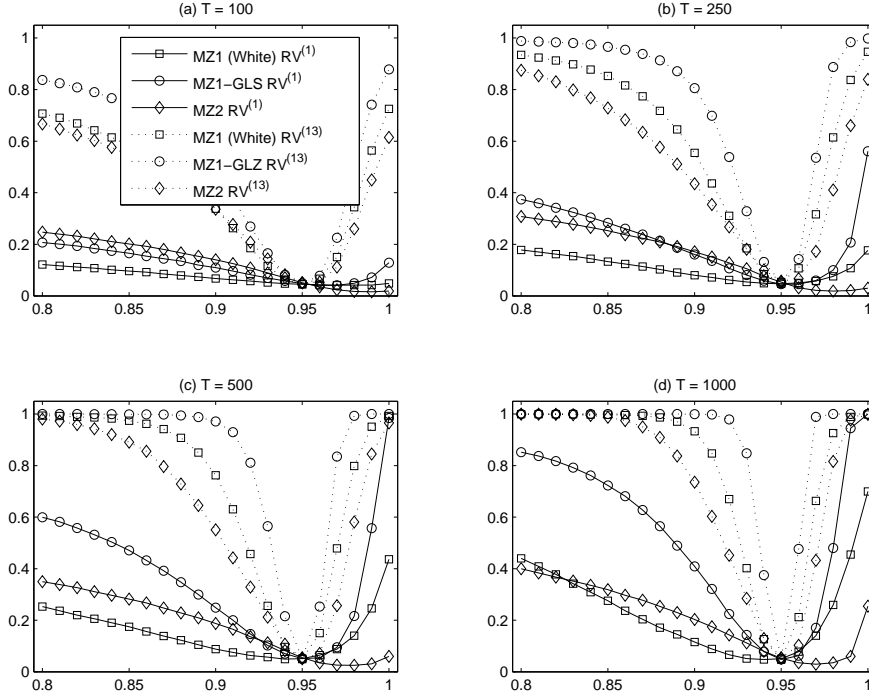
Figure 1: The four panels of this figure contain the size-adjusted power for MZ1 and MZ2 tests applied to volatility forecasts. Two proxies were used for the latent volatility: daily squared returns ($RV^{(1)}$, *solid line*) and a 13-sample realised variance estimator ($RV^{(13)}$, *dotted line*). MZ1 tests regress the volatility proxy on the forecast using either OLS, indicated by □, or GLS, indicated by ○, and MZ2 tests regress the standardised volatility proxy on a lag of the standardised volatility proxy, indicated by ◊

where $\lambda = 78/m$. We considered three values for m, the number of intra-daily returns. $m = 1$ corresponds to the use of daily squared returns as a proxy; $m = 13$ corresponds to using half-hourly returns for a stock traded on the NYSE; and $m = 78$ corresponds to using five-minute returns for a stock traded on the NYSE.

The forecast model was designed to reflect the fact that most volatility forecasting models are able to closely match the unconditional variance. Forecasts predominantly differ in their specification for the dynamics of conditional variance. The model used is also a GARCH(1,1), with the correct unconditional variance, but with differing persistence. The setup is such that the ratio of the coefficient on r_{t-1}^2 to the coefficient on h_{t-1} is always equal to 10/85. We take these parameters as fixed, and do not consider the forecaster's problem of estimating the model parameters from the data. The null corresponds to the case that $k = 0.95$.

$$Model \quad : \quad h_t = (1-k) + \frac{0.85}{0.95}k \times h_{t-1} + \frac{0.10}{0.95}k \times r_{t-1}^2 \tag{27}$$
$$k = \{0.80, 0.81, \ldots, 0.99, 1\}$$

We studied varying degrees of persistence (k) and 4 different sample sizes (T) designed to reflect realistic evaluation samples, ranging from 100 to 1000, each with 10,000 replications. The power curves are presented in Fig. 1, and the finite-sample size properties are summarised in Table 1.

Three main conclusions can be drawn from our small Monte Carlo study: Firstly, MZ1 tests are generally more powerful than the MZ2 tests, particularly for larger sample sizes. Secondly, of the MZ1 tests, MZ1-GLS

9

Size Comparison of MZ-type Tests for Volatility Forecasts						
	T = 100			T = 250		
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
MZ1 (White)	0.23	0.09	0.07	0.16	0.07	0.07
MZ1-GLS	0.11	0.07	0.06	0.08	0.06	0.05
MZ1-GLS (White)	0.15	0.07	0.06	0.10	0.06	0.06
MZ2	0.07	0.06	0.05	0.05	0.05	0.05
MZ2 (White)	0.15	0.07	0.07	0.10	0.06	0.06
	T = 500			T = 1000		
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
MZ1 (White)	0.12	0.07	0.06	0.10	0.06	0.06
MZ1-GLS	0.06	0.05	0.05	0.06	0.05	0.05
MZ1-GLS (White)	0.07	0.05	0.05	0.06	0.05	0.05
MZ2	0.05	0.05	0.05	0.05	0.05	0.05
MZ2 (White)	0.08	0.05	0.06	0.07	0.05	0.05

Table 1: Rejection frequency of test statistics using a nominal size of 5% based on asymptotic critical values. All test statistics were computed using Wald tests using either White's heteroskedasticity-consistent covariance estimator or the standard OLS variance covariance formula as indicated. Three proxies were used for the latent volatility: the daily squared returns ($RV^{(1)}$), a 13-sample realised variance estimator ($RV^{(13)}$) and a 78-sample realised variance ($RV^{(78)}$)

has better power, often substantially, than the standard feasible MZ test. The difference between MZ1 and MZ1-GLS was particularly striking when using the daily squared return as a proxy. Furthermore, it has good size properties even in small samples. That MZ1-GLS has better size properties than tests relying on robust standard errors is not too surprising, given that robust standard error estimators are known to often perform poorly in finite samples. Finally, the use of high-frequency data provides substantial gains in power: using just 13 intra-daily observations (corresponding to 30-minute returns for a stock traded on the NYSE) yields a marked improvement in power over a proxy based on daily returns. Using only $T = 100$ observations based on realised volatility with $m = 13$ produced similar power to using $T = 1000$ observations and squared daily returns ($m = 1$).

The size and power of the tests of covariance forecasts were examined in a similar study. Using the natural analogue of the GARCH(1,1), the DGP was specified as a bivariate scalar diagonal *vech*,

$$
\begin{aligned}
DGP \quad : \quad & \mathbf{r}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\varepsilon}_t, \, t = 1, 2, \ldots, T \\
\text{where } E_{t-1}\left[\boldsymbol{\varepsilon}_t\right] \quad = \quad & \mathbf{0} \text{ and } E_{t-1}\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right] = \mathbf{I}_k \\
\boldsymbol{\Sigma}_t \quad = \quad & 0.05\bar{\boldsymbol{\Sigma}} + 0.85\boldsymbol{\Sigma}_{t-1} + 0.10\mathbf{r}_{t-1}\mathbf{r}_{t-1}' \\
T \quad = \quad & \{100, 250, 500, 1000\}
\end{aligned}
\tag{28}
$$

where $\bar{\boldsymbol{\Sigma}}$ is a bivariate matrix with unit diagonals and off-diagonal values of 0.3 (see [12]). The standardised innovations were specified to allow DGP-consistent realised covariances to be computed,

$$
\begin{aligned}
\boldsymbol{\varepsilon}_t \quad &= \quad \sum_{m=1}^{78} \xi_{mt} \\
\xi_{mt} \quad &\overset{\text{IID}}{\sim} \quad N\left(0, 1/78\mathbf{I}_2\right).
\end{aligned}
\tag{29}
$$

Size Comparison of MZ-type Tests for Covariance Forecasts					
	T = 100			T = 250	
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
MZ1 (White)	0.18	0.09	0.08	0.13	0.07	0.07
MZ1-GLS	0.10	0.06	0.06	0.07	0.05	0.05
MZ1-GLS (White)	0.12	0.07	0.07	0.08	0.06	0.06
MZ1-Approx GLS (White)	0.13	0.07	0.07	0.09	0.06	0.06
MZ2	0.05	0.05	0.05	0.05	0.05	0.05
MZ2 (White)	0.06	0.08	0.08	0.07	0.06	0.06
	T = 500			T = 1000	
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
MZ1 (White)	0.11	0.06	0.06	0.09	0.06	0.06
MZ1-GLS	0.06	0.05	0.06	0.06	0.05	0.05
MZ1-GLS (White)	0.07	0.05	0.06	0.06	0.05	0.05
MZ1-Approx GLS (White)	0.07	0.05	0.06	0.06	0.05	0.05
MZ2	0.05	0.05	0.05	0.05	0.05	0.05
MZ2 (White)	0.05	0.05	0.05	0.05	0.05	0.05

Table 2: Rejection frequency of test statistics using a nominal size of 5% based on asymptotic critical values. All test statistics were computed using Wald tests using either White's heteroskedasticity-consistent covariance estimator or the standard OLS variance covariance formula as indicated. Three proxies were used for the latent covariance matrix: daily outer-products of returns ($RV^{(1)}$), a 13-sample realised covariance estimator ($RV^{(13)}$) and a 78-sample realised covariance estimator ($RV^{(78)}$)

The forecasting model was also a scalar diagonal *vech* parameterised in the spirit of (23),

$$\mathbf{H}_t = (1-k)\bar{\Sigma} + \frac{0.85}{0.95}k\mathbf{H}_{t-1} + \frac{0.10}{0.95}k\mathbf{r}_{t-1}\mathbf{r}'_{t-1}$$
$$k = \{0.80, 0.81, \ldots, 0.99, 1\}.$$

When $k = 0.95$ the model corresponds to the DGP. Daily cross-products and 13- and 78-sample realised covariance, $RC^{(13)}$ and $RC^{(78)}$, respectively, were used to proxy for the latent covariance and 10,000 replications were conducted.

