Consistent Preordering with an Estimated Criterion Function,

with an Application to the

Evaluation and Comparison of Volatility Models

Peter Reinhard Hansen[1]

and

Asger Lunde[2]

Working Paper 2003-01

January, 2003

Brown University

Department of Economics

[1]Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912, USA, Phone: (401) 863 9864, Email: Peter_Hansen@brown.edu.

[2]The Aarhus School of Business, Department of Information Science, Fuglesangs Allé 4 DK-8210 Aarhus V, Phone (+45) 89486688, Email: alunde@asb.dk.

Abstract[3]

When alternatives are compared using an estimated criterion function, this may introduce a discrepancy between the true and the estimated criterion.

In this paper, we consider a situation where a preordering (ranking) of stochastic sequences is defined from expected loss/gain, using a parametric criterion function. Evaluation based on estimated parameters induces a second preordering, and using sample averages in place of expectations induces a third (empirical) preordering, and we derive conditions that ensure equivalence of the three preorderings.

We apply the framework to the comparison of ARCH-type models. In practice, the conditional variance, σ_t^2, $t = 1, 2, \ldots$ is unobserved, such that evaluation must be based on a proxy for σ_t^2. We show that some commonly used criteria for evaluation of volatility models, may induce a different preordering than the one intended. This problem is cause by the measurement error of σ_t^2, which defines (part of) the empirical criterion. An empirical analysis and a simulation study show the practical relevance of this inconsistency problem. The results provide an additional argument for using intra-day data to approximate σ_t^2, such as realized volatility.

JEL Classification: C22; C52; C53; D0;

Keywords: Consistent Preordering; Model Comparison; Model Selection; Volatility Models.

[3] We thank Sean Campbell and Andrew Patton for comments. We are responsible for all errors. Financial support from the Danish Research Agency, grant no. 24-00-0363 and the Salomon Research Award at Brown University is gratefully acknowledged.

1 Introduction

The criterion function that define preferences is not always observed, so the unknown criterion function is often substituted by an approximate criterion function in empirical studies. Example include studies that involve economic entities, such as individual and firms, where the parameters of the unknown utility functions or production functions are estimated from observables. Another example is in the evaluation of volatility that involves an unobserved conditional variance, σ_t^2, which is substituted by a proxy for σ_t^2.

We consider a framework where the criterion function is parameterized such that the uncertainty about it, is expressed in terms of parameter-uncertainty. The expected values of the true criterion function and the approximate criterion function induce two preorderings of alternatives, and we refer to these preorderings as the *true preordering* and the *approximate preordering*, respectively. A third preordering, called the *empirical preordering*, is induced by an empirical comparison of alternatives, where the expected value have been approximated by a suitable sample average.

The substitution of a proxy for true parameters affects the ranking of alternatives in two ways. One effect is that noise is added to the evaluation, which makes it more difficult to identify the best of two alternative, and we refer to this effect as the *added-noise effect*. This effect creates a discrepancy between the approximate and the empirical preordering, however, this discrepancy vanishes asymptotically, under standard regularity conditions. The second effect is called the *inconsistency effect*, as the substitution creates a discrepancy between the true and the approximate preorderings of alternatives, unless certain conditions are meet.

In this paper, we are mainly concerned about the inconsistency effect. In empirical comparisons, where the inconsistency effect applies, the implication is that the approximate preordering, in part, is defined by the measurement errors of the criterion function's parameters. Hence, the actual preordering will differ from the intended preordering, and the discrepancy is defined by random measurement errors, which is highly unfortunate for obvious reasons.

In this setting, we formulate a sufficient set of conditions that ensure the equivalence of the true and the approximate preordering, and another set of conditions, which ensure that the empirical preordering is consistent for the approximate preordering. When these conditions are not meet, the empirical preordering may be inconsistent, e.g., what is asymptotically the best alternative according to the empirical preordering may not be the best alternative as prescribed by the true preordering.

We illustrate this problem with an empirical comparison of volatility models. Some criteria that are commonly used to compare volatility model do not satisfy the required conditions. An example, is the R^2-criterion when the R^2s are calculated from Mincer-Zarnowitz regressions using logarithmically transformed variables. When an empirical ranking of volatility models is based on a criterion, which do not satisfy the required conditions, it it possible that a volatility model, which is inferior to other volatility models, is found to be "significantly" better than all other models, with a probability that converges to one.

In the context of volatility models, it has been shown that using high-frequency data for the construction of precise measures of σ_t^2, can greatly reduce the added-noise effect, see e.g. Andersen and Bollerslev (1998). When the criterion, which is used to evaluate the volatility models, do not satisfy the required conditions, the inconsistency effect provides an additional argument for using high-frequency data to measure the conditional variance, σ_t^2. It will typically be the case, that the smaller is the variance of the measurement error, the smaller is the discrepancy between the true and the approximate ranking of volatility models. So an inconsistency may be avoided by using a precise measure of σ_t^2 in the model evaluation.

We use the following notation. All random variables are defined on the probability space, $(\Omega, \mathcal{F}, P)$. Thus, a random variable, $X(\omega)$, is a measurable mapping, $X : (\Omega, \mathcal{F}) \curvearrowright (\Gamma, \mathcal{G})$, where $(\Gamma, \mathcal{G})$ is a measurable space, and we will often suppress the dependence on ω, and simply write X in place of $X(\omega)$. Statements that are said to hold *almost surely*, (*a.s.*), refer to the existence of a set $G \in \mathcal{F}$, with $P(G) = 1$, for which the statement is true for all $\omega \in G$.

This paper is organized as follows. In Section 2 we define the set of alternatives and the various preorderings of alternatives, and we derived sufficient conditions under which the preorderings are equivalent. In Section 3, we apply our framework to evaluations and comparison of volatility models, and Section 4 contains an empirical and a simulation-based comparison of volatility models, and both comparisons show the practical relevance of the inconsistency effect. Section 5 contains concluding remarks.

2 Theoretical Framework

Let $\mathcal{X}$ be a random variable that is evaluated through the expected value of some loss function, $L(\mathcal{X})$. We consider a situation where the loss function, L, is unobserved, such that the evaluation of $\mathcal{X}$ must be based on some approximation of L. We denote the proxy for the true loss function by $\tilde{L}$, and we seek to establish conditions that will ensure that,

$$E(L(\mathcal{X})) \geq E(L(\mathcal{Y})), \quad \text{if and only if} \quad E(\tilde{L}(\mathcal{X})) \geq E(\tilde{L}(\mathcal{Y})), \quad \text{for all } \mathcal{X} \text{ and } \mathcal{Y}.$$

We formalize this idea in a setting where $\mathcal{X}$ and $\mathcal{Y}$ represent sequences of random variables that are being evaluated and compared in terms of their average expected loss.

Definition 1 (Set of Alternatives) *The set of alternatives, $\mathbb{A}$, is a set of random sequences. A typical element of $\mathbb{A}$ is $\mathcal{X}(\omega) = \{X_1(\omega), X_2(\omega), \ldots\}$, which is defined on a probability space $(\Omega, \mathcal{F}, P)$ and takes values in $(\mathbb{S}^\infty, \mathcal{B}_\infty)$, where $\mathbb{S}^\infty \equiv \mathbb{S} \times \mathbb{S} \times \cdots$, where $\mathbb{S} \subseteq \mathbb{R}^l$, and where $\mathcal{B}_\infty$ is the Borel σ-algebra on $\mathbb{S}^\infty$.*[1]

Initially, we make the following assumptions about the two loss functions.

Assumption 1 *Let L_t and $\tilde{L}_t$ be mappings from $\mathbb{S}$ into $\mathbb{R}$, $t = 1, 2, \ldots$, and define the random variable $\hat{\psi}_n(\mathcal{X}) \equiv n^{-1} \sum_{t=1}^n \tilde{L}_t(X_t)$.*

For all $\mathcal{X} \in \mathbb{A}$ it holds that:

(i) *$L_t(X_t)$ and $\tilde{L}_t(X_t)$ are integrable for all t;*

(ii) *The limits,*

$$\psi(\mathcal{X}) \equiv \lim_{n\to\infty} n^{-1} \sum_{t=1}^n E\left[L_t(X_t)\right], \quad \text{and} \quad \tilde{\psi}(\mathcal{X}) \equiv \lim_{n\to\infty} n^{-1} \sum_{t=1}^n E\left[\tilde{L}_t(X_t)\right],$$

exist and are finite;

(iii) *The limit $\hat{\psi}(\mathcal{X}) \equiv \lim_{n\to\infty} \hat{\psi}_n(\mathcal{X})$ exists and is finite a.s.*

This assumption allows us to define two non-stochastic preorderings on $\mathbb{A}$, and a sequence of stochastic preorderings. We shall refer to the non-stochastic preorderings as the *true preordering*, $\succeq$, and the *approximate preordering*, $\overset{a}{\succeq}$, where the latter can be thought of as an approximation of the former. The stochastic preorderings are referred to as the *empirical preorderings*, $\overset{e}{\succeq}_n$, $n = 1,\ 2,\ \ldots$.

