Forecasting Volatility with a GARCH(1,1) Model: Some New Analytical and Monte Carlo Results

Theologos Pantelidis
National University of Ireland, Maynooth

Nikitas Pittis
University of Piraeus

Preliminary Version

Abstract

This paper aims at explaining the poor forecasting performance of the GARCH(1,1) model reported in many empirical studies. A number of previous applied studies compare the conditional volatility forecasts of the GARCH(1,1) model to those of a model that assumes constant conditional variance (homoscedasticity) and conclude that the GARCH(1,1) model does not outperform the homoscedastic one. In these studies, the comparison of the relative forecasting accuracy of the two models is usually based on a selected formal statistical test that compares the Mean Squared Errors (MSEs) of the volatility forecasts of the two models. Some recent theoretical studies argue that the poor forecasting performance of the GARCH(1,1) model found in empirical studies is deceptive, since the utilization of the squared shocks as a proxy for the unobserved true conditional volatility increases the MSE of the volatility forecasts (MSE-inflation), rendering the MSE criterion invalid for proper evaluation of the forecasting accuracy of GARCH(1,1). However, the analytical results presented in this paper show that the utilization of the squared shocks as a proxy for the unobserved true conditional volatility "inflates" equally the MSEs of the forecasts of the GARCH(1,1) and the homoscedastic models and thus the MSE criterion remains capable to evaluate properly the relative forecasting performance of the two models. We also provide useful suggestions for proper statistical evaluation of the volatility forecasts of GARCH(1,1), together with an empirical illustration based on five bilateral exchange rates for the US dollar.

Keywords: GARCH, forecasting accuracy, volatility, MSE, nested models, exchange rates.

JEL Classification: C22, C53

1 Introduction

During the last two decades, emphasis has been given in forecasting the volatility of financial time series. This issue is crucial for policy makers, option traders and investors, since volatility forecasts are important for many financial decisions, such as implementation and evaluation of asset pricing theories and risk management (i.e. calculation of the Value-at-Risk). The ARCH class of models, pioneered by Engle (1982) and generalized by Bollerslev (1986), is by far the most popular class of econometric models for describing a series with time-varying conditional variance.

The out-of-sample forecasting accuracy of the GARCH models is questioned by a number of studies (see, e.g. West *et al.*, 1993; West and Cho, 1995; Franses and van Dijk, 2000). In most of these papers, the evaluation of the GARCH models is based on the Mean Squared Error (MSE) of the forecasts. However, Andersen and Bollerslev (1998) and Christodoulakis and Satchell (1998, 2003) have recently argued that the poor forecasting performance of the GARCH models in empirical studies is deceptive. More specifically, these studies reveal that the seemingly poor out-of-sample forecasting behavior of the GARCH models is due to the utilization of the squared shocks as a proxy for the unobserved true conditional volatility. This approximation results in an increase in the MSE of the forecasts (MSE-inflation), delivering a misleading evaluation of the GARCH models.

In this paper, we examine this issue more thoroughly by deriving analytical formulas of the MSE of the volatility forecasts of the popular GARCH(1,1) model when (i) the true conditional volatility is used and (ii) the squared shocks are used as a proxy for the true conditional volatility. Thus, we provide an analytical formula for the MSE-inflation in the case of a GARCH(1,1) model. Furthermore, we go one step forward and examine whether the poor forecasting performance of the GARCH(1,1) model reported in many empirical studies can be attributed to the utilization of a proxy for the unobserved conditional volatility. In most empirical studies, the evaluation of the forecasting accuracy of GARCH(1,1) is based on the comparison of the volatility forecasts of GARCH(1,1) to the volatility forecasts of a model that assumes homoscedasticity, which serves as a benchmark. The comparison of the two models (i.e. GARCH(1,1) and homoscedastic) is usually performed in the context of a formal statistical test for evaluating the relative forecasting performance of alternative econometric models. In order to reveal the effect of the utilization of the squared shocks as a proxy for the unobserved conditional volatility on the comparison of the relative forecasting performance of the GARCH(1,1) and the homoscedastic models, we derive analytical formulas for the MSE of the forecasts of an homoscedastic model.

We reach to two important conclusions with significant implications for empirical studies. First, the difference in the forecast accuracy between the GARCH(1,1) model and the homoscedastic model in terms of the MSE criterion is not affected by the utilization of the squared shocks as a proxy for the conditional variance. This means that the MSE criterion remains capable to evaluate properly the relative forecast performance of the two models. Second, the difference between the two models is maximized for one-step ahead forecasts, while the two models become equivalent as the forecast horizon tends to infinity. Our second finding suggests that applied researchers should base the statistical testing of the relative forecast accuracy of the two models on sequences of one-step ahead forecasts, rather than sequences of $s-$step ahead forecasts ($s > 1$), in order to increase the power of the tests.