In addition to the MZ1, MZ1-GLS and MZ2 specification studied in the volatility evaluation Monte Carlo, the approximate GLS specification in (21) was included in the study and is indicated by MZ1-Approx. GLS. Table 2 contains summary information about the size and Fig. 2 contains size-adjusted power curves for the specifications examined. The results are in line with those of the volatility tests: the tests have better size when GLS is used, particularly when the use of robust standard errors can be avoided, and the size is improved by using a more precise proxy. The power curves also show that MZ2 is less powerful than MZ1.

Combining the results of these two simulations, two main conclusions emerge. Firstly, the gains from using intra-daily data to construct a realised volatility proxy are large, even when only using a few intra-daily samples. Tests based solely on daily data are often oversized and have low power, even for $T = 1000$. The fact that substantial gains from intra-daily data may be obtained even when using just 30-minute returns is a positive result, given that prices sampled at this frequency are generally believed to be free from microstructure noise and non-synchronous trading problems; something that is not true for prices sampled a 5-minute intervals (see, e.g. [37, 33, 53]).
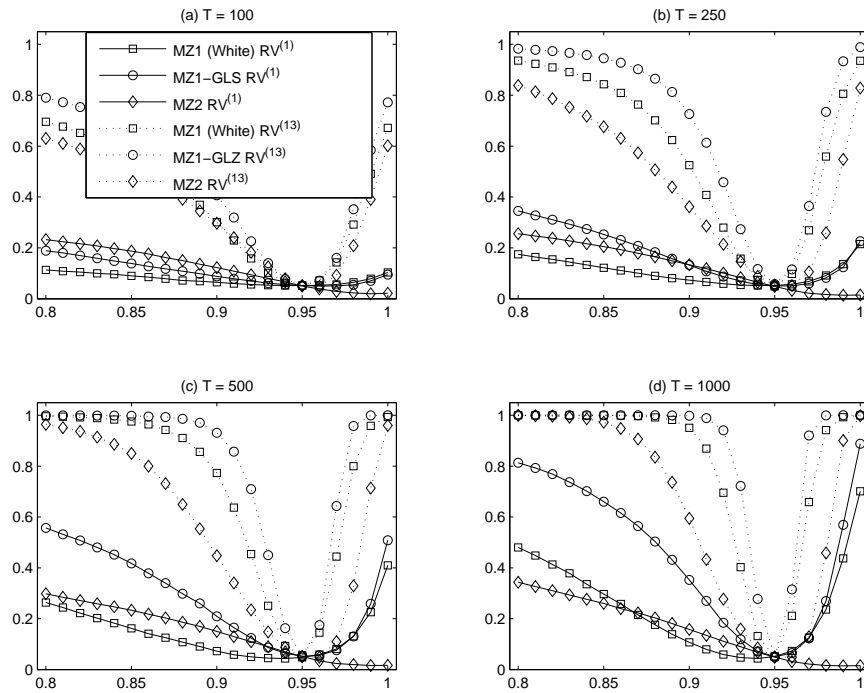
11

Figure 2: The four panels of this figure contain the size-adjusted power for MZ1 and MZ2 tests applied to covariance forecasts. Two proxies were used for the latent covariance matrix: daily outer-products of returns ($RV^{(1)}$, *solid line*) and a 13-sample realised covariance estimator ($RV^{(13)}$, *dotted line*). MZ1 tests regress the volatility proxy on the forecast using either OLS, indicated by □, or GLS, indicated by ○, and MZ2 tests regress the standardised volatility proxy on a lag of the standardised volatility proxy, indicated by ◊

The second main finding is that the use of a GLS estimator produces substantial improvements in finite-sample size and distinct increases in power. For example, when using daily squared returns as the volatility proxy, the use of the a GLS regression is approximately as powerful as a standard MZ regression with twice as many observations.

3   Direct Comparison of Volatility Forecasts

This section reviews methods for comparing two or more competing volatility forecasts using a volatility proxy.

Direct comparisons of competing volatility forecasts can be done in a number of ways. One popular approach for univariate volatility forecasts is to compare forecasts using the R^2 from MZ regressions of a proxy on the forecast [3, 6]. If the forecasts are unbiased, this is equivalent to ranking the forecasts on the basis of their mean square error (MSE). The significance of any difference in MSE, or R^2, can be tested via a Diebold-Mariano and West (henceforth DMW) test (see [23, 57, 58]).[11]

3.1   Pair-wise Comparison of Volatility Forecasts

The DMW test can be used to compare two forecasts using general loss functions, including those other than MSE. Consider the general case, with the loss function defined over the true conditional covariance, Σ_t, or a proxy, $\hat{\Sigma}_t$, and the covariance forecast, $\mathbf{H}_t$, $L : \mathbb{M}_+^K \times \mathcal{H}^K \to \mathbb{R}_+$. The DMW tests the null of equal predictive accuracy against composite alternatives that indicate which forecast performs better:

$$
\begin{aligned}
H_0 &: E\left[L(\Sigma_t, \mathbf{H}_t^A)\right] = E\left[L(\Sigma_t, \mathbf{H}_t^B)\right] \\
\text{vs.}\quad H_1 &: E\left[L(\Sigma_t, \mathbf{H}_t^A)\right] > E\left[L(\Sigma_t, \mathbf{H}_t^B)\right] \\
H_2 &: E\left[L(\Sigma_t, \mathbf{H}_t^A)\right] < E\left[L(\Sigma_t, \mathbf{H}_t^B)\right]
\end{aligned}
\tag{30}
$$

Since Σ_t is unobservable, this test is implemented using a statistic computed on the difference in the losses measured via a volatility proxy:

$$
d_t = L(\hat{\Sigma}_t, \mathbf{H}_t^A) - L(\hat{\Sigma}_t, \mathbf{H}_t^B)
\tag{31}
$$

The test is computed using a standard t-test:

$$
\begin{aligned}
DMW_T &= \frac{\sqrt{T}\bar{d}_T}{\sqrt{\widehat{avar}\left[\sqrt{T}\bar{d}_T\right]}} \\
\text{where } \bar{d}_T &\equiv \frac{1}{T}\sum_{t=1}^{T} d_t
\end{aligned}
\tag{32}
$$

and $\widehat{avar}\left[\sqrt{T}\bar{d}_T\right]$ is some consistent estimator of the asymptotic variance of the re-scaled average, $\sqrt{T}\bar{d}_T$, such as the Newey-West estimator [49]. Under the null hypothesis the test statistic is asymptotically Normally distributed.

---

[11]West [57] explicitly considers forecasts that are based on estimated parameters, whereas the null of equal predictive accuracy is based on population parameters. Diebold and Mariano [23], on the other hand, take the forecasts as given and do not allow for estimation error. In this chapter we also take the forecasts as given, and so these two approaches coincide.

Giacomini and White [29] recently proposed comparing forecasts based on their expected loss *conditional* on variables thought to be important for relative forecast performance. The null hypothesis of interest in such a comparison replaces the unconditional expectation in (30) with a conditional expectation. For example, in a volatility forecasting application, one could test whether one volatility forecast out-performs another more in times of high volatility than in times of low volatility, or during bull markets compared to bear markets. However, we have not seen such an application in the literature to date.

3.2 Comparison of Many Volatility Forecasts

When interest lies in comparing many competing forecasts there are two main approaches: the "Reality Check" [62], and a modified version with better power properties [34], or the "Model Confidence Set" [38]. The Reality Check tests the null hypothesis that no forecast outperforms, according to some loss function, a given benchmark forecast. The null hypothesis in this test is

$$H_0 \quad : \quad E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^A)\right] \leq \min_{i \in \{B, C, \dots\}} E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^i)\right]$$

$$\text{vs.} \quad H_1 \quad : \quad E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^A)\right] > \min_{i \in \{B, C, \dots\}} E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^i)\right]$$

Hansen and Lunde use this type of test to determine whether a GARCH(1,1) model was out-performed by any one of over three hundred competing volatility forecasts, for IBM equity return volatility and for the volatility of the log-difference of the DM-USD exchange rate [35].

The Model Confidence Set (MCS) is useful when there is no benchmark forecast. The outcome of this approach is a subset of forecasts that are not distinguishable from the best forecast across the complete set of forecasts. Defining the set of all competing forecasts as $\mathcal{M} = \{A, B, C, \dots\}$, the MCS tests the null that no forecast is distinguishable against an alternative that at least one of the forecasts has a higher expected loss,

$$H_0 \quad : \quad E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^i)\right] = E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^j)\right] \text{ for all } i, j \in \mathcal{M}$$

$$\text{vs.} \quad H_1 \quad : \quad E\left[L(\boldsymbol{\Sigma}_i, \mathbf{H}_t^i)\right] > E\left[L(\boldsymbol{\Sigma}_t, \mathbf{H}_t^j)\right] \text{ for some } i \in \mathcal{M}, \text{for all } j \in \mathcal{M} \setminus i.$$

The MCS operates by iteratively deleting poorly performing forecasts to construct a set, $\mathcal{M}^*$, that contains the forecast producing the lowest expected loss with probability weakly greater than the level of the test (e.g. 0.05), with the property that the probability that this set contains a sub-optimal forecast asymptotes to zero with the sample size. The MCS resembles in many respects a confidence interval for a parameter.