Definition 2 (Preordering of Alternatives) *For $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$ we write: $\mathcal{X} \succeq \mathcal{Y}$ if $\psi(\mathcal{X}) \leq \psi(\mathcal{Y})$; $\mathcal{X} \overset{a}{\succeq} \mathcal{Y}$ if $\tilde{\psi}(\mathcal{X}) \leq \tilde{\psi}(\mathcal{Y})$; and we write $\mathcal{X} \overset{e}{\succeq}_n \mathcal{Y}$ if $\hat{\psi}_n(\mathcal{X}) \leq \hat{\psi}_n(\mathcal{Y})$, where $\overset{e}{\succeq}_n$ is a stochastic preorderings, as it depends on the realized value of ω, $n = 1, 2, \ldots$.*

It is easy to verify that the preorderings of Definition 2 are complete preorderings, and we shall follow standard notation and write $\mathcal{X} \sim \mathcal{Y}$ if "$\mathcal{X} \succeq \mathcal{Y}$ and $\mathcal{X} \preceq \mathcal{Y}$", and we write $\mathcal{X} \succ \mathcal{Y}$ if "$\mathcal{X} \succeq$

[1] Thus $\mathcal{X}$ is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{S}^\infty, \mathcal{B}_\infty)$, P is a probability on $(\Omega, \mathcal{F})$, and $\mathcal{B}_\infty$ is the smallest σ-algebra that contains all open subsets of $\mathbb{S}^\infty$ under the euclidian norm.

$\mathcal{Y}$ and $\mathcal{X} \not\preceq \mathcal{Y}$", and similarly for the approximate preordering, $\overset{a}{\succeq}$, and the empirical preorderings, $\overset{e}{\succeq}_n$, $n = 1,\ 2,\ \ldots.$

We define equivalence of preorderings, and the interesting situation is when the preorderings are equivalent in some sense.

Definition 3 (Equivalent and Weakly Equivalent) *Let $\succeq'$ and $\succeq''$ be preorderings and let $\succeq''_n$, $n = 1, 2, \ldots$ be a sequence of stochastic preorderings of $\mathbb{A}$. If it, for all $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$, holds that:*

$a)$	$\mathcal{X} \succeq' \mathcal{Y} \Leftrightarrow \mathcal{X} \succeq'' \mathcal{Y}$,	*then* $\succeq'$ *and* $\succeq''$ *are* equivalent *on* $\mathbb{A}$;
$b)$	$\mathcal{X} \succ' \mathcal{Y} \Leftrightarrow \mathcal{X} \succ'' \mathcal{Y}$,	*then* $\succeq$ *and* $\succeq'$ *are* weakly equivalent *on* $\mathbb{A}$;
$c)$	$\mathcal{X} \succeq' \mathcal{Y} \Rightarrow P(\mathcal{X} \succeq''_n \mathcal{Y}) \underset{n\to\infty}{\to} 1$,	*then* $\succeq''_n$ *is* asymptotically equivalent *to* $\succeq$ *on* $\mathbb{A}$;
$d)$	$\mathcal{X} \succ' \mathcal{Y} \Rightarrow P(\mathcal{X} \succ''_n \mathcal{Y}) \underset{n\to\infty}{\to} 1$,	*then* $\succeq''_n$ *is* asymptotically weakly equivalent *to* $\succeq'$ *on* $\mathbb{A}$.

The difference between equivalence and weak equivalence is that the former makes a statement about alternative for which $\mathcal{X} \sim \mathcal{Y}$, whereas the latter does not. The concept of asymptotic equivalence is useful for the analysis of empirical preorderings. It should be noted that the definitions are specific to the set of alternatives, $\mathbb{A}$, under considerations. E.g., two preorderings that are equivalent on $\mathbb{A}$, may not be equivalent on a larger set of preorderings $\mathbb{A}'$.

Lemma 4 *Define $\gamma(\mathcal{X}) \equiv \psi(\mathcal{X}) - \tilde{\psi}(\mathcal{X})$ and $\gamma_n(\mathcal{X}) \equiv \tilde{\psi}(\mathcal{X}) - \hat{\psi}_n(\mathcal{X})$, (where the latter is random). (i) If $\delta(\mathcal{X}, \mathcal{Y}) \equiv \gamma(\mathcal{X}) - \gamma(\mathcal{Y}) = 0$ for all $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$, then $\succeq$ and $\overset{a}{\succeq}$ are equivalent. (ii) If $\delta_n(\mathcal{X}, \mathcal{Y}) \equiv \gamma_n(\mathcal{X}) - \gamma_n(\mathcal{Y}) \overset{a.s.}{\to} 0$, as $n \to \infty$, for all $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$, then $\overset{e}{\succeq}_n$ is asymptotically weakly equivalent to $\overset{a}{\succeq}$ on $\mathbb{A}$.*

So $\delta(\mathcal{X}, \mathcal{Y})$ can be interpreted as a measure of discrepancy between $\succeq$ and $\overset{a}{\succeq}$, and similarly, $\delta_n(\mathcal{X}, \mathcal{Y})$ can be interpreted as a measure of discrepancy between $\overset{a}{\succeq}$ and the limit of $\overset{e}{\succeq}_n$. Note that mild variation in δ need not distort the equivalence of preorderings, as long as $|\delta(\mathcal{X}, \mathcal{Y})| < |\gamma(\mathcal{X}) - \gamma(\mathcal{Y})|$ for all $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$. So the conditions of Lemma 4 are sufficient conditions, but need not be necessary conditions.

2.1 Equivalence under Parametric Specification

We make additional assumptions about the true criterion function, L_t, the observed criterion function, $\tilde{L}_t$, and the relation between the two. For example, we assume that the loss functions have the same parametric form.

Assumption 2 *Let θ_t and $\tilde{\theta}_t$ denote two (possibly random) variables.*

(i) *For all* $\mathcal{X} \in \mathbb{A}$, *it holds that* $L_t(X_t) \overset{a.s.}{=} L(\theta_t, X_t)$ *and* $\tilde{L}_t(X_t) \overset{a.s.}{=} L(\tilde{\theta}_t, X_t)$, $t = 1, 2, \ldots$.

Define $\eta_t \equiv \tilde{\theta}_t - \theta_t$ *and let* $\{\mathcal{F}_t\}$ *be a filtration, such that for all* $\mathcal{X} \in \mathbb{A}$, *it holds that* X_t *and* θ_t *are* $\mathcal{F}_{t-1}$*-measurable,* $t = 1, 2, \ldots$.

(ii) *Either,*

(a) $L'(\theta, X) \equiv \frac{\partial L(\theta, X)}{\partial \theta}$ *exists and does not depend on* X; *or*

(b) $L''(\theta, X) \equiv \frac{\partial^2 L(\theta, X)}{\partial \theta \partial \theta'}$ *exists, does not depend on* X, *and* $\{\eta_t, \mathcal{F}_t\}$ *is a martingale difference sequence*.

Assumption 2 (i) requires that L_t and $\tilde{L}_t$ have the same parametric form, such that the uncertainty about L_t is entirely expressed in terms of uncertainty about the parameter θ_t. We call θ_t and $\tilde{\theta}_t$ parameters although both may be random variables, much like a conditional mean or a conditional variance. Assumptions 2 $(ii.a)$ and 2 $(ii.b)$ are assumptions about the functional form of L, linear and quadratic, respectively.

Theorem 5 *Under Assumptions 1* $(i$*-*$ii)$ *and 2, the true and the approximate preorderings,* $\succeq$ *and* $\overset{a}{\succeq}$, *are equivalent. Assumptions 1 and 2* (i) *alone, are not sufficient conditions for* $\succeq$ *and* $\overset{a}{\succeq}$ *to be equivalent.*

It is interesting to elaborate on the situation where Assumption 2 (ii) does not hold. A consequence is that an increase in the measurement error, as measured by $\mathrm{var}(\eta_t)$, will tend to increase the discrepancy between $\succeq$ and $\overset{a}{\succeq}$.

Corollary 6 *Let Assumptions 1* $(i$*-*$ii)$ *and 2* (i) *hold, and suppose that Assumption 2* (ii) *is violated. Let the approximate preordering,* $\tilde{\Psi}_\lambda$, *be defined by* $\tilde{\theta}_{\lambda,t} \equiv \theta_t + \lambda \eta_t$, $t = 1, \ldots, n$, *where* $E(\eta_t | \mathcal{F}_{t-1}) \overset{a.s.}{=} 0$ *and* $\mathrm{var}(\eta_t | \mathcal{F}_{t-1}) > 0$, $a.s.$ *and define the discrepancy measure*

$$\delta_\lambda(\mathcal{X}, \mathcal{Y}) \equiv [\psi(\mathcal{X}) - \tilde{\psi}_\lambda(\mathcal{X})] - [\psi(\mathcal{Y}) - \tilde{\psi}_\lambda(\mathcal{Y})].$$

(i) *If the second derivative,* $\partial^2 L(\theta, X)/\partial\theta\partial\theta'$, *is bounded a away from zero, uniformly in* X, $a.s.$, *then for some alternatives,* $\mathcal{X}$ *and* $\mathcal{Y}$, *it holds that* $|\delta_\lambda(\mathcal{X}, \mathcal{Y})| \to \infty$ *as* $\lambda \to \infty$.

(ii) *Under certain regularity conditions, see Assumption 4 in the Appendix, it holds that* $|\delta_\lambda(\mathcal{X}, \mathcal{Y})|$ *is strictly increasing in* $|\lambda|$, *for some* $\mathcal{X}, \mathcal{Y} \in \mathbb{A}$.

The broad message of Corollary 6, is that an increase in the conditional variance, $\mathrm{var}(\theta_t - \tilde{\theta}_{\lambda,t} | \mathcal{F}_{t-1})$, is likely to cause an inconsistency between $\succeq$ and $\overset{a}{\succeq}$, when Assumption 2 (ii) does not hold. For a detailed form of $\delta_\lambda(\mathcal{X}, \mathcal{Y})$ see (11) in the appendix.

2.2 Consistency of the Empirical Preordering

Without knowledge about the probability measure, P, it is not possible to evaluate expected values, such as $E\left[L(\theta, X)\right] = \int L(\theta, X) dP$, and it is therefore not directly possible to rank alternatives in terms of $\succeq$ or $\overset{a}{\succeq}$. However, under regularity conditions the expected value can be approximated by a sample average, such that the empirical preordering, $\overset{e}{\succeq}_n$, asymptotically resemble $\overset{a}{\succeq}$.