Moreover, by means of Monte Carlo simulations, we calculate the empirical size and the size-adjusted power of a variety of statistical procedures to select the GARCH(1,1) model over the misspecified homoscedastic one when the MSE is used for the evaluation of the volatility forecasts. In the literature, there are numerous statistical procedures properly designed to compare either non-nested models, such as the Diebold-Mariano statistic (Diebold and Mariano, 1995), or nested models, such as the test statistics proposed by McCracken (2004). In our case, we are interested in comparing two nested models. However, we consider tests designed for comparing both nested and non-nested models for comparison reasons.[1]

The simulation results indicate that the majority of the test statistics under scrutiny suffer from significant size distortions, with the exception of the two out of the three statistics proposed by McCracken (2004). For these two statistics only minor finite sample size distortions are observed. As far as the power of the statistics is concerned, the simulation results reveal two important issues. First, the power of the tests is only marginally affected by the sample size used in the estimation procedure as long as the number of observations is sufficient to ensure reliable estimates of the GARCH parameters. On the other hand, the power of the tests heavily depends on the number of out-of-sample forecasts. More specifically, the power of the tests increases substantially with the number of the out-of-sample forecasts. This suggests that applied researchers should use as many observations as possible for the out-of-sample

[1] Christodoulakis and Satchell (2003) show that the utilization of the squared shocks as a proxy for the true but unobserved conditional volatility invalidates the Diebold-Mariano statistic, even if we are interested in comparing two non-nested GARCH models.

forecast exercise. Second, the power of the tests increases with the persistence of the series under scrutiny. This is a favorable result for practitioners, since conditional volatilities are usually highly persistent in empirical studies. In summary, the simulation results show that McCracken's statistics are in general able to indicate whether the CARCH(1,1) model provides more accurate volatility forecasts than the homoscedastic model.

Finally, we illustrate the usefulness of our results with an application to five bilateral exchange rates for the US dollar. We consider the data used by West and Cho (1995). Contrary to the findings of West and Cho (1995), our results indicate the superiority of the GARCH(1,1) model over the homoscedastic model for forecasting the conditional variance in three out of the five exchange rates considered.

The rest of the paper is organized as follows. In Section 2, we derive the MSEs of the volatility forecasts for the GARCH(1,1) model and the homoscedastic model. Section 3 describes briefly selected statistical procedures for comparing the forecast accuracy of two models and presents the results of a Monte Carlo simulation that examines the empirical size and the size-adjusted power of these tests to choose the GARCH(1,1) model over the homoscedastic one. Section 4 contains an illustrative example and Section 5 concludes the paper.

2 Evaluation of the Forecasting Performance of the GARCH(1,1) Model

We examine the following GARCH(1,1) model:

$$\begin{aligned} e_t &= h_t z_t \\ h_t^2 &= w + a e_{t-1}^2 + b h_{t-1}^2 \end{aligned} \tag{1}$$

where w, a and b are positive and $z_t \sim iid(0,1)$. The unit variance assumption for the innovation process, z_t, ensures that h_t^2 is the variance of e_t, conditional on the information set, F_{t-1}, that contains the past history of the process up to period $t-1$. We do not assume a specific distribution for the innovations. However, similarly to Baillie and Bollerslev (1992) (BB henceforth) and for simplicity, we assume that the conditional distribution of e_t given F_{t-1} is symmetric with all the existing even-ordered moments[2] proportional to the corresponding powers of the

[2]We are particularly interested in the second and fourth conditional moments.

conditional variance, that is

$$E_{t-1}(e_t^{2r+1}) = 0, \qquad r = 0, 1, ..., K-1$$
$$E_{t-1}(e_t^{2r}) = k_r(h_t^{2r}), \quad r = 0, 1, ..., K$$

where k_r is the $rth-$order cumulant for the conditional density of e_t.[3]