3.3 'Robust' Loss Functions for Forecast Comparison

Common to all approaches for comparing volatility forecasts is the focus on expected loss using the true, latent, covariance. In practice, however, the actual quantity computed is the difference in expected losses evaluated at some volatility proxy. Patton [52] defines a loss function as "robust" if it yields the same ranking of competing forecasts using an unbiased volatility proxy, $E_{t-1}[\hat{\boldsymbol{\Sigma}}_t] = \boldsymbol{\Sigma}_t$ as would be obtained using the (unobservable) conditional variance. Patton's focus was on the evaluation of univariate volatility forecasts, but the extension of his definition of loss function "robustness" is clear:

Definition 0.1 *A loss function, L, is "robust" if the ranking of any two (possibly imperfect) volatility forecasts,* $\mathbf{H}_t^A$ *and* $\mathbf{H}_t^B$, *by expected loss is the same whether the ranking is done using the true conditional variance,* $\boldsymbol{\Sigma}_t$,

or some conditionally unbiased volatility proxy, $\hat{\mathbf{\Sigma}}_t$:

$$E\left[L\left(\mathbf{\Sigma}_t, \mathbf{H}_t^A\right)\right] \gtreqless E\left[L\left(\mathbf{\Sigma}_t, \mathbf{H}_t^B\right)\right] \tag{33}$$
$$\Leftrightarrow E\left[L\left(\hat{\mathbf{\Sigma}}_t, \mathbf{H}_t^A\right)\right] \gtreqless E\left[L\left(\hat{\mathbf{\Sigma}}_t, \mathbf{H}_t^B\right)\right]$$

For univariate volatility forecast comparison, Meddahi showed that the ranking of forecasts on the basis of the R^2 from a standard Mincer-Zarnowitz regression is robust to noise in $\hat{\sigma}_t^2$ [47]. Hansen and Lunde showed that the R^2 from a regression of $\log\left(\hat{\sigma}_t^2\right)$ on a constant and $\log(h_t)$ is not robust to noise [36]. Moreover, they showed a sufficient condition for a loss function to be robust is that $\partial^2 L\left(\hat{\sigma}_t^2, h_t\right)/\partial\left(\hat{\sigma}_t^2\right)^2$ does not depend on h_t. Patton generalised this result by providing necessary and sufficient conditions for a univariate loss function to be robust [52].

3.4 Problems Arising rom 'Non-robust' Loss Functions

In this section we investigate the problems caused by the use of non-robust loss functions in univariate volatility forecast comparison, and the reduction in the magnitude of these problems achieved through the use of higher frequency data (such as realised volatility). Patton showed that if a loss function is robust, then the optimal forecast under L, defined as

$$h_t^* \equiv \arg\min_{h\in\mathcal{H}} E_{t-1}\left[L\left(\hat{\sigma}_t^2, h\right)\right]$$

must be the conditional variance (see [52]). One measure of the degree of distortion caused by the use of a loss function in combination with a noisy volatility proxy is the degree of bias in the optimal forecast under that loss function. Patton analytically derived the bias caused by nine widely-used loss functions, see equations (34) to (42) below, when combined with the range or realised variance as the volatility proxy [52]. Under a simple zero-drift Brownian motion assumption for the return process he found that the (multiplicative) bias ranged from 0.28 to 3 using daily squared returns as the volatility proxy, but shrank to 0.98 to 1.03 if a realised variance estimator based on 5-minute returns was available.

To investigate whether these dramatic reductions in bias when using volatility proxies based on high frequency data hold under more realistic assumptions on the data generating process (DGP), we conduct a small simulation study. We consider three data generating processes, using the same models and parameter values as the simulation study of Gonçalves and Meddahi [31]. The first model is a GARCH diffusion, as in [3]:

$$\begin{aligned} \text{d}\log P_t &= 0.0314\text{d}t + v_t\left(-0.576\text{d}W_{1t} + \sqrt{1-0.576^2}\text{d}W_{2t}\right) \\ \text{d}v_t^2 &= 0.035\left(0.636 - v_t^2\right)\text{d}t + 0.144v_t^2\text{d}W_{1t} \end{aligned}$$

The second model is a log-normal diffusion, as in [1]:

$$\begin{aligned} \text{d}\log P_t &= 0.0314\text{d}t + v_t\left(-0.576\text{d}W_{1t} + \sqrt{1-0.576^2}\text{d}W_{2t}\right) \\ \text{d}\log v_t^2 &= -0.0136\left(0.8382 + \log v_t^2\right)\text{d}t + 0.1148\text{d}W_{1t} \end{aligned}$$

15

The final model is the two-factor diffusion [16]:

$$
\begin{aligned}
\mathrm{d}\log P_t &= 0.030\mathrm{d}t + v_t\left(-0.30\mathrm{d}W_{1t} - 0.30\mathrm{d}W_{2t}\right.\\
&\qquad\left. + \sqrt{1 - 0.3^2 - 0.3^2}\mathrm{d}W_{3t}\right)\\
v_t^2 &= \text{s-exp}\left\{-1.2 + 0.04v_{1t}^2 + 1.5v_{2t}^2\right\}\\
\mathrm{d}v_{1t}^2 &= -0.00137v_{1t}^2\mathrm{d}t + \mathrm{d}W_{1t}\\
\mathrm{d}v_{2t}^2 &= -1.386v_{2t}^2\mathrm{d}t + \left(1 + 0.25v_{2t}^2\right)\mathrm{d}W_{2t}\\
\text{where}\quad \text{s-exp}\left\{x\right\} &= \begin{cases} \exp\left\{x\right\}, & x \le x_0 \\ \exp\left\{x_0\right\}\sqrt{1 - x_0 + x^2/x_0}, & x > x_0 \end{cases}
\end{aligned}
$$

In simulating from these processes we use a simple Euler discretization scheme, with the step size calibrated to one second (i.e., with 23,400 steps per simulated day). The loss functions we consider are:

$$MSE \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left(\hat{\sigma}_t^2 - h_t\right)^2 \tag{34}$$

$$QLIKE \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \log h_t + \frac{\hat{\sigma}_t^2}{h_t} \tag{35}$$

$$MSE\text{-}LOG \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left(\log\hat{\sigma}_t^2 - \log h_t\right)^2 \tag{36}$$

$$MSE\text{-}SD \quad : \quad L\left(\hat{\sigma}_t^2, h\right) = \left(\hat{\sigma}_t - \sqrt{h_t}\right)^2 \tag{37}$$

$$MSE\text{-}prop \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left(\frac{\hat{\sigma}_t^2}{h_t} - 1\right)^2 \tag{38}$$

$$MAE \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left|\hat{\sigma}^2 - h_t\right| \tag{39}$$

$$MAE\text{-}LOG \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left|\log\hat{\sigma}_t^2 - \log h_t\right| \tag{40}$$

$$MAE\text{-}SD \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left|\hat{\sigma}_t - \sqrt{h_t}\right| \tag{41}$$

$$MAE\text{-}prop \quad : \quad L\left(\hat{\sigma}_t^2, h_t\right) = \left|\frac{\hat{\sigma}^2}{h_t} - 1\right| \tag{42}$$

In the presence of stochastic volatility, as in the three DGPs above, the appropriate volatility concept changes from the conditional variance of daily returns to the expected *integrated variance*, see [4]:

$$
E_{t-1}\left[IV_t\right] \equiv E_{t-1}\left[\int_{t-1}^{t} v_\tau^2 d\tau\right]
$$

We consider three realised volatility proxies, based on $m = 1, 13$ and 78 intra-daily observations. We also consider the use of the adjusted squared range as a volatility proxy, which is defined as

$$
\begin{aligned}
RG_t^{*2} &= \frac{1}{4\log 2}RG_t^2\\
\text{where}\quad RG_t &\equiv \max_\tau \log P_\tau - \min_\tau \log P_\tau, \ t - 1 < \tau \le t
\end{aligned}
$$

The adjustment factor $\left(\frac{1}{4}\log 2\right)$ is required so as to make the squared range unbiased for the daily volatility for a Brownian motion with no drift [51]. Christensen and Podolskij show that this adjustment factor leads to an asymptotically unbiased proxy for general Brownian semi-martingales (as the number of intra-daily observations, m, goes to infinity, see [17]).

16

	Range	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	Range	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
				Biases in optimal volatility forecasts				
Loss function		Brownian Motion				GARCH-SV		
MSE	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
QLIKE	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
MSE-LOG	0.85	0.28	0.91	0.98	0.83	0.28	0.92	0.98
MSE-SD	0.92	0.56	0.96	0.99	0.91	0.63	0.96	0.99
MSE-prop	1.41	3.00	1.15	1.03	1.40	3.02	1.16	1.03
MAE	0.83	0.45	0.95	0.99	0.82	0.46	0.94	0.99
MAE-log	0.83	0.45	0.95	0.99	0.82	0.46	0.94	0.99
MAE-SD	0.83	0.45	0.95	0.99	0.82	0.46	0.94	0.99
MAE-prop	0.99	2.36	1.10	1.02	1.18	2.37	1.10	1.01
		Log-normal SV				Two-factor SV		
MSE	0.99	1.00	1.00	1.00	1.00	1.01	1.00	1.00
QLIKE	0.99	1.00	1.00	1.00	1.00	1.01	1.00	1.00
MSE-LOG	0.83	0.28	0.92	0.98	0.35	0.12	0.37	0.41
MSE-SD	0.91	0.63	0.96	0.99	0.57	0.40	0.58	0.62
MSE-prop	1.40	3.03	1.16	1.03	9.79	20.60	9.03	6.70
MAE	0.82	0.46	0.94	0.99	0.31	0.17	0.32	0.35
MAE-LOG	0.82	0.46	0.94	0.99	0.31	0.17	0.32	0.35
MAE-SD	0.82	0.46	0.94	0.99	0.31	0.17	0.32	0.35
MAE-prop	1.18	2.37	1.10	1.02	3.47	6.60	3.33	2.98

Table 3: Bias, as a percentage of the underlying integrated variance, when using robust and non-robust loss functions. A coefficient of 1 indicates the loss function does not generate a biased optimal forecast. The results in the Brownian motion case are analytical while the results in the three other cases are the result of a simulation

We present the results of our simulation in Table 3, and for ease of comparison we also present the analytical results from Patton for the Brownian motion case [52] . This table contains the (multiplicative) biases in the optimal forecasts under the loss functions listed above, for the various DGPs. An unbiased forecast will have coefficient of one. The MSE and QLIKE loss functions, as expected, did not generate bias for any volatility proxy. These loss functions are easily shown to be "robust", and so lead to zero bias as long as the volatility proxy is unbiased.