Theorem 7 *Let Assumptions 1 and 2.i hold and suppose that $L(\tilde{\theta}_t, X_t)$, $t = 1, 2, \ldots,$ is stationary and ergodic for all $\mathcal{X} \in \mathbb{A}$. Then $\overset{e}{\succeq}_n$ is asymptotically weakly equivalent to $\overset{a}{\succeq}$ almost surely.*

For practical implementations, the interesting situation is when, $\overset{a}{\succeq}$ is equivalent to $\succeq$, such that $\overset{e}{\succeq}_n$ is asymptotically equivalent to the true preferences $\succeq$, with probability one.

3 Comparison of Volatility Models

In this section, we show that our framework for preordering of stochastic sequences, yields valuable insight to the problem of comparing volatility models. Some, but not all, of the popular criteria for evaluating volatility models do satisfy the conditions we formulated in the previous section. For these criteria, it holds that an empirical ranking of alternatives is consistent for the intended ranking of alternatives. On the other hand, if a criterion does not satisfy the needed conditions, an inconsistency may arise, and there are very strong arguments for using a proxy of σ_t^2, which is unbiased and has the smallest possible conditional variance. The reason is that the distortion of the empirical evaluation increases with the variance of the measurement error $\eta_t = \tilde{\sigma}_t^2 - \sigma_t^2$.

The literature contains a vast number of studies that evaluate and compare volatility models, see, e.g., Poon and Granger (202) that contains a review of 93 papers. Most papers apply a loss function to compare model-based predictions of volatility, $\{h_t^2\}$, to proxies for volatility, $\{\tilde{\sigma}_t^2\}$, where the latter are measured ex-post. Common loss functions are: mean square (prediction) error (MSE), root mean squared error (RMSE), mean absolute error (MAD), and logarithmic versions of these, where the loss functions take log-volatilities as the arguments. Another approach to the evaluation is to base it on the R^2 from simple regressions, as suggested by Mincer and Zarnowitz (1969). In this approach, a proxy for the conditional variance is regressed on a model forecast of volatility and a constant. Pagan and Schwert (1990) noted that this regression is sensitive to "outliers" of the proxies, assuming that estimation is made using the least squares method. This point was also made by Engle and Patton (2001). The problem is that the parameter estimates are affected disproportionately by the largest realizations (outliers) of the proxy, $\tilde{\sigma}_t^2$. For this reason, Pagan and Schwert (1990) suggest to use a log-regression, where $\log(\sigma_t^2)$ is regressed on $\log(h_t^2)$ and a constant, as this regression is less sensitive to "outliers". An influential paper that applied Mincer-Zarnowitz regressions is Andersen and Bollerslev (1998), and Hansen and Lunde (2001) contains an extensive evaluation of volatility models using loss functions as well as Mincer-Zarnowitz regressions.

3.1 THE FRAMEWORK

From a continuously compounded price process, $\{p_t\}$, $t \geq 0$, we define daily returns $r_t \equiv p_t - p_{t-1}$, $t = 1, 2, \ldots$ and the σ-algebra, $\mathcal{F}_t \equiv \sigma(r_t, r_{t-1}, \ldots)$, such that r_t is adapted to $\mathcal{F}_t$. We assume that r_t has finite second moment such that it is meaningful to define $\sigma_t^2 \equiv \text{var}(r_t|\mathcal{F}_{t-1})$. For simplicity, we also assume that $E(r_t|\mathcal{F}_{t-1}) = 0$, such that $\{r_t, \mathcal{F}_t\}$ is a martingale difference sequence.

We consider volatility models that are designed to describe the variation in the conditional variance σ_t^2, $t = 1, 2, \ldots$. Each model produces a "forecast", h_t^2, of σ_t^2, where h_t^2 is $\mathcal{F}_{t-1}$-measurable, $t = 1, 2, \ldots$. So our set of alternatives, $\mathbb{A}$, consists of different sequences, $\mathcal{H} = (h_1^2, h_2^2, \ldots)$, (one sequence for each model).

An immediate obstacle for evaluating a volatility model is the fact that σ_t^2 is unobserved. The solution is to substitute a proxy for σ_t^2, such as squared daily returns, $\tilde{\sigma}_t^2 = r_t^2$, or $\tilde{\sigma}_t^2 = (r_t - \hat{\mu}_t)^2$, where $\hat{\mu}_t$ is an estimate of the conditional mean, $E(r_t|\mathcal{F}_{t-1})$. It is not surprising that squared daily returns produce a rather noisy measure of σ_t^2, $t = 1, 2, \ldots$. In fact, when volatility models are evaluated using squared daily returns, it results in, (what appears to be), a very poor performance. Better choices for $\tilde{\sigma}_t^2$ are measures that incorporate the additional information that intra-day returns have to offer about σ_t^2. The simplest extension is the range-based proxy for σ_t^2, which is based on the "open", "low", "high", and "close" prices for day t, see, e.g., Parkinson (1983), Garman and Klass (1983), Beckers (1983), Ball and Torous (1984), Rogers and Satchell (1991), Wiggins (1991), Kunitomo (1992), Gallant, Hsu, and Tauchen (1999), and McLeish (2002).

A better, but also more computational intensive, measure of daily volatility is the *realized volatility*, see e.g., Andersen and Bollerslev (1998). Realized volatility is constructed by taking the sum of squared intra-day returns, which produces an unbiased measure of the conditional variance, σ_t^2, (conditional on $\mathcal{F}_{t-1}$) under suitable regularity conditions.

It is clear that this problem fits into the framework of the previous section, where σ_t^2 and $\tilde{\sigma}_t^2$ play the role of θ_t and $\tilde{\theta}_t$, respectively, and where the set of alternative, $\mathbb{A}$, is given by the sequences of forecasts, $\mathcal{H} = (h_1^2, h_2^2, \ldots)$.

3.2 Evaluation Based on Loss Functions

We consider two loss functions, the mean squared error (MSE) loss function, $L(\sigma_t^2, h_t^2) = (\sigma_t^2 - h_t^2)^2$, and the logarithmic version of it (MSE*), which is given by $L(\sigma_t^2, h_t^2) = \left[\log\left(\sigma_t^2\right) - \log(h_t^2)\right]^2$. Both are common in the literature, and in the context of evaluating volatility models, it is sometimes argued that the latter is better than former because it is less sensitive to outliers.

MSE: Mean Squared Error Loss

Consider the loss function

$$L(\sigma_t^2, h_t^2) = (\sigma_t^2 - h_t^2)^2. \tag{MSE}$$

A Taylor expansion of the approximating loss function, $L(\tilde{\sigma}_t^2, h_t^2) = (\tilde{\sigma}_t^2 - h_t^2)^2$, is given by

$$L(\tilde{\sigma}_t^2, h_t^2) = L(\sigma_t^2, h_t^2) + 2(\sigma_t^2 - h_t^2)\eta_t + \eta_t^2,$$

where $\eta_t = \tilde{\sigma}_t^2 - \sigma_t^2$. So Assumption 2 $(ii.b)$ is satisfied whenever $\tilde{\sigma}_t^2$ is conditionally unbiased for σ_t^2, and we can conclude that $L(\sigma^2, \cdot)$ and $L(\tilde{\sigma}^2, \cdot)$ induce the same preordering. In particular, if we index the competing volatility models by k, we have that

$$\arg\min_k E\left[n^{-1}\sum_{t=1}^{n} L(\sigma_t^2, h_{k,t}^2)\right] = \arg\min_k E\left[n^{-1}\sum_{t=1}^{n} L(\tilde{\sigma}_t^2, h_{k,t}^2)\right],$$

i.e., the volatility model (or models) with the smallest population loss for the true and the approximate loss functions, coincide.

MSE*: Mean Squared Log Relative Error Loss

Consider now, the loss function given by

$$L(\sigma_t^2, h_t^2) = \left[\log\left(\frac{h_t^2}{\sigma_t^2}\right)\right]^2 = \left[\log\left(\sigma_t^2\right) - \log(h_t^2)\right]^2. \tag{MSE*}$$

The relevant derivatives are given by,

$$\frac{\partial L}{\partial \sigma_t^2} = 2\frac{\log(\sigma_t^2/h_t^2)}{\sigma_t^2}, \qquad \text{and} \qquad \frac{\partial^2 L}{\partial \sigma_t^2 \partial \sigma_t^2} = 2\frac{1 - \sigma_t^2 \log(\sigma_t^2/h_t^2)}{\sigma_t^4},$$

which do not satisfy Assumption 2 (ii).

Given the failure to satisfy Assumption 2 (ii), we conclude that evaluation based on this loss function may result in an inconsistent ranking of volatility models, even if an conditionally unbiased proxy for the unobserved conditional variance, σ_t^2, is employed ($E(\eta_t|\mathcal{F}_{t-1}) = 0$). Inconsistencies are more likely to arise the larger is $E(\eta_t^2|\mathcal{F}_{t-1})$, in fact a large conditional variance causes the the approximate evaluation to favor models for which $E[\log h^2]$ is relatively small. This can be seen from the Taylor expansion of this loss function. Naturally, if $\log(\tilde{\sigma}_t^2)$ is conditionally unbiased for $\log(\sigma_t^2)$, then we have a situation that is identical to that of the MSE loss function, where Assumption 2 (ii) is satisfied.

3.3 Regression Based Evaluation

An alternative to using loss functions for evaluating volatility models, is to use the R^2 of the Mincer-Zarnowitz regressions, which take the form

$$\varphi(\sigma_t^2) = \alpha + \beta\varphi(h_t^2) + u_t, \qquad t = 1, \dots, n. \tag{1}$$

Common choices for the function, φ, are: the identity, $\varphi(x) = x$, and the logarithmic transformation, $\varphi(x) = \log(x)$, which leads to the two regression equations

$$\sigma_t^2 = \alpha + \beta h_t^2 + u_t, \qquad t = 1, \dots, n, \tag{2}$$

$$\log \sigma_t^2 = \alpha + \beta \log h_t^2 + u_t, \qquad t = 1, \dots, n, \tag{3}$$

respectively. As mentioned earlier, Pagan and Schwert (1990) have argued in favor of the log-regression (3) as the level-regression (2) may be sensitive to extreme (large) values of r_t^2.

It is simple to verify that the R^2-criterion is equivalent to the criterion given by,

$$\psi(\mathcal{H}) \equiv \mathrm{cov}(\varphi(\sigma_t^2), \varphi(h_t^2)) \left[\mathrm{var}(\varphi(h_t^2))\right]^{-1} \mathrm{cov}(\varphi(h_t^2), \varphi(\sigma_t^2)). \tag{4}$$

When $\tilde{\sigma}_t^2$ is substituted for σ_t^2, the approximate criterion is proportional to

$$\tilde{\psi}(\mathcal{H}) \equiv \mathrm{cov}(\varphi(\tilde{\sigma}_t^2), \varphi(h_t^2)) \left[\mathrm{var}(\varphi(h_t^2))\right]^{-1} \mathrm{cov}(\varphi(h_t^2), \varphi(\tilde{\sigma}_t^2)), \tag{5}$$

The underlying criterion of (4) and (5) does not fit directly into our framework of Section 2. Nevertheless, we shall derive condition that ensure consistency, which are analog to those of the previous section.

Assumption 3 *Let $\eta_t \equiv \sigma_t^2 - \tilde{\sigma}_t^2$. The function, φ, in (1) is infinite differentiable,*[2] *and it holds that*

$$E\left((\eta_t)^j \,|\mathcal{F}_{t-1}\right) \left.\frac{\partial^j \varphi(x)}{\partial (x)^j}\right|_{x=\sigma_t^2} \overset{a.s.}{=} c_j,$$

for some constant $c_j, \in \mathbb{R}$, for all $t = 1, 2, \dots,$ and all $j = 1, 2, \dots$.

[2]By infinite differentiable is meant that the pth derivative of φ exists for any integer, p.

This assumption is simply to interpret in special cases. Below we consider the three cases where φ is linear, quadratic, and logarithmic respectively.

Quadratic φ. Suppose that φ is quadratic, such that the second order derivative, φ'', is constant and higher order derivatives are all zero. The conditions in Assumption 3, then simplifies to $E\left(\eta_t|\mathcal{F}_{t-1}\right)\varphi'(\sigma_t^2)$ and $E\left(\eta_t^2|\mathcal{F}_{t-1}\right)$ being equal to some constants, almost surely.

Theorem 8 *The criteria, $\psi(\mathcal{H})$ and $\tilde{\psi}(\mathcal{H})$, are equivalent under Assumption 3.*

Linear φ. Suppose that φ is linear, such that the first derivative, φ', constant and higher order derivatives are zero. Assumption 3 simply requires $E\left(\eta_t|\mathcal{F}_{t-1}\right)$ to be constant (almost surely).

So the R^2-criterion of (2) is not affected by a conditional bias of $\tilde{\sigma}_t^2$. This robustness does not come without a cost, as the criterion is unable to distinguish between the volatility models, $\mathcal{H}_1 = \{h_t^2\}$ and $\mathcal{H}_2 = \{a + bh_t^2\}$, (for any value of a and any value of $b \neq 0$).

Logarithmic φ. Suppose that the models are compared using the R^2s from the regressions (3). If the proxy, $\tilde{\sigma}_t^2$, is conditionally unbiased for σ_t^2, then it is unlikely that the R^2 from the feasible regressions, (3), will induce the same ranking of volatility models as the R^2 from the infeasible regression $\log\sigma_t^2 = a + b\log h_t^2 + u_t$, as Assumption 3 is not satisfied. This is clearly an unfortunate property of the log-regression criterion, in particular if squared returns, r_t^2, are substituted for the unobserved volatility measure, σ_t^2, as this leads to an $\eta_t = r_t^2 - \sigma_t^2$ with a large variance. Analog to the case with additive loss function, see Corollary 6, the distortion is aggravate as the conditional variance of η_t increases. The last point is easily seen from a closer inspection of the proof of Theorem 8.

Suppose now that $\tilde{\sigma}_t^2 = (1 - \upsilon_t)\sigma_t^2$, for some random variable, υ_t, [3] where the conditional moments of υ_t, given by $\kappa_j \equiv E(\upsilon_t^j|\mathcal{F}_{t-1})$, $j = 1, 2, \ldots$, are finite and constant. This implies that the conditional bias of $\log\tilde{\sigma}_t^2$, relative to $\log\sigma_t^2$, is constant. The measurement error is then given by $\eta_t = \upsilon_t\sigma_t^2$, and $E(\eta_t^j|\mathcal{F}_{t-1}) = E(\upsilon_t^j|\mathcal{F}_{t-1})(\sigma_t^2)^j$. Since $\frac{\partial^j \log(x)}{\partial x^j} = \frac{(-1)^{j-1}}{(j-1)!}x^{-j}$, we have that

$$E\left((\eta_t)^j|\mathcal{F}_{t-1}\right)\left.\frac{\partial^j\varphi(x)}{\partial(x)^j}\right|_{x=\sigma_t^2} = \kappa_j\frac{(-1)^{j-1}}{(j-1)!},$$

which is constant. So an measurement error of this kind will not create an inconsistency for this criterion.

The multiplicative structure of the error appear to be more appropriate, than the additive. For example, Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b), have shown that the variance of the realized volatility estimator increases with σ_t^2, see also Meddahi (2002).[4]

[3] One may impose that $\upsilon_t \leq 1$ $(a.s)$ to ensure that $\tilde{\sigma}_t^2$ is non-negative $(a.s)$.

[4] The correct statement is that the variance of realized volatility increases with intergrated volatility. However, the integrated

4 EMPIRICAL AND SIMULATION-BASED COMPARISONS OF ARCH-TYPE MODELS

In this section we explore the empirical relevance of the theoretical inconsistencies that we studied in the previous section. This is done by evaluating and comparing ARCH-type volatility models using a real data set and a large number of artificial data.

4.1 EMPIRICAL COMPARISON BASED ON IBM EQUITY RETURNS

We evaluate and compare eight ARCH-type volatility models, using IBM stock price data. These data were extracted from the Trade and Quote (TAQ) database.[5] In our estimation of intra-day volatility, we have used the mid-quotes between 9:30 AM and 4 PM. The sample period is January 3, 1995 through February 21, 2002, which adds up to a total of 1795 trading days. The models were estimated using the first 1250 observation (up until December 13, 1999) and the last 545 observations were used for the evaluation and comparison.

Our set of competing ARCH-type models comprises the ARCH of Engle (1982), the GARCH model by Bollerslev (1986), the threshold GARCH model (Thr.-GARCH) by Zakoian (1994), the EGARCH of Nelson (1991), the A-PARCH model that was proposed in Ding, Granger, and Engle (1993), the FIGARCH suggested by Baillie, Bollerslev, and Mikkelsen (1996), and the FIAPARCH of Tse (1998). Each model is estimated using inter-day returns and using the MSE and the MSE* loss functions and two Mincer-Zarnowitz regressions, the level-regression (2) and the log-regression (3).

To make the evaluation feasible, we need to substitute a proxy for the unobserved conditional variance, σ_t^2, and we apply three different proxies, $\tilde{\sigma}_{j,t}^2$, $j = 1, 2, 3$, that differs in terms of the associated variance of $\eta_{jt} \equiv \tilde{\sigma}_{j,t}^2 - \sigma_t^2$. The three proxies are denoted by

$$\begin{aligned}
\tilde{\sigma}^2_{[\text{intra}]t} &\equiv \hat{c} \cdot \tilde{\sigma}^2_{(f/m),t}, \\
\tilde{\sigma}^2_{[\text{+on.}]t} &\equiv \tilde{\sigma}^2_{(f/m),t} + (p_t^{open} - p_{t-1}^{close})^2, \\
\tilde{\sigma}^2_{[\text{inter}]t} &\equiv (p_t^{close} - p_{t-1}^{close})^2.
\end{aligned}$$

We shall not give a lengthy discussion of these proxies and how they are estimated. Details are presented in the Appendix, however, for the illustration of the results regarding consistent ranking of alternatives, it suffices to know that:

volatility is a conditionally unbiased estimator of the conditional variance σ_t^2.

[5] The TAQ database contains all trades and quotes in the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotation (Nasdaq) securities.

$\tilde{\sigma}^2_{[\mathbf{intra}]t}$ The term, $\tilde{\sigma}^2_{(f/m),t}$, is a measure of realized volatility during the hours that the market is open (only f our of m possible intra-day returns are available). A Fourier method is implemented for the calculation of $\tilde{\sigma}^2_{(f/m),t}$. The constant, $\hat{c}$, is an estimated scaling (a correction factor) that accounts for the partial availability of intra-day returns.

$\tilde{\sigma}^2_{[\mathbf{+on.}]t}$ Is also based on realized volatility, but rather than scaling, it adds the squared over-night return to account for the partial availability.

$\tilde{\sigma}^2_{[\mathbf{inter}]t}$ The simple squared daily return, r_t^2.