The first three panels of Table 3 reveal that allowing for stochastic volatility through a GARCH diffusion or a log-Normal diffusion does not substantially change the degree of bias in optimal forecasts under various loss function/volatility proxy combinations relative to the simple Brownian motion case. In fact, almost all of the differences occur only in the second decimal place. This suggests that the biases computed under the simplistic assumption of constant intra-daily volatility are a good approximation to those obtained under GARCH or log-Normal SV DGPs.

The situation is quite different when the two-factor SV model [16] is considered. This model was developed to replicate the jump-like features observed in some data without actually introducing a jump component into the model and can generate extreme observations and excess kurtosis. Patton found that excess kurtosis generally exacerbated any biases that were present under normally distributed returns, and this is reinforced by our simulation results: the two-factor diffusion generates biases ranging from 0.35 to 6.70 even when using a realised volatility estimator based on 5-minute returns [52]. That is, the biases from 5-minute realised volatility under the two-factor diffusion are greater than the biases from using *daily* squared returns when returns are conditionally normally distributed.

This simulation study shows that the use of volatility proxies with less noise (such as those based on higher-frequency data) ameliorates but does not eliminate the biases caused by the use of non-robust loss functions. The remaining biases can still be large depending on the form of the data generating process. This suggests that using robust loss functions in forecast comparison tests is important, even when higher quality volatility proxies are available.

3.5 Choosing a "Robust" Loss Function

Patton derives a parametric class of univariate loss functions that are both homogeneous and robust to the use of a noisy volatility proxy [52]. Homogeneity is a desirable property of a loss function that ensures the choice of normalisation (for example using raw returns versus 100 times returns) does not affect the optimal forecast beyond a known scale factor. The class can be described as[12]

$$L(\hat{\sigma}_t^2, h_t; b) = \begin{cases} \frac{1}{(b+1)(b+2)}(\hat{\sigma}_t^{2(b+2)} - h_t^{b+2}) \\ \qquad - \frac{1}{b+1} h_t^{b+1}(\hat{\sigma}_t^2 - h_t) & , b \neq -1, -2 \\ \\ h_t - \hat{\sigma}_t^2 + \hat{\sigma}_t^2 \log \frac{\hat{\sigma}_t^2}{h_t} & , b = -1 \\ \\ \frac{\hat{\sigma}_t^2}{h_t} - \log \frac{\hat{\sigma}_t^2}{h_t} - 1 & , b = -2 \end{cases} \tag{43}$$

The ability of these loss functions to distinguish between competing volatility forecasts in a DMW test may vary substantially with the choice of "shape" parameter, b. To provide some insight into this problem, we

---

[12]All loss functions have been normalised to 0 when $h_t = \hat{\sigma}_t^2$. In applications where $\Pr(\hat{\sigma}_t^2 = 0) > 0$ the normalised loss is not always well-defined. The normalisation terms can be removed without affecting any results with respect to robustness. Homogeneity is also preserved, effectively, as the removal of these normalising terms means that re-scaling the data adds a constant to the loss, which does not affect the optimum and drops out in forecast comparisons.

conduct a small simulation study of the size and power of a DMW test of the null that two competing forecasts have equal predictive power. The models we consider are based on those used in the Mincer-Zarnowitz Monte Carlo study in Sect. 2.5. The returns were generated according to (23). The first forecast, h_t^A, is set equal to the conditional variance of the process multiplied by an iid error term, z_t^A, distributed as a standardised χ^2 random variable: $v z_t^A \overset{IID}{\sim} \chi_v^2$. This error term can be thought of as representing (in a simple and computationally convenient way) estimation error or model mis-specification in the volatility forecast, for example. The second volatility forecast, h_t^B, is generated by (27), and is also multiplied by an iid error term, z_t^B, independent of the first, distributed as a standardised χ^2 random variable: $v z_t^B \overset{IID}{\sim} \chi_v^2$. We set the degree of freedom parameter, v, to 500, which implies that the iid error terms have unit mean and standard deviations of 0.06. Although neither of these forecasts is perfect, h_t^A is weakly preferred to h_t^B. The point where h_t^B is also "correctly specified" ($k = 0.95$) corresponds to the case when the two forecasts are equally accurate.

Using the two forecasts, the loss was computed using the "robust" loss function with shape parameter $b \in \{-5, -3, -2, -1, 0, 2\}$, and a DMW test statistic was computed using the difference in the losses

$$d_t(b) \quad = \quad L(\hat{\sigma}_t^2, h_t^A; b) - L(\hat{\sigma}_t^2, h_t^B; b) \tag{44}$$

$$DMW_T(b) \quad = \quad \frac{\sqrt{T} \bar{d}_T(b)}{\sqrt{\widehat{avar}\left[\sqrt{T} \bar{d}_T(b)\right]}} \tag{45}$$

The asymptotic variance of the average was computed using a Newey-West variance estimator with the number of lags set to $\lceil T^{1/3} \rceil$.

Three proxies were used to compare the performance of the two forecasts: the daily squared return, $RV^{(1)}$, a 30-minute realised variance $RV^{(13)}$, and a 5-minute realised variance $RV^{(78)}$. The finite-sample size of the test is reported in Table 4. Overall the size of the test is good, however for larger values of b the test is undersized. This may be due to a lack of moments in the d_t series, which results in a non-standard distribution of the DMW test statistic. This non-standard behaviour was foreshadowed in Patton, who showed that $8 + \delta$ moments of returns are needed for a DMW test using squared returns as the volatility proxy [52]. Larger values of b require even more moments of returns. In many realistic cases, including the model use in the simulation study, returns will not be sufficiently well behaved for tests based loss functions $b \geq 0$ to be reliable.

The power of the test was studied using a model with mis-specified dynamics where the alternative forecasts were generated from either over- or under-persistent models. Figure 3 contains four views into the power of the test using $T = 100$ and $T = 250$ using daily squared returns and 13-sample realised volatility. Panels (c) and (d) confirm that improved proxies for volatility have a distinct effect on the ability to discern superior forecast performance.

Three of the four panels show that the QLIKE loss function ($b = -2$) has the highest power. Using either daily returns or 30-minute realised volatility, power drops off markedly when using large b loss functions. Even when using a precise proxy and a long sample there is a distinct decrease in power for the loss functions with b furthest from -2.

Figure 4 contains the average size-adjusted power (averaging the power curves in Fig. 3 across all $k \in \{0.80, 0.81, \ldots, 1.00\}$) as a function of b using daily returns squared and 30-minute realised variance. Both power plots exhibit peaks near -2 in all cases except the smallest sample size test statistic computed from

	Finite-Sample Size of DMW Tests					
	T = 100			T = 250		
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
b=-5	0.05	0.06	0.05	0.05	0.05	0.05
b=-3	0.06	0.07	0.07	0.06	0.06	0.06
b=-2 (QLIKE)	0.06	0.07	0.07	0.06	0.06	0.06
b=-1	0.06	0.07	0.07	0.05	0.06	0.06
b=0 (MSE)	0.05	0.06	0.05	0.04	0.04	0.05
b=2	0.03	0.03	0.03	0.02	0.02	0.02
	T = 500			T = 1000		
	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$	$RV^{(1)}$	$RV^{(13)}$	$RV^{(78)}$
b=-5	0.05	0.05	0.05	0.05	0.05	0.05
b=-3	0.05	0.05	0.06	0.05	0.05	0.05
b=-2 (QLIKE)	0.05	0.05	0.06	0.05	0.05	0.05
b=-1	0.05	0.05	0.05	0.05	0.05	0.05
b=0 (MSE)	0.04	0.04	0.04	0.04	0.04	0.04
b=2	0.02	0.02	0.02	0.02	0.02	0.02