Under quite reasonable assumptions it holds that

$$\mathrm{var}(\tilde{\sigma}^2_{[\mathrm{intra}]t}|\mathcal{F}_{t-1}) \leq \mathrm{var}(\tilde{\sigma}^2_{[\mathrm{+on.}]t}|\mathcal{F}_{t-1}) \leq \mathrm{var}(\tilde{\sigma}^2_{[\mathrm{inter}]t}|\mathcal{F}_{t-1}).$$

So for criteria where the true and the approximate preferences do not coincide, the discrepancy should increase with the conditional variance of $\tilde{\sigma}^2_t$, that underlies the approximate evaluation.

4.2 Empirical Results

The empirical results of the evaluation of the eight volatility models are given in Tables 1 and 2. Table 1 contains the results based of the two loss functions and the analogous results for the regression-based criteria are presented in Table 2.

We shall refer to the MSE loss function and the regression in levels as *robust criteria*, as they satisfy conditions, where the true and the approximate preordering are equivalent, provided that the proxy for σ^2_t is conditionally unbiased. The two other criteria, MSE* and the $R^{\star 2}$ from the regression in logs, which do not satisfy the relevant conditions are referred to as *sensitive criteria*.

It is striking that the robust criteria all point to the same model as the best volatility model, for all three choices of $\tilde{\sigma}^2_t$, whereas the sensitive criteria points to different models. The MSE* points to the ARCH model as the best volatility models when squared returns are substituted for σ^2_t. Most people with knowledge about volatility models would not believe this result, as the ARCH model is unable to capture the persistence in the conditional variance. Given the large standard errors of the performance measures, it is not clear that any pair of models have a performance that is significantly different from another. However, the benefits from using precise measures of the conditional variance are clearly illustrated by the fact that the standard errors are much smaller for the realized volatility measure than is the case for the measure based on squared returns (in relative terms).

4.3 Simulation Based Comparison

Generating artificial data to evaluate and compare volatility models has two advantages. Firstly, by simulating the data we know the true conditional variance, which can be used in the evaluation; Secondly, a large number of samples can be generated such that an extensive comparison can be made.

In our simulation study, we generate artificial data from a GARCH(1,1) model and an EGARCH(1,1) model, where the models population parameter values were set to estimates we found in the analysis of IBM data[6]. The artificial samples are used to estimate and evaluate seven ARCH-type models. The models are evaluated in terms of out-of-sample loss using the MSE and MSE* loss function, and we use both squared returns, r_t^2, and the true conditional variance, σ_t^2, in the evaluation. Our set of competing ARCH-type models comprises the ARCH of Engle (1982), the GARCH model by Bollerslev (1986), the threshold GARCH model (Thr.-GARCH) by Zakoian (1994), the EGARCH of Nelson (1991), the A-PARCH model that was proposed in Ding, Granger, and Engle (1993). We also include a two-component GARCH(1,1) model (2Comp) and a two-component Thr.-GARCH(1,1) model (2CompThr) of Ding and Granger (1996).[7] The FIGARCH model is computational expensive to estimate, so we did not include this model in the simulation experiments.

The results are presented in Figures 1 and 2. Figure 1 contains the results from the artificial data that where generate using the GARCH model, and Figure 2 contains the analog results from the data that were generated with an EGARCH model. The simulations are based on 500 artificial samples, where 1000 observations were used for the initial estimation, and an additional 100, 250, or 500 observations were used for the out-of-sample evaluation. The models were estimated recursively.

From Figure 1 it can be seen that the true GARCH model is often found to have the smallest loss when the true conditional variance is used in the evaluation, which is true for both loss functions. However, with squared returns are substituted for the conditional variance in the evaluation, the GARCH model is less likely to have the smallest loss and there is a noticeable difference between the two loss functions. When the evaluation is based on 250 and 500 observations the GARCH is more likely to have a smaller

[6] The estimated parameter values used in the simulations are as follows:

$$\text{GARCH(1,1):}\quad \sigma_t^2 = \underset{(0.0075)}{0.038} + \underset{(0.01)}{0.094}\epsilon_{t-1}^2 + \underset{(0.0129)}{0.875}\,\sigma_{t-1}^2 \tag{6}$$

$$\text{EGARCH(1,1):}\quad \sigma_t^2 = -\underset{(0.011)}{0.105} + \underset{(0.014)}{0.141}\,(|\epsilon_{t-1}| - \underset{(0.07)}{0.512}\epsilon_{t-1})\sigma_{t-1}^{-1} + \underset{(0.004)}{0.974}\sigma_{t-1}^2 \tag{7}$$

[7] The actual form to these models are those implemented in S+, which are sligthly modified versions of those in Ding and Granger (1996), see Zivot and Wang (2003, p. 234) for details.

sample loss than any other model using the MSE loss function. This is in sharp contrast to the results for the MSE* loss function, where the Component Thr.-GARCH model very frequently out-performs all other models, and in particular the GARCH(1,1) model. The results from the simulations based on the EGARCH model that are presented in Figure 2, similarly shows that an inconsistency can arise.

5 Conclusion

We have considered evaluations and comparisons that are based on an estimated criterion function. Uncertainty about parameters that characterize the criterion functions have two effects, an added-noise effect and an inconsistency effect. The former increases the sample uncertainty, which makes it more difficult to tell good an bad models apart, and the latter may introduce an inconsistency. In this paper, we have separated the two effects by distinguishing between the true, the approximate, and the empirical preordering, and we have mainly been concerned about the inconsistency effect. We derived conditions under which the true and the approximate preordering are equivalent, and conditions that ensure that the empirical preordering is asymptotically weakly equivalent to the approximate preordering. Thus, under both sets of assumptions, the limit of the empirical preordering coincide with the true preordering.

As we indicated in the introduction, this framework applies to many econometric problems that involves (a large number of) economic entities, such as individuals or firms. This will be the case in situations where the individuals' utility functions or the firms' production functions are estimated (somewhat imprecisely) and the object of interest is an aggregate of the individual utility/production functions.

In the context of evaluation and comparison of volatility models, we have shown that certain criteria that are commonly applied in the evaluation of volatility models, do not meet the needed conditions, and may for this reason produce a different ranking of models, than the one intended, and may do so with a probability that approaches one, as the evaluation period is increased. The inconsistency, typically, increases with the conditional variance of the parameter estimators. This result provides an additional argument for using high frequency data when volatility models are being evaluated and compared.

In an empirical analysis of IBM equity return, we compared eight volatility models using three different proxies for the unobserved conditional variance, and found that criteria that are sensitive to estimation uncertainty, pointed to different models as the best volatility model, whereas the robust criteria all pointed to the same model. Simulation-based studies similarly showed that inconsistencies can arise for certain loss functions. These findings strongly indicate that the inconsistency problem does have practical relevance, as the empirical results are consistent with the predictions of the theoretical results.

A Appendix: Estimation of Daily Volatility

Let p_t be the log-price on some asset in which dividends (if any) are accumulated. For any t and any $\Delta > 0$, we define $r_{t,\Delta} \equiv p_t - p_{t-\Delta}$, which is the return over the time interval with length Δ, that preceded time t. We let the unit of time be a trading day (i.e. the length from close to close), such that $r_{t,\Delta}$ corresponds to a daily return when $\Delta = 1$. For integer values of t, which corresponds to the time of "close", we have that $r_{t,1} = r_t$ (daily returns that were defined previously).

Realized volatility (at frequency m), $\tilde{\sigma}^2_{m,t}$, is defined by

$$\tilde{\sigma}^2_{m,t} \equiv \sum_{i=0}^{m-1} r^2_{t-\frac{i}{m},\frac{1}{m}}, \qquad t = 1, 2, \ldots, \tag{8}$$

and for some integer m. (So $\Delta = \frac{1}{m}$ in this notation). Similarly the realized volatility over a period of partially available intra-day returns is define by

$$\tilde{\sigma}^2_{(f,m),t} \equiv \sum_{i=0}^{f-1} r^2_{t-\frac{i}{m},\frac{1}{m}}, \qquad t = 1, 2, \ldots.$$

Under the assumption that $E(r_t) = 0$, it holds that the squared inter-day return, $r_t^2 = \hat{\sigma}^2_{t,1}$, is a conditionally unbiased estimator of σ^2_t, and for $m \geq 2$, the realized volatility, $\tilde{\sigma}^2_{m,t}$, is conditionally unbiased for σ^2_t under the additional assumption that the intra-day returns, $r_{t-\frac{i}{m},\frac{1}{m}}$, $i = 0, \ldots, m-1$, are uncorrelated.

Realized volatility can also be calculated from unevenly spaced intra-day returns, where one example is

$$\tilde{\sigma}^2_{(f,m,+\text{on}),t} \equiv r^2_{t-\frac{f}{m},\frac{m-f}{m}} + \sum_{i=0}^{f-1} r^2_{t-\frac{i}{m},\frac{1}{m}},$$

which is also unbiased for σ^2_t, provided that intra-day returns are uncorrelated.

If $\tilde{\sigma}^2_{(f,m),t}$ is proportional to σ^2_t, then $c\cdot\tilde{\sigma}^2_{(f,m),t}$ is conditionally unbiased for σ^2_t, where $c = E(\sigma^2_t/\tilde{\sigma}^2_{(f,m),t})$, which is consistently estimated by $n^{-1}\sum_{t=1}^{n} r_t^2/\tilde{\sigma}^2_{(f,m),t}$, under certain regularity conditions. Note that some correlation across intra-day returns, is tolerated under the proportionality assumption. See also Hansen and Lunde (2001).

A.1 The Fourier Method

We apply a Fourier method to estimate realized volatilities. This method was suggested by Malliavin and Mancino (2002) and has previously been applied by Barucci and Reno (2002a) and Barucci and Reno (2002b). A short description of the method is the following.

Let the price process be defined from a diffusion process with bounded quadratic variation,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t),$$

where μ and σ are time-varying functions, and where $W(t)$ is a standard Brownian motion. The integrated volatility over an interval $[a, b]$ is defined by $\int_a^b \sigma(t)dt$, which, in general, is a random variable. If we let the unit of time be 24 hours, then $\int_t^{t+1} \sigma(t)dt$ is an unbiased estimator of the conditional variance, σ_{t+1}^2 (for integer values of t). As is the case for realized volatility, $\tilde{\sigma}_{m,t}^2$, defined in (8), the Fourier method approximates the integrated volatility. Let $p(t)$ be observed in the interval $[t_0, t_0 + 1]$ at the discrete points in time, $t_1 < t_2 < \ldots < t_N$. These points in time are mapped into the interval $[0, 2\pi]$, by defining, $\tau_i = 2\pi(t_i - t_0)/(t_N - t_0)$ for $i = 1, \ldots, N$.

The Fourier method is based on the identity

$$\frac{1}{2\pi}\int_0^{2\pi} \sigma(t)dt = a_0(\sigma^2),$$

where

$$a_0(\sigma^2) = \lim_{m\to\infty} \frac{\pi}{2m} \sum_{k=1}^{m} \left(a_k^2(dp) + b_k^2(dp)\right), \tag{9}$$

and where

$$\begin{aligned} a_k(dp) &= \frac{p(\tau_N) - p(\tau_1)}{\pi} + \frac{1}{\pi}\sum_{i=1}^{N-1} p(\tau_i)\left[\cos(k\tau_i) - \cos(k\tau_{i+1})\right], \\ b_k(dp) &= -\frac{1}{\pi}\sum_{i=1}^{N-1} p(\tau_i)[\sin(k\tau_i) - \sin(k\tau_{i+1})]. \end{aligned}$$

The estimate of the Fourier method is given by

$$\breve{\sigma}_{m,t}^2 \equiv \frac{\pi}{2m}\sum_{k=1}^{m}\left(a_k^2(dp) + b_k^2(dp)\right),$$

where we applied $m = 80$ in our estimation.

B Appendix of Proofs

Proof of Lemma 4. (i) Suppose that $\mathcal{X} \succeq \mathcal{Y}$ then $\psi(\mathcal{X}) \leq \psi(\mathcal{Y})$. But $\tilde{\psi}(\mathcal{X}) = \psi(\mathcal{X}) - \gamma(\mathcal{X}) = \psi(\mathcal{X}) - \gamma(\mathcal{X}) + \gamma(\mathcal{Y}) - \psi(\mathcal{Y}) + \tilde{\psi}(\mathcal{Y}) = \psi(\mathcal{X}) - \psi(\mathcal{Y}) + \tilde{\psi}(\mathcal{Y}) \leq \tilde{\psi}(\mathcal{Y})$, and this implies $\mathcal{X} \overset{a}{\succeq} \mathcal{Y}$. The other implication of (i) is shown similarly. (ii) Suppose that $\mathcal{X} \succ \mathcal{Y}$. Then there exists $\varepsilon > 0$, such that $\psi(\mathcal{X}) + \varepsilon \leq \psi(\mathcal{Y})$. Similar calculations to those in the proof of (i) leads to $\hat{\psi}_n(\mathcal{X}) + \varepsilon - \delta_n(\mathcal{X}, \mathcal{Y}) \leq \hat{\psi}_n(\mathcal{Y})$, and since $\delta_n(\mathcal{X}, \mathcal{Y}) \overset{a.s.}{\to} 0$ it holds for almost surely that $\hat{\psi}_n(\mathcal{X}) \leq \hat{\psi}_n(\mathcal{Y}) - \varepsilon/2 < \hat{\psi}_n(\mathcal{Y})$, for n sufficiently large. The other implication is proven similarly. ■

Proof of Theorem 5. Under Assumption $(ii.a)$ we consider the first order Taylor expansion of L, given by $L(\tilde{\theta}_t, X_t) = L(\theta_t, X_t) + L'(\theta_t^*, X_t)\eta_t$, where θ_t^* lies between $\tilde{\theta}_t$ and θ_t. Taking expected value yields

$$E\left[L(\tilde{\theta}_t, X_t)\right] = E\left[L(\theta_t, X_t)\right] + E\left[L'(\theta_t^*, X_t)\eta_t\right],$$

since the last term does not depend on X_t under Assumption $(ii.a)$, it holds that

$$E[L(\tilde{\theta}_t, X_t)] - E[L(\tilde{\theta}_t, Y_t)] = E[L(\theta_t, X_t)] - E[L(\theta_t, Y_t)], \tag{10}$$

for all (X_t, Y_t), which shows the equivalence in this case. Under Assumption $(ii.b)$ we consider the second order Taylor expansion of L, given by

$$L(\tilde{\theta}_t, X_t) = L(\theta_t, X_t) + L'(\theta_t, X_t)\eta_t + \frac{1}{2}\eta' L''(\theta_t^{**}, X_t)\eta_t,$$

where θ_t^{**} lies between $\tilde{\theta}_t$ and θ_t. Taking expected value yields

$$E\left[L(\tilde{\theta}_t, X_t)\right] = E\left[L(\theta_t, X_t)\right] + E\left[L'(\theta_t, X_t)\eta_t\right] + \frac{1}{2}E\left[\eta_t' L''(\theta^{**}, X_t)\eta_t\right],$$

where the last term does not depend on X_t, and where the second term is zero, as $E[L'(\theta_t, X_t)\eta_t|\mathcal{F}_{t-1}] = L'(\theta_t, X_t)E[\eta_t|\mathcal{F}_{t-1}] = 0$. So once again we have established the identity (10), which shows the equivalence of $\succeq$ and $\overset{a}{\succeq}$ in this case. ■

Proof of Theorem 7. The stationarity assumption implies that $E[L(\tilde{\theta}_1, X_1)] = n^{-1}\sum_{t=1}^{n} E[L(\tilde{\theta}_t, X_t)]$, and the ergodic theorem states that $n^{-1}\sum_{t=1}^{n} L(\tilde{\theta}_t, X_t) \overset{a.s.}{\to} E[L(\tilde{\theta}_1, X_1)]$, which completes the proof. ■

Let $\tilde{\theta}_{\lambda,t} \equiv \theta_t + \lambda\eta_t$, $t = 1, \ldots, n$, and suppose that L is twice differentiable with continuous derivatives. Consider the Taylor expansion,

$$L(\tilde{\theta}_{\lambda,t}, X_t) = L(\theta_t, X_t) + L'(\theta_t, X_t)\lambda\eta_t + L''(\theta_{\lambda,t}^*, X_t)\lambda^2\eta_t^2,$$

where $\theta_{\lambda,t}^* \in [\theta_t, \theta_t + \lambda\eta_t]$. Suppose that $E(\eta_t|\mathcal{F}_{t-1}) \overset{a.s.}{=} 0$ and that $\mathrm{var}(\eta_t|\mathcal{F}_{t-1}) > 0$, $a.s.$, such that

$$\begin{aligned} E[L(\tilde{\theta}_{\lambda,t}, X_t) - L(\theta_t, X_t)] &= E[L'(\theta_t, X_t)\lambda\eta_t] + E[L''(\theta_{\lambda,t}^*, X_t)\lambda^2\eta_t^2] \\ &= 0 + \lambda^2 E[L''(\theta_{\lambda,t}^*, X_t)\eta_t^2]. \end{aligned}$$

The last term need not simplify to $\lambda^2 E[L''(\theta_{\lambda,t}^*, X_t)]\mathrm{var}(\eta_t|\mathcal{F}_{t-1})$, since $\theta_{\lambda,t}^*$ is not $\mathcal{F}_{t-1}$-measurably (it depends on η_t).

Assumption 4 *For all* $\mathcal{X} \in \mathbb{A}$, *it holds that* (i) $L''(\theta^*_{\lambda,t}, X_t) > 0$, *almost surely; and*

$$(ii) \qquad L''(\tilde{\theta}^*_{\lambda_1,t}, X_t)/L''(\tilde{\theta}^*_{\lambda_2,t}, X_t) \leq \left(\frac{\lambda_2}{\lambda_1}\right)^2,$$

almost surely, for all $0 \leq \lambda_1 < \lambda_2 < \infty$.