Table 4: Finite-sample size of the DMW tests using different volatility proxies and sample sizes. Data was simulated according to a GARCH(1,1) and the series of forecasts, $\{h_t\}$, was produced. The size was computed by comparing the performance of a DMW test comparing the loss of $h_t^A = z_t^A h_t$ and $h_t^B = z_t^B h_t$ where z_t^A and z_t^B were independent standardised χ_ν^2 random variables with $\nu = 500$
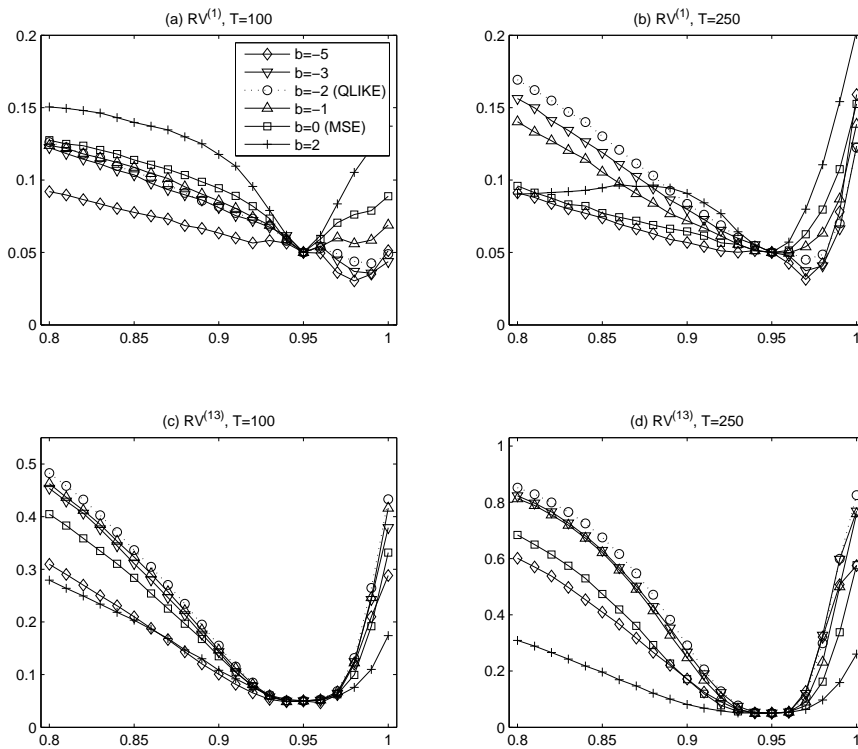
Figure 3: Plots of size-adjusted power for DMW tests using two different proxies for latent volatility, daily returns squared ($RV^{(1)}$) and 30-minute realised volatility ($RV^{(13)}$). All loss functions are members of the "robust" loss function family, equation (43). Note: The scale of the y-axis changes in each panel
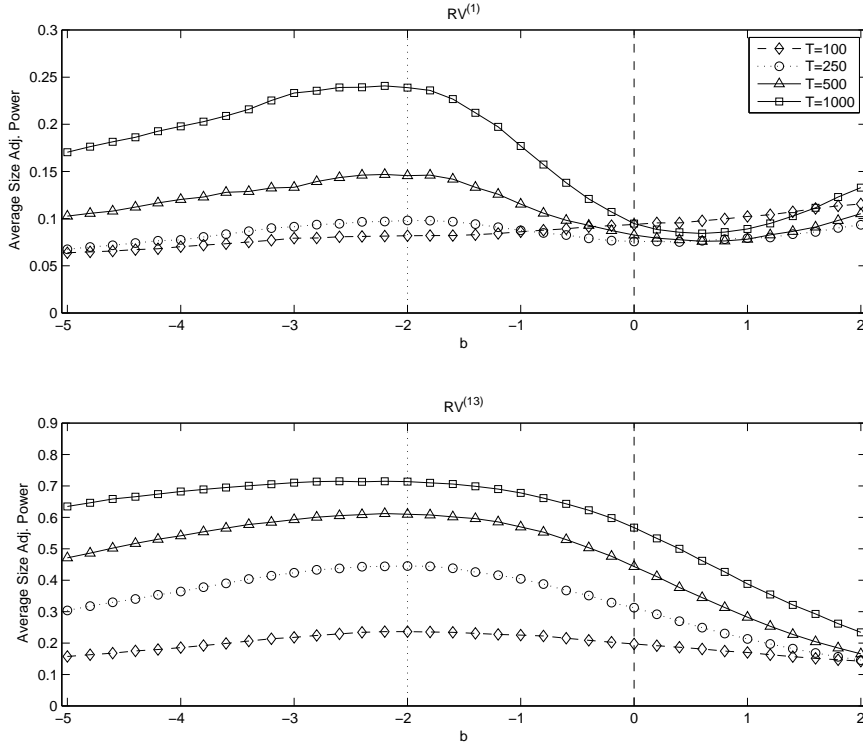
Figure 4: Average size-adjusted power curves for DMW tests as a function of b. DMW tests were computed using daily squared returns ($RV^{(1)}$, top panel) or 13-sample realised volatility ($RV^{(13)}$, bottom panel) using the "robust" loss function family in equation (43). Note: The scale of the y-axis changes in each panel

daily returns. These results all point to QLIKE as the preferred choice among the loss functions that are both homogeneous and robust to noise in the proxy[13].

3.6 Robust Loss Functions for Multivariate Volatility Comparison

The homogeneous, robust loss functions of Patton [52] all have a first-order condition of the form:

$$\frac{\partial L\left(\hat{\sigma}_t^2, h_t; b\right)}{\partial h_t} = -h_t^b(\hat{\sigma}_t^2 - h_t).$$ (46)

From this first-order condition it is simple to see that when $h_t = \sigma_t^2 \, a.s.$ and $h_t > 0$, the expected score is zero and the second derivative is positive. As a result, the true conditional variance is the solution to the expected loss minimisation problem. Extending this analysis to the evaluation of covariance forecasts is straightforward: The direct analogue of (46) for conditional covariance forecasts is

$$\frac{\partial L\left(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b\right)}{\partial \mathbf{H}_t} = -\mathbf{H}_t^b(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t).$$ (47)

While this is not the FOC of any standard expression, it does provide guidance to one class of FOCs, namely

---

[13]It is worth noting that the results presented for $b = -2$ are likely a close-to-ideal case. Returns in practice are decidedly non-normal exhibiting fat-tails, skewness and/or jumps. The increased propensity for actual asset returns to produce large observations should produce even worse performance of loss functions with $b \geq 0$.

$$\frac{\partial L\left(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b\right)}{\partial \mathbf{H}_t} = -\mathbf{C}_1\left(\mathbf{H}_t\right)(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t)\mathbf{C}_2\left(\mathbf{H}_t\right) \ . \tag{48}$$

where $\mathbf{C}_1\left(\mathbf{H}_t\right)$ and $\mathbf{C}_2\left(\mathbf{H}_t\right)$ are positive definite matrix valued functions, $\mathbf{C} : \mathbb{M}_{++}^K \to \mathbb{M}_{++}^K$ for $i = 1, 2$, that do not depend on $\hat{\boldsymbol{\Sigma}}_t$. Using the "vec" function to express this FOC as a column vector, it can be equivalently written

$$\frac{\partial L(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b)}{\partial \text{vec}(\mathbf{H}_t)} = -(\mathbf{C}_2\left(\mathbf{H}_t\right) \otimes \mathbf{C}_1\left(\mathbf{H}_t\right))\text{vec}(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t) \ . \tag{49}$$

It is simple to verify that this FOC will be zero as long as $E_{t-1}[\hat{\boldsymbol{\Sigma}}_t] = \boldsymbol{\Sigma}_t = \mathbf{H}_t \ a.s.$ The second order derivative with respect to $\mathbf{H}_t$ is $\mathbf{C}_2\left(\mathbf{H}_t\right) \otimes \mathbf{C}_1\left(\mathbf{H}_t\right)$ since $\partial \text{vec}\mathbf{H}_t / \partial \mathbf{H}_t = \mathbf{I}_{K^2}$ and is positive semi-definite by construction. The natural analogue to the class of robust, homogeneous loss functions introduced in Patton [52] is thus

$$L(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b) = \begin{cases} \frac{2}{(b+2)}\text{tr}(\boldsymbol{\Sigma}_t^{(b+2)} - \mathbf{H}_t^{b+2}) - \text{tr}(\mathbf{H}_t^{b+1}(\boldsymbol{\Sigma}_t - \mathbf{H}_t)) & b \neq -1, -2 \\ \text{tr}(\mathbf{H}_t^{-1}\boldsymbol{\Sigma}_t) - \log|\mathbf{H}_t^{-1}\boldsymbol{\Sigma}_t| - K & b = -2 \end{cases} \tag{50}$$

As in the univariate case, this class nests both the multivariate QLIKE and multivariate MSE classes. To verify that this class satisfies the condition of (47), consider the first order conditions when b is an integer[14]

$$\frac{\partial L(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b)}{\partial \mathbf{H}_t} = \begin{cases} -\sum_{j=0}^{b} \mathbf{H}^j \boldsymbol{\Sigma}_t \mathbf{H}^{b-j} - \mathbf{H}_t^{b+1} & b \in \mathbb{Z}_+ \\ -\mathbf{H}_t^{-1}\boldsymbol{\Sigma}_t \mathbf{H}_t^{-1} + \mathbf{H}_t^{-1} & b = -2 \end{cases} \tag{51}$$

and, transforming to column vectors,

$$\frac{\partial L(\hat{\boldsymbol{\Sigma}}_t, \mathbf{H}_t; b)}{\partial \text{vec}(\mathbf{H}_t)} = - \begin{cases} -\sum_{j=0}^{b} \left(\mathbf{H}^{b-j} \otimes \mathbf{H}^j\right) \text{vec}(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t) & b \in \mathbb{Z}_+ \\ -\left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \text{vec}(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t) & b = -2 \end{cases} \tag{52}$$

It should be noted that, unlike the univariate case, this class of loss functions does not encompass all that are robust and homogeneous. The expanded possibilities arise naturally because there are many functions $\mathbf{C}$ that can be used to weight the forecast errors, $\text{vec}(\hat{\boldsymbol{\Sigma}}_t - \mathbf{H}_t)$.