Lemma 9 *Under Assumption 4, the criterion function for the approximate preordering, is given by*

$$E(L(\theta_t, X_t)) + \lambda^2 E[L''(\theta^*_{\lambda,t}, X_t)\eta_t^2],$$

and the measure of discrepancy

$$\delta_\lambda(\mathcal{X}, \mathcal{Y}) \equiv \lambda^2 \lim_{n\to\infty} n^{-1} \sum_{t=1}^n E\{[L''(\theta^*_{\lambda,t}, X_t) - L''(\theta^*_{\lambda,t}, Y_t)]\eta_t^2\}. \tag{11}$$

Proof of Theorem 8. Consider the Taylor expansions under each of the three conditions, (i) $\varphi(\tilde{\sigma}_t^2) = \varphi(\sigma_t^2) + \dot{\varphi}\eta_t$, where $\dot{\varphi} \equiv \partial\varphi/\partial\sigma_t^2$ is a constant; (ii) $\varphi(\tilde{\sigma}_t^2) = \varphi(\sigma_t^2) + \varphi'(\sigma_t^2)\eta_t + \frac{1}{2}\ddot{\varphi}\eta_t^2$, where $\ddot{\varphi} \equiv \partial^2\varphi/\partial\sigma_t^2\partial\sigma_t^2$ is a constant; and (iii) $\varphi(\tilde{\sigma}_t^2) = \varphi(\sigma_t^2) + \varphi'(\sigma_t^2)\eta_t + 2^{-1}\varphi''(\sigma_t^2)\eta_t^2 + 6^{-1}\dddot{\varphi}\eta_t^3$, where $\dddot{\varphi} \equiv \partial^3\varphi/(\partial\sigma_t^2)^2$ is a constant. It now follows that for each of the three cases, $\mathrm{cov}(\varphi(\tilde{\sigma}_t^2), \varphi(h_t^2)) - \mathrm{cov}(\varphi(\sigma_t^2), \varphi(h_t^2))$ equals $\dot{\varphi}c_1\sigma_{hh}$, $2^{-1}\ddot{\varphi}c_2\sigma_{hh}$, and $2^{-1}\sigma_\eta^2 c_3\sigma_{hh}$, respectively. For example under Assumption (iii) we have

$$\begin{aligned} \mathrm{cov}(\varphi(\tilde{\sigma}_t^2), \varphi(h_t^2)) &= \mathrm{cov}(\varphi(\sigma_t^2), \varphi(h_t^2)) + \mathrm{cov}(\varphi'(\sigma_t^2)\eta_t, \varphi(h_t^2)) \\ &\quad + 2^{-1}\mathrm{cov}(\varphi''(\sigma_t^2)\eta_t^2, \varphi(h_t^2)) + 6^{-1}\mathrm{cov}(\dddot{\varphi}\eta_t^3, \varphi(h_t^2)) \\ &= \mathrm{cov}(\varphi(\sigma_t^2), \varphi(h_t^2)) + 0 + 2^{-1}\sigma_\eta^2 c_3\sigma_{hh}^2 + 0, \end{aligned}$$

where $c_3 = \mathrm{cov}(\varphi''(\sigma_t^2), \varphi(h_t^2))$.

Thus $\tilde{\psi}(\mathcal{H}) = \psi(\mathcal{H}) + C$, where C under assumption (i), (ii), and (iii) equals $\dot{\varphi}^2c_1^2$, $\ddot{\varphi}^2c_2^2$, and $\dddot{\varphi}^2c_3^2$ respectively. ■

References

Andersen, T. G., and T. Bollerslev (1998): "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39(4), 885–905.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

Ball, C. A., and W. N. Torous (1984): "The maximum likelihood estimation of security price volatility: Theory evidence, and application to option pricing," *Journal of Business*, 57, 97–112.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002a): "Econometric analysis of realised volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society* **B**, 64, Forthcoming.

——— (2002b): "Estimating quadratic variation using realised volatility," *Journal of Applied Econometrics*, Forthcoming.

BARUCCI, E., AND R. RENO (2002a): "On measuring volatility and GARCH forecasting performance," *Journal of International Financial Markets*, 12, 183–200.

——— (2002b): "On measuring volatility of diffusion processes with high frequency data," *Economics Letters*, 74, 371–378.

BECKERS, S. (1983): "Variances of security price returns based on high, low, and closing prices," *Journal of Business*, 56(1), 97–112.

BOLLERSLEV, T. (1986): "Generalized autoregressive heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

DING, Z., AND C. W. J. GRANGER (1996): "Modelling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185–215.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

ENGLE, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica*, 45, 987–1007.

ENGLE, R. F., AND A. J. PATTON (2001): "What good is a volatility model?," *Quantitative Finance*, 1(2), 237–245.

GALLANT, A. R., C.-T. HSU, AND G. E. TAUCHEN (1999): "Using daily range data to calibrate volatility diffusions and extract the forward integrated variance," *Review of Economics & Statistics*, 81(4), 617–631.

GARMAN, M. B., AND M. J. KLASS (1983): "On the estimation of security volatilities from historical data," *Journal of Business*, 53(1), 67–78.

HANSEN, P. R., AND A. LUNDE (2001): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?," *Brown University Working Paper, 04*.

KUNITOMO, N. (1992): "Improving the Parkinson method of estimating security price volatilities," *Journal of Business*, 64(2), 295–302.

MALLIAVIN, P., AND M. E. MANCINO (2002): "Fourier series method for measurement of Multivariate Volatilities," *Finance and Stochastics*, 6(1), 49–61.

McLeish, D. L. (2002): "Highs and lows: Some properties of the extremes of a diffusion and applications to finance," *Canadian Journal of Statistics*, 30(2), 243–267.

Meddahi, N. (2002): "A Theoretical Comparison Between Integrated and Realized Volatility," *Journal of Applied Econometrics*, 17, 479–508.

Mincer, J., and V. Zarnowitz (1969): "The evaluation of economic forecasts and expections," in *Economic Forecasts and Expections*, ed. by J. Mincer. New York: National Bureau of Economic Research.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Retruns: A New Approach," *Econometrica*, 59, 347–370.

Pagan, A. R., and G. W. Schwert (1990): "Alternative models for conditional volatility," *Journal of Econometrics*, 45, 267–290.

Parkinson, M. (1983): "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, 53(1), 61–65.

Poon, S.-H., and C. Granger (202): "Forecasting volatility in financial markets: A review," *Forthcomming in Journal of Economic Literature*.

Rogers, L. C. G., and S. E. Satchell (1991): "Estimating variances from high, low, and closing prices," *Annals of Applied Probability*, 1(4), 504–512.

Tse, Y. (1998): "The conditional heteroskedasticity of the Yen-Dollar exchange rate," *Journal of Applied Econometrics*, 13(1), 49–55.

Wiggins, J. B. (1991): "Empirical tests of the baus and efficiency of the extreme-value estimator for common stocks," *Journal of Business*, 64(3), 417–432.

Zakoian, J.-M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.

Zivot, E., and J. Wang (2003): *Modeling Financial Times Series with S-Plus*. Springer-Verlag New York, Inc.

Appendix: Tables and Figures

Table 1: Mean Squared Errors Loss Functions for GARCH Models of IBM Stock Returns.

	Mean Squared Error			Mean Squared Error logs		
Model	MSE_{intra}	$MSE_{\text{+on.}}$	MSE_{inter}	$MSE^{\star}_{\text{intra}}$	$MSE^{\star}_{\text{+on.}}$	$MSE^{\star}_{\text{inter}}$
ARCH(1)	39.924	168.30	305.06	0.492	0.599	**11.882**
	(7.159)	(126.67)	(150.21)	(0.030)	(0.046)	(3.639)
GARCH(1,1)	30.722	159.77	297.01	0.298	0.479	12.441
	(5.820)	(120.33)	(142.98)	(0.019)	(0.033)	(3.718)
EGARCH(1,1)	25.434	155.25	289.37	**0.260**	**0.459**	12.506
	(5.389)	(120.01)	(142.64)	(0.016)	(0.032)	(3.733)
A-PARCH(1,1)	**23.358**	**153.66**	**286.52**	0.269	0.485	12.644
	(5.147)	(120.01)	(142.54)	(0.015)	(0.031)	(3.736)
THR-GARCH(1,1)	24.711	155.00	288.49	0.261	0.471	12.594
	(5.111)	(119.39)	(141.86)	(0.015)	(0.031)	(3.741)
FIGARCH(0,0)	31.460	161.89	299.03	0.338	0.543	12.573
	(6.376)	(123.15)	(146.01)	(0.020)	(0.036)	(3.724)
FIGARCH(1,1)	32.057	161.89	299.51	0.344	0.564	12.765
	(6.350)	(121.08)	(143.76)	(0.021)	(0.036)	(3.744)
FIAPARCH(1,1)	24.334	155.53	288.11	0.293	0.533	12.883
	(4.854)	(119.69)	(141.96)	(0.017)	(0.032)	(3.768)

In this table MSE_{intra}, $MSE_{\text{+on.}}$, MSE_{inter} is the MSE of the volatility forecasts less the estimated realized volatility, less the estimated realized volatility plus the squarred overnight return, and less the squarred log return. $MSE^{\star}_{\text{intra}}$, $MSE^{\star}_{\text{+on.}}$, and $MSE^{\star}_{\text{inter}}$ are the corresponding MSE of the log quantities. Standard errors are given in parentheses.

Table 2: R^2 from Mincer-Zarnowitz Regressions for GARCH Models of IBM Stock Returns.

	M.-Z. Regression			M.-Z. Log-regression		
Model	R^2_{intra}	$R^2_{\text{+on.}}$	R^2_{inter}	$R^{\star 2}_{\text{intra}}$	$R^{\star 2}_{\text{+on.}}$	$R^{\star 2}_{\text{inter}}$
ARCH(1)	0.067	0.010	0.003	0.131	0.102	0.009
GARCH(1,1)	19.907	4.328	1.808	43.301	35.048	1.531
EGARCH(1,1)	30.575	6.365	3.596	52.007	42.366	**2.625**
A-PARCH(1,1)	**36.052**	**7.329**	**4.525**	52.182	42.816	2.483
THR-GARCH(1,1)	32.425	6.665	3.871	**52.442**	42.700	2.622
FIGARCH(0,0)	16.867	3.112	1.191	37.827	30.773	1.614
FIGARCH(1,1)	16.338	3.420	1.226	38.245	31.471	1.314
FIAPARCH(1,1)	33.784	6.737	4.037	51.791	**42.931**	2.532

In this table R^2_{intra}, $R^2_{\text{+on.}}$, R^2_{inter} are the the coefficients of determination from the Mincer-Zarnowitz regressions of the volatility forecasts on the estimated realized volatility, the estimated realized volatility plus the squarred overnight return, and the squarred log return. $R^{\star 2}_{\text{intra}}$, $R^{\star 2}_{\text{+on.}}$, and $R^{\star 2}_{\text{inter}}$ are from the corresponding log regression.