3.7 Direct Comparison via Encompassing Tests

Encompassing tests are an alternative to DMW tests for comparing the performance of two or more forecasts. Rather than compare the loss of using one forecast to the loss of using the other, encompassing tests examine whether some function, generally affine, of the forecasts produces a smaller loss than a single forecast. Restricting attention to only two forecasts, the null and alternative tested are

$$H_0^A \ : \ E[L(\hat{\sigma}_t^2, h_t^A)] = E[L(\hat{\sigma}_t^2, f(h_t^A, h_t^B; \boldsymbol{\theta}))] \tag{53}$$

$$\text{vs.} \ H_1^A \ : \ E[L(\hat{\sigma}_t^2, h_t^A)] > E[L(\hat{\sigma}_t^2, f(h_t^A, h_t^B; \boldsymbol{\theta}))] \tag{54}$$

and

$$H_0^B \ : \ E[L(\hat{\sigma}_t^2, h_t^B)] = E[L(\hat{\sigma}_t^2, f(h_t^A, h_t^B; \boldsymbol{\theta}))] \tag{55}$$

$$\text{vs.} \ H_1^B \ : \ E[L(\hat{\sigma}_t^2, h_t^B)] > E[L(\hat{\sigma}_t^2, f(h_t^A, h_t^B; \boldsymbol{\theta}))] \tag{56}$$

---

[14]For clarity of exposition, we only show the FOC when b is an integer greater than 0. The cases with non-integer b involve eigenvalue decompositions and are too lengthy to present here, although the loss function analysis goes through unmodified for any $b \neq -1$.

where $\boldsymbol{\theta}$ is a vector of unknown parameters. The "forecast combination function", $f(h_t^A, h_t^B; \boldsymbol{\theta})$ is typically specified as a linear combination of the forecasts, thus $f(h_t^A, h_t^B; \boldsymbol{\theta}) = \beta_1 h_t^A + \beta_2 h_t^B$. When using the MSE loss function, the encompassing test reduces to a standard augmented MZ regressions (see (9)) where the $\mathcal{F}_t$ measurable "instruments" are the competing forecasts,

$$\hat{\sigma}_t^2 = \beta_1 h_t^A + \beta_2 h_t^B + u_t \tag{57}$$

and testing

$$H_0^A \quad : \quad \beta_1 = 1 \cap \beta_2 = 0$$
$$H_1^A \quad : \quad \beta_1 \neq 1 \cup \beta_2 \neq 0$$

To test whether forecast B encompasses forecast A we test:

$$H_0^B \quad : \quad \beta_1 = 0 \cap \beta_2 = 1$$
$$H_1^B \quad : \quad \beta_1 \neq 0 \cup \beta_2 \neq 1$$

Like the augmented MZ, the performance of this specification will generally be adversely affected by heteroskedastic errors. A GLS version can be specified

$$\frac{\hat{\sigma}_t^2}{\tilde{h}_t} = \beta_1 \frac{h_t^A}{\tilde{h}_t} + \beta_2 \frac{h_t^B}{\tilde{h}_t} + \tilde{u}_t \tag{58}$$

where $\tilde{h}_t = h_t^A$ to test H_0^A, and $\tilde{h}_t = h_t^B$ to test H_0^B. If there is no natural null hypothesis, there are two alternative choices for the weights, $\tilde{h}_t$: The first employs an average of h_t^A and h_t^B, either geometric $\tilde{h}_t = \sqrt{h_t^A h_t^B}$ or arithmetic $\tilde{h}_t = \frac{1}{2} h_t^A + \frac{1}{2} h_t^B$. The second uses a two-step feasible GLS (FGLS) where (57) is initially estimated and then (58) is estimated using $\tilde{h}_t = \hat{\beta}_1 h_t^A + \hat{\beta}_2 h_t^B$, although care is needed to ensure the fitted volatilities are positive[15].

An alternative specification for $f(h_t^A, h_t^B; \boldsymbol{\theta})$, one that avoids any problems of negative fit volatilities, uses a geometric-average inspired form:

$$f(h_t^A, h_t^B; \boldsymbol{\theta}) = \exp(\beta_1 \ln h_t^A + \beta_2 \ln h_t^B) \tag{59}$$

The null and alternative hypotheses are identical, although non-linear least squares estimation is needed to implement this form of an encompassing test.

Extending encompassing tests to covariance forecasts is straightforward. The only remaining choice is the combination function. The natural candidate is a linear function:

$$f(\mathbf{H}_t^A, \mathbf{H}_t^B; \boldsymbol{\theta}) = \beta_1 \mathbf{H}_t^A + \beta_2 \mathbf{H}_t^B . \tag{60}$$

Using a linear combination function, encompassing can be tested using an augmented MZ version of the pooled panel regression (see 16). An alternative choice for the combination covariance would be to use a geometric average,

$$f(\mathbf{H}_t^A, \mathbf{H}_t^B; \boldsymbol{\theta}) = \mathrm{expm} \left(\beta_1 \mathrm{logm} \mathbf{H}_t^A + \beta_2 \mathrm{logm} \mathbf{H}_t^B \right) \tag{61}$$

where "expm" and "logm" are matrix exponentiation and logarithm respectively, see [45]. This alternative specification removes any restrictions on the estimated parameters while ensuring that the combination is strictly positive definite.

---

[15] While it does not affect the asymptotic distribution, there may be finite sample gains to using improved first-stage estimates from a standard GLS assuming pre-specified weights, such as $\tilde{h}_t = h_t^A$ or $\tilde{h}_t = h_t^B$.

4 Indirect Evaluation of Volatility Forecasts

While statistical evaluation of forecasts is useful for ranking volatility forecasts, most forecasts are designed to aid in economic applications, see Andersen, Bollerslev, Christoffersen and Diebold for a recent survey of volatility and correlation forecasting [2]. Economic evaluation of volatility forecasts is an important metric for assessing the performance of models.

Volatility and covariance forecasts are fundamental inputs into many decisions in financial economics: mean-variance portfolio optimisation, hedging, risk measurement, option pricing and utility maximisation all rely on forecast variances and covariance as inputs. While volatility forecasts have been evaluated using all of these applications, mean-variance portfolio choice and hedging are unique in that the correct conditional volatility or covariance, $\Sigma_t = H_t$ $a.s.$, will lead to improved performance without strong auxiliary assumptions. In general, economic evaluation of volatility and correlation forecasts relies in an important way on other assumptions, such as the utility function of the hypothetical investor (in portfolio choice or hedging applications), the density of the standardised returns (as in Value-at-Risk and Expected Shortfall forecasting, density forecasting, portfolio choice applications with non-quadratic utility), the derivative pricing model (in option, and other derivative securities, pricing applications). Thus these approaches are "non-robust", in the sense described in the Introduction, however with strong economic motivations they can yield valuable information on competing volatility and correlation forecasts.

Economic evaluation of covariance forecasts has traditionally focused on the out-of-sample performance of portfolios formed using the forecast as an input in a mean-variance framework. Noting the sensitivity of portfolio weights to small changes in the conditional mean, many have questioned the wisdom of conditional mean-variance optimisation. However, recent work by Engle and Colacito (EC, henceforth) has clarified the role that the uncertain mean plays in comparing the performance of two or more forecasts [24]. The EC framework establishes that the correct covariance will produce a smaller expected portfolio variance for *any* assumed non-zero vector of means. As a result, evidence of superiority based on ex-post measurements of portfolio variance can be attributed to the covariance forecast.

4.1 Portfolio Optimisation

Portfolio optimisation is a natural application for covariance forecasts. To avoid specifying or estimating asset mean returns, many authors have focussed on the problem of finding the global minimum variance portfolio. It is computed as the solution to

$$\min_{\mathbf{w}_t} \mathbf{w}_t' \Sigma_t \mathbf{w}_t \tag{62}$$
$$\text{subject to } \mathbf{w}_t' \iota = 1$$

It is simple to show that if the portfolio weights, $\mathbf{w}_t$, are constructed from the true covariance $H_t = \Sigma_t$ $a.s.$, then the variance of a portfolio computed using the GMVP from any other forecast, $\tilde{\mathbf{w}}_t$, must be larger:

$$V[\tilde{\boldsymbol{w}}_t' \Sigma_t \tilde{\boldsymbol{w}}_t] = (\mathbf{w_t} + \mathbf{c_t})' \Sigma_t (\mathbf{w}_t + \mathbf{c_t})$$

$$= (\frac{\boldsymbol{\iota}' \Sigma_t^{-1}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} + \mathbf{c}_t') \Sigma_t (\frac{\Sigma_t^{-1} \boldsymbol{\iota}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} + \mathbf{c}_t)$$

$$= \frac{\boldsymbol{\iota}' \Sigma_t^{-1}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} \Sigma_t \frac{\Sigma_t^{-1} \boldsymbol{\iota}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} + \frac{\boldsymbol{\iota}' \Sigma_t^{-1}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} \Sigma_t \mathbf{c}_t + \mathbf{c}_t' \Sigma_t \frac{\Sigma_t^{-1} \boldsymbol{\iota}}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} + \mathbf{c}_t' \Sigma_t \mathbf{c}_t$$

$$= \frac{1}{\boldsymbol{\iota}' \Sigma_t^{-1} \boldsymbol{\iota}} + \mathbf{c}_t' \Sigma_t \mathbf{c}_t > \frac{1}{\boldsymbol{\iota} \Sigma_t^{-1} \boldsymbol{\iota}}$$

since $\boldsymbol{\iota}' \mathbf{c}_t = 0$ follows from $\boldsymbol{\iota}' \mathbf{w}_t = \boldsymbol{\iota}' \tilde{\mathbf{w}}_t = 1$. Using this result, it is then possible to compare two competing covariance forecasts by comparing the volatility of the minimum variance portfolio constructed using each forecast. Let $\mathbf{w}^* (\mathbf{H}_t)$ denote the solution to (62) for covariance matrix $\mathbf{H}_t$, then let

$$d_t \equiv \mathbf{w}^* (\mathbf{H}_t^B)' \mathbf{r}_t \mathbf{r}_t' \mathbf{w}^* (\mathbf{H}_t^B) - \mathbf{w}^* (\mathbf{H}_t^A)' \mathbf{r}_t \mathbf{r}_t' \mathbf{w}^* (\mathbf{H}_t^A)$$

This then allows for a DMW forecast comparison test, as in (30): if the mean of d_t is significantly positive (negative) then forecast A (B) is the better forecast. If more than one competing forecast is being considered, the "reality check" or MCS can be used to test for superior performance (see Sect. 3.2).