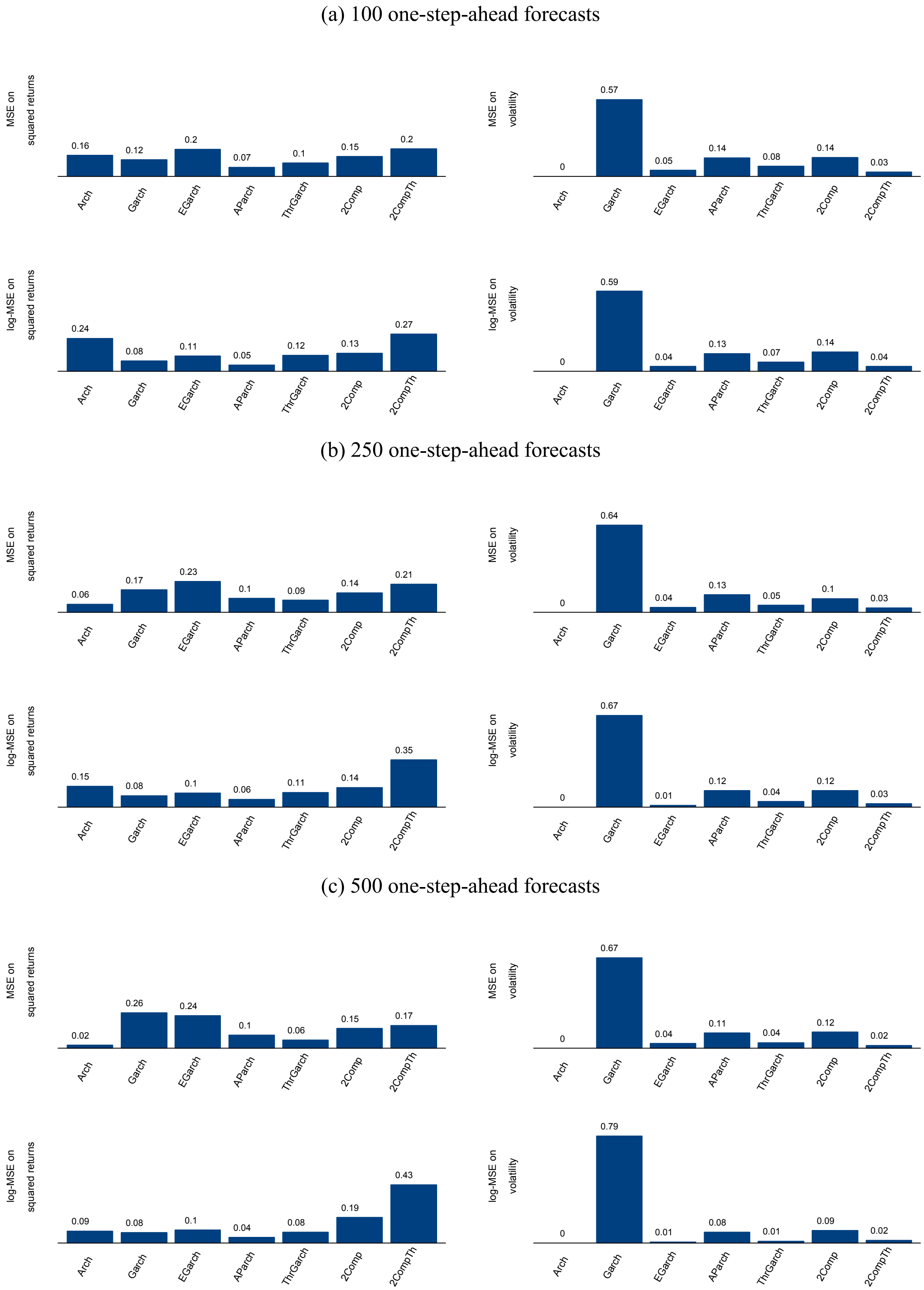
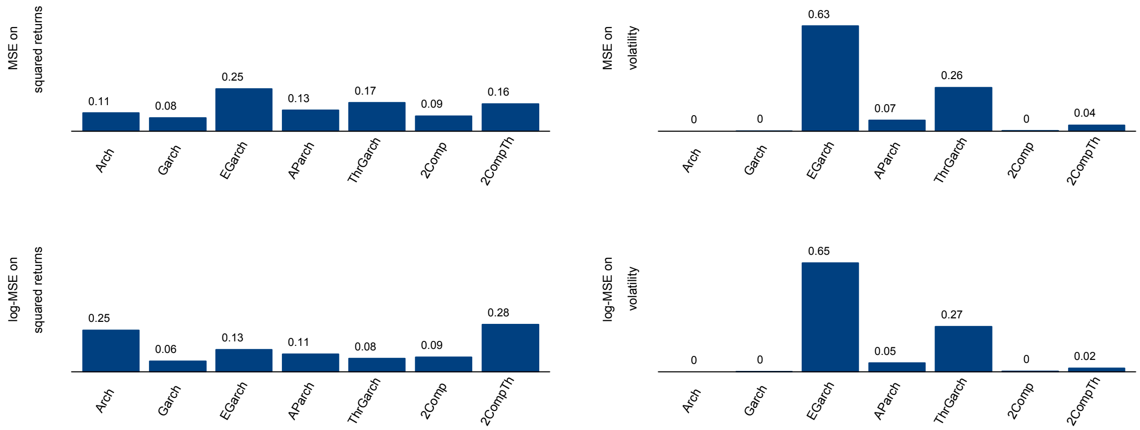
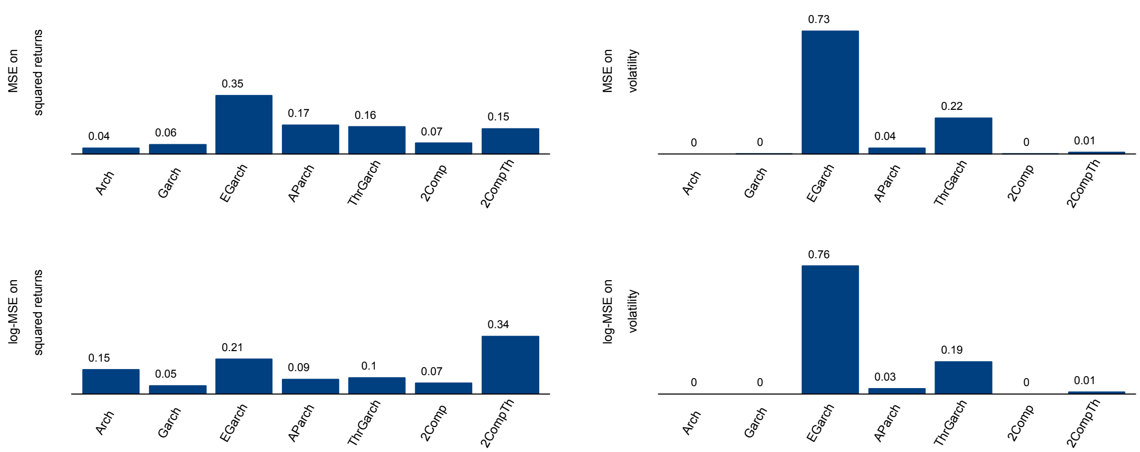


Figure 1: The figure shows the frequency that each model had the smallest out-of-sample loss. The artificial data was simulated from a GARCH(1,1) model. The three panels contain the results for 100, 250, and 500 out-of-sample observations. The first (second) column contains the evaluation based on squared returns (the true conditional variance); and the MSE (MSE*) loss function was applied in rows 1, 3, and 5 (2, 4, and 6).

(a) 100 one-step-ahead forecasts



(b) 250 one-step-ahead forecasts



(c) 500 one-step-ahead forecasts

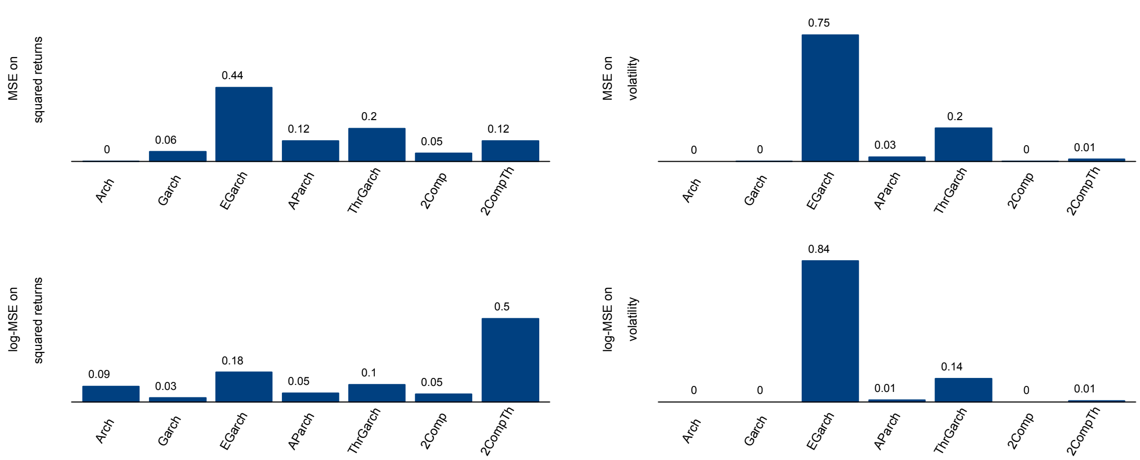


Figure 2: The figure shows the frequency that each model had the smallest out-of-sample loss. The artificial data was simulated from an EGARCH(1,1) model. The three panels contain the results for 100, 250, and 500 out-of-sample observations. The first (second) column contains the evaluation based on squared returns (the true conditional variance); and the MSE (MSE*) loss function was applied in rows 1, 3, and 5 (2, 4, and 6).