The general mean-variance portfolio optimisation problem is

$$\min_{\mathbf{w}_t} \mathbf{w}_t' \Sigma_t \mathbf{w}_t$$
$$\text{subject to } \mathbf{w}_t' \boldsymbol{\mu}_t = \mu_0$$

where $\boldsymbol{\mu}_t = E_{t-1}[\mathbf{r}_t]$. If returns are expressed in excess of the risk free rate, optimal portfolio weights can be computed as

$$\mathbf{w}_t = \frac{\Sigma_t^{-1} \boldsymbol{\mu}_t}{\boldsymbol{\mu}_t \Sigma_t^{-1} \boldsymbol{\mu}_t} \mu_0$$

where the weight on the risk free asset is $1 - \mathbf{w}_t' \boldsymbol{\iota}$.[16] Recent work by Engle and Colacito [24] has shown that it is possible to rank covariance forecasts using this more general mean-variance optimisation problem. These authors show that if expected returns are constant, $E_{t-1}[\mathbf{r}_t] = \boldsymbol{\mu} \forall t$, then for *any* assumed mean vector $(\boldsymbol{\mu})$ the true covariance will produce a portfolio with a variance less than or equal to that of any forecast where $\mathbf{H}_t \neq \Sigma_t$. Engle & Colacito suggest using only positive values for $\boldsymbol{\mu}$, and, rather than testing using a single value, a quasi-Bayesian method can be used to integrate over a range of plausible values for $\boldsymbol{\mu}$.

Forecast evaluation tests, as opposed to comparison tests, are more difficult in this application due to the presence of a non-negligible mean. Using squared returns as a proxy for the unobservable conditional variance relies on expected returns being zero, an implausible assumption for investors holding risky assets. Engle and Colacito argue that if the measurement frequency is high enough, then expected returns are of a smaller order of magnitude than volatility, and so the mean can be ignored. In that case, the squared portfolio return, $\mathbf{w}_t' \mathbf{r}_t \mathbf{r}_t' \mathbf{w}_t$ or, better yet, the realised volatility of the portfolio, may be used as a volatility proxy and the estimated portfolio variance, $\mathbf{w}_t' \mathbf{H}_t \mathbf{w}_t$, can be evaluated using the univariate methods described in Sects. 2.1 and 2.4.

---

[16]If the returns are not computed as the difference between nominal returns and the risk-free rate, mean-variance analysis can still be used although with more complication. For details, see [14, pp. 184-185].

One important downside of minimum portfolio variance optimisation is that the mapping from $\mathbb{M}_{++}^K \to \mathbb{R}^K$ is many to one. There will generally be a continuum of forecasts that will produce the set of weights corresponding to the minimum variance portfolio. One direct method to address this deficiency is to test at least $K(K+1)/2$ portfolios using subsets of 2^K possible collections of assets.

4.2 Tracking Error Minimisation

Conditional covariance also plays a crucial role in estimating time-varying weights for tracking portfolios. Suppose a portfolio of $K-1$ stocks, with returns denoted $\tilde{\mathbf{r}}_t$ is to be used track the return of another asset, r_t, by minimising the mean squared tracking error:

$$\min_{\mathbf{w}} \mathrm{E}[(r_t - \mathbf{w}'\tilde{\mathbf{r}}_t)^2] \tag{63}$$

It is well known that the optimal linear projection satisfies

$$\mathbf{w} = E\left[\tilde{\mathbf{r}}_t \tilde{\mathbf{r}}_t'\right]^{-1} E[\tilde{\mathbf{r}}_t r_t] \tag{64}$$

which are just the usual regression coefficients, see [40]. Suppose the vector of returns is partitioned such that the first return, r_t, is the asset to be tracked so the covariance can be expressed

$$\mathbf{\Sigma}_t = \left[\begin{array}{cc} \sigma_{1,t}^2 & \mathbf{\Sigma}_{12,t}' \\ \mathbf{\Sigma}_{12,t} & \mathbf{\Sigma}_{22,t} \end{array} \right]$$

where $V_{t-1}[r_t] = \sigma_{1,t}^2$ and $V_{t-1}[\tilde{\mathbf{r}}_t] = \mathbf{\Sigma}_{22,t}$ The weights in the minimum tracking error portfolio can be computed by

$$\mathbf{w}_t = \mathbf{\Sigma}_{22,t}^{-1}\mathbf{\Sigma}_{12,t}. \tag{65}$$

Using the set of portfolio weights, the accuracy of the tracking portfolio weights can be tested with a MZ-GLS regression and the relative performance of two competing forecasts can be assessed using a Diebold-Mariano-West test. However, unlike the previously described minimum variance portfolio problems, it is not clear that utilising a forecast that satisfies H_0^* will lead to the smallest tracking error. However, the tracking error minimisation problem can be recast into the Engle-Colacito framework *if* the expected returns on all assets are assumed to be the same. In that case, the minimum variance portfolio problem becomes

$$\min_{\mathbf{w}_t} \mathbf{w}_t'\mathbf{\Sigma}_t\mathbf{w}_t$$

$$\text{subject to } w_1 = 1$$

which can be re-written as the search for a global minimum variance portfolio by substituting in the constraint and expressing the returns on all assets as excesses above the return on the first asset.

4.3 Other Methods of Indirect Evaluation

Volatility forecasts play a critical role in many other financial decisions, and accordingly their performance in these decisions is of much interest. Unfortunately, but perhaps not surprisingly, most financial decisions also depend on inputs beyond a volatility forecast. Influential assumptions about these other inputs will be required and these evaluations are "non-robust" using the definitions of this chapter. Nevertheless, we review the most common applications in this section.

One key application of volatility forecasts is in derivatives pricing, given the sensitivity of these securities to the volatility of the underlying asset price. These applications are generally univariate in nature; derivatives with multiple underlying assets (thus requiring a covariance forecast) are much less studied in the literature (see [9, 28] for surveys). In any applications to derivatives, the volatility forecast must be combined with a pricing model – the most commonly-used such model is the Black-Scholes model for European options. If the pricing model is mis-specified then the ranking of two volatility forecasts by pricing errors will not necessarily lead to the true conditional variance being ranked above imperfect forecasts. However if the volatility forecast *is* to be used in a particular pricing model, then the interest of the forecast user is not necessarily in finding the true conditional variance, rather it is in finding the forecast that produces the smallest pricing errors, and so this "distortion" is not a cause for concern. The evaluation and comparison of volatility forecasts via derivative pricing problems has been previously considered (see [50, 30, 21, 20, 32]).

Another important financial application involving volatility forecasts is portfolio decisions. We reviewed two special cases in the previous section where only the covariance was required for a portfolio decision. For most utility functions and general returns distributions, other inputs are needed to choose portfolio weights, and the specification of these inputs can affect the ranking of volatility forecasts by the out-of-sample utility of portfolio returns. Applications of volatility forecasts in portfolio decisions are have been widely explored (see [59, 27, 46, 32], amongst others).

In the past decade or so, measures of risk beyond standard deviation have gained increasing attention. Most prominent among these is Value-at-Risk (VaR), which can be defined as the α-quantile of the conditional distribution of the return on a given asset or portfolio[17]. In most applications, α is set to 0.05 or 0.01. See [19] for more on VaR. It is possible to produce conditional VaR forecasts by directly modelling the conditional quantile of interest, see [25]. However, the majority of conditional VaR forecasts are produced by first specifying a model for the conditional variance, and then specifying a model for the quantile of the standardised residual. This link between conditional VaR forecasts and conditional variance forecasts has lead some authors to suggest testing variance forecasts by testing the performance of VaR forecasts based on the variance forecast(s), although mis-specification of the distribution for the standardised residuals can lead to the rejection of a forecast satisfying H_0^*. VaR-based evaluation and comparison of volatility forecast has been examined in [44, 32, 26, 43], amongst others.

5 Conclusion

This chapter provided an overview of the methods available for evaluating and comparing forecasts of the conditional variance of an asset return or the conditional covariance of a set of asset returns. We paid particular attention to the problems that arise due to the fact that volatility is unobservable. We emphasised the importance of using tests (a) that are robust to the "noise" in the volatility proxy, if a proxy used, and (b) that require only minimal assumptions on the distribution of the returns. Many widely-employed volatility forecast evaluation tests, such as those using VaR or option pricing, fail one or both of these criteria.

In addition to presenting the theory for methods that satisfy these two criteria, we also presented the results of some small Monte Carlo studies to provide guidance for empirical work. We suggested a modification of the widely-used Mincer-Zarnowitz regression for testing volatility forecast optimality, which exploits the additional structure that holds under the null hypothesis. Our suggested "MZ-GLS" test has good size

---

[17]Another prominent measure of risk is "expected shortfall", which is defined as the conditional mean of the return on an asset given that the return is less than the Value-at-Risk, ie: $ES_t \equiv E\left[r_t | \mathcal{F}_{t-1} \cap r_t \leq VaR_t\right]$.

and much better power in finite samples than other MZ tests. Our simulations also clearly demonstrated the value of higher-precision volatility proxies, such as realised variance [5]. Even simple estimators based on 30-minute returns provide large gains in power and improvements in finite-sample size.

In Monte Carlo studies using realistic stochastic volatility processes, we studied the choice of loss function in Diebold-Mariano-West (DMW) tests [23, 57]. We showed that the use of loss functions that are "non-robust", in the sense of Patton, can yield perverse rankings of forecasts, even when very accurate volatility proxies are employed [52]. Amongst the class of robust and homogeneous loss functions in Patton and the multivariate generalisation of these loss functions provided in this chapter, our small simulations suggested that the "QLIKE" loss functions yield the greatest power in DMW tests.

References

[1] Andersen, T. G., Benzoni, L. and Lund, J. [2002], 'An empirical investigation of continuous-time equity return models', *Journal of Finance* **57**(3), 1239–1284.

[2] Andersen, T. G., Bollerselv, T., Christoffersen, P. F. and Diebold, F. X. [2006], Volatility and correlation forecasting, *in* G. Elliott, C. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting', North Holland Press, Amsterdam.

[3] Andersen, T. G. and Bollerslev, T. [1998], 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**(4), 885–905.

[4] Andersen, T. G., Bollerslev, T. and Diebold, F. X. [2006], Parametric and nonparametric volatility measurement, *in* L. P. Hansen and Y. Aït-Sahalia, eds, 'Handbook of Financial Econometrics', North-Holland, Amsterdam. *Forthcoming*.

[5] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. [2003], 'Modeling and Forecasting Realized Volatility', *Econometrica* **71**(1), 3–29.

[6] Andersen, T. G., Bollerslev, T. and Meddahi, N. [2005], 'Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities', *Econometrica* **73**(1), 279–296.

[7] Arellano, M. [2003], *Panel Data Econometrics*, Oxford University Press, Oxford.

[8] Barndorff-Nielsen, O. E. and Shephard, N. [2004], 'Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics', *Econometrica* **73**(4), 885–925.

[9] Bates, D. S. [2003], 'Empirical option pricing: a retrospection', *Journal of Econometrics* **116**(1-2), 387–404.

[10] Bierens, H. J. [1990], 'A consistent conditional moment test of functional form', *Econometrica* **58**(6), 1443–1458.

[11] Bierens, H. J. and Ploberger, W. [1997], 'Asymptotic theory of integrated conditional moment tests', *Econometrica* **65**(5), 1129–1152.

[12] Bollerslev, T., Engle, R. F. and Wooldridge, J. M. [1988], 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

[13] Bollerslev, T. and Wright, J. H. [2001], 'High-frequency data, frequency domain inference, and volatility forecasting', *The Review of Economics and Statistics* **83**(4), 596–602.

[14] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. [1997], *The Econometrics of Financial Markets*, Princeton Univeristy Press, Princeton.

[15] Chernov, M. [2007], 'On the role of risk premia in volatility forecasting', *Journal of Business and Economic Statistics*. forthcoming.

[16] Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, G. [2003], 'Alternative models for stock price dynamics', *Journal of Econometrics* **116**(1-2), 225–257.

[17] Christensen, K. and Podolskij, M. [2006], 'Realized range-based estimation of integrated variance', *Journal of Econometrics*. Forthcoming.

[18] Christodoulakis, G. A. and Satchell, S. E. [2004], 'Forecast evaluation in the presence of unobserved volatility', *Econometric Reviews* **23**(3), 175–198.

[19] Christoffersen, P. [2007], Estimation of value-at-risk, *in* 'Handbook of Financial Time Series', Springer. *Forthcoming*.

[20] Christoffersen, P. F. and Jacobs, K. [2004*a*], 'The importance of the loss function in option valuation', *Journal of Financial Economics* **72**(3), 291–318.

[21] Christoffersen, P. F. and Jacobs, K. [2004*b*], 'Which garch model for option valuation?', *Management Science* **50**, 1204–1221.

[22] de Jong, R. M. [1996], 'The Bierens test under data dependence', *Journal of Econometrics* **72**(1-2), 1–32.

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

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

[25] Engle, R. F. and Manganelli, S. [2004], 'Caviar: conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**, 367–381.

[26] Ferreira, M. A. and Lopez, J. A. [2005], 'Evaluating interest rate covariance models within a value-at-risk framework', *Journal of Financial Econometrics* **3**, 126–168.

[27] Fleming, J., Kirby, C. and Ostdiek, B. [2003], 'The economic value of volatility timing using "realized" volatility', *Journal of Financial Economics* **67**(3), 473–509.

[28] Garcia, R., Ghysels, E. and Renault, E. [2008], The econometrics of option pricing, *in* Y. Aït-Sahalia and L. P. Hansen, eds, 'Handbook of Financial Econometrics', Elsevier-North Holland, Amsterdam. Forthcoming.

[29] Giacomini, R. and White, H. [2006], 'Tests of conditional predictive ability', *Econometrica* **74**(6), 1545–1578.

[30] Gibson, M. S. and Boyer, B. H. [1997], Evaluating forecasts of correlation using option pricing. Board of Governors of the Federal Reserve System (U.S.).

[31] Gonçalves, S. and Meddahi, N. [2005], Bootstrapping realized volatility. Département de sciences économiques, CIREQ and CIRANO, Université de Montréal.

[32] Gonźalez-Rivera, G., Lee, T.-H. and Mishra, S. [2004], 'Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood', *International Journal of Forecasting* **20**(4), 629–645.

[33] Griffin, J. E. and Oomen, R. C. [2006], Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Mimeo.

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

[35] Hansen, P. R. and Lunde, A. [2005], 'A forecast comparison of volatility models: does anything beat a GARCH(1,1)?', *Journal of Applied Econometrics* **20**(7), 873–889.

[36] Hansen, P. R. and Lunde, A. [2006*a*], 'Consistent ranking of volatility models', *Journal of Econometrics* **127**(1-2), 97–121.

[37] Hansen, P. R. and Lunde, A. [2006*b*], 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218.

[38] Hansen, P. R., Lunde, A. and Nason, J. M. [2005], Model confidence sets for forecasting models. Federal Reserve Bank of Atlanta Working Paper 2005-7.

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

[40] Hayashi, F. [2000], *Econometrics*, Princeton University Press.

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

[42] Koopman, S. J. [2007], Parameter estimation and practical aspects of modelling stochastic volatility, *in* 'Handbook of Financial Time Series', Springer. *Forthcoming*.

[43] Kuester, K., Mittnik, S. and Paolella, M. S. [2006], 'Value-at-Risk Prediction: A Comparison of Alternative Strategies', *Journal of Financial Econometrics* **4**(1), 53–89.

[44] Lopez, J. A. [2001], 'Evaluating the predictive accuracy of volatility models', *Journal of Forecasting* **20**(2), 87–109.

[45] Magnus, J. R. and Neudecker, H. [2002], *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester, England.

[46] Marquering, W. and Verbeek, M. [2004], 'The economic value of predicting stock index returns and volatility', *Journal of Financial and Quantitative Analysis* **39**(2), 407–429.

[47] Meddahi, N. [2001], A theoretical comparison between integrated and realized volatilities. Manuscript, Département de sciences économiques, CIREQ and CIRANO, Université de Montréal.

[48] Mincer, J. and Zarnowitz, V. [1969], The evaluation of economic forecasts, *in* J. Mincer, ed., 'Economic Forecasts and Expectations', Columbia University Press.

[49] Newey, W. K. and West, K. D. [1987], 'A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

[50] Noh, J., Engle, R. F. and Kane, A. [1994], 'Forecasting volatility and option prices of the S&P 500 index', *Journal of Derivatives* **2**(1), 17–30.

[51] Parkinson, M. [1980], 'The extreme value method for estimating the variance of the rate of return', *The Journal of Business* **53**(1), 61–65.

[52] Patton, A. J. [2006], Volatility forecast comparison using imperfect volatility proxies. Quantitative Finance Research Centre, University of Technology Sydney, Research Paper 175.

[53] Sheppard, K. [2006], Realized covariance and scrambling. Univeristy of Oxford.

[54] Silvennoinen, A. and Teräsvirta, T. [2007], Multivariate garch models, *in* 'Handbook of Financial Time Series', Springer. *Forthcoming*.

[55] Theil, H. [1958], *Economic Forecasts and Policy*, North-Holland, Rotterdam.

[56] Tse, Y. K. [2000], 'A test for constant correlations in a multivariate GARCH model', *Journal of Econometrics* **98**(1), 107–127.

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

[58] West, K. D. [2006], Forecast evaluation, *in* G. Elliott, C. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting', North Holland Press, Amsterdam.

[59] West, K. D., Edison, H. J. and Cho, D. [1993], 'A utility-based comparison of some models of exchange rate volatility', *Journal of International Economics* **35**(1–2), 23–45.

[60] White, H. [1980], 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**(4), 817–838.

[61] White, H. [1996], *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, Cambridge University Press, Cambridge.

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

[63] Zivot, E. [2007], Practical aspects of garch-modelling, *in* 'Handbook of Financial Time Series', Springer. *Forthcoming*.