The optimal predictor, $\widehat{h}_{t+s}^2$, of the conditional variance for forecast horizon s is the conditional expected value of h_{t+s}^2, that is $E_t(h_{t+s}^2)$.[4] It is easy to show that:

$$\widehat{h}_{t+s}^2 = E_t(h_{t+s}^2) = w\sum_{i=1}^{s-1}(a+b)^{i-1} + (a+b)^{s-1}h_{t+1}^2 \tag{2}$$

where $h_{t+1}^2 = w + ae_t^2 + bh_t^2$ is known at time t. The necessary and sufficient condition for the existence of the unconditional variance $\sigma^2 := Var(e_t) = \frac{w}{1-a-b}$ of e_t is $a+b<1$. In that case, (2) can be written as follows:

$$\widehat{h}_{t+s}^2 = \sigma^2 + (a+b)^{s-1}(h_{t+1}^2 - \sigma^2)$$

It is obvious that as $s \to \infty$, the predictor tends to the unconditional variance σ^2. On the other hand, if $a+b=1$, we obtain the following predictor:

$$\widehat{h}_{t+s}^2 = w(s-1) + h_{t+1}^2$$

2.1 Non-Feasible and Feasible MSEs of the Forecasts of the GARCH(1,1) Model

Initially, we calculate the non-feasible conditional MSE, $MSE_{G,t}^{NF}(s) := E_t[(h_{t+s}^2 - \widehat{h}_{t+s}^2)^2]$, of the GARCH(1,1) model for $s-$periods ahead based on the true conditional variance h_{t+s}^2.[5] BB provide the following formula for the calculation of $MSE_{G,t}^{NF}$ of the volatility forecasts based on $\widehat{h}_{t+s}^2$:

$$MSE_{G,t}^{NF}(s) = (k_2-1)a^2\sum_{i=1}^{s-1}(a+b)^{2(i-1)}E_t(h_{t+s-i}^4)$$

[3]By definition $k_0 = k_1 = 1$. Under conditional normality, $k_r = \prod_{i=1}^{r}(2i-1)$, $r = 1, 2, ...$, while under conditional t_n-distribution, $k_r = (n-2)^r\Gamma(r+\frac{1}{2})\Gamma(\frac{n}{2}-r)\Gamma^{-1}(\frac{1}{2})\Gamma^{-1}(\frac{n}{2})$, $r = 1, 2, .., K$ where $\Gamma(\cdot)$ denotes the gamma function and $K = int(\frac{n}{2})$.

[4]By definition $E_t(h_{t+s}^2) = E_t(e_{t+s}^2)$ for $s > 0$.

[5]In practice, the calculation of this MSE is non-feasible, since h_{t+s}^2 is unobservable.

where $E_t(h^4_{t+s}) = w^2 + \sum_{i=1}^{2} \begin{pmatrix} 2 \\ i \end{pmatrix} \pi_i w^{2-i} E_t(h^{2i}_{t+s-1})$, $\pi_1 = (a+b)$ and $\pi_2 = k_2 a^2 + ab + b^2$. Upon repeated substitution, we obtain that

$$
\begin{aligned}
MSE^{NF}_{G,t}(s) \;&=\; (k_2-1)a^2 \sum_{i=1}^{s-1} (a+b)^{2(i-1)} \{ \pi_2^{s-i-1} h^4_{t+1} + w^2 \sum_{q=1}^{s-i-1} \pi_2^{q-1} + \\
&\quad + 2\pi_1 w \sum_{q=1}^{s-i-1} \{ \pi_2^{q-1} w \sum_{p=1}^{s-i-q-1} [(a+b)^{p-1} + (a+b)^{s-i-q-1} h^2_{t+1}]\}\}
\end{aligned}
$$

In order to obtain the non-feasible unconditional MSE, $MSE^{NF}_G := E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2]$, we need to assume that the fourth moment of e_t is finite. Among others, Hafner (2003) proves that the necessary and sufficient condition for the finiteness of $E(e^4_t)$ is that $3ca^2 + 2ab + b^2 < 1$ where $c := \frac{E(z^4_t)}{3}$. In that case:

$$E(e^4_t) = 3cw\sigma^2 \frac{1+a+b}{1-3ca^2-2ab-b^2}$$

We can now derive the formula for the unconditional MSE^{NF}_G as follows: $MSE^{NF}_G(s) := E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2] = E\{E_t[(h^2_{t+s} - \widehat{h}^2_{t+s})^2]\} = E\{(k_2-1)a^2 \sum_{i=1}^{s-1} (a+b)^{2(i-1)} E_t(h^4_{t+s-i})\}$. After some algebra,[6] we obtain the following formula:

$$MSE^{NF}_G(s) = \frac{k_2-1}{k_2} a^2 [3cw\sigma^2 \frac{1+a+b}{1-3ca^2-2ab-b^2}] \frac{1-(a+b)^{2(s-1)}}{1-(a+b)^2}$$

So far, we have calculated the MSE of the forecasts of the conditional variance assuming that h^2_{t+s} is known. However, we cannot observe h^2_{t+s}. Therefore, in empirical studies we evaluate the forecasting accuracy of GARCH models by using the squared shocks e^2_{t+s} as a proxy for h^2_{t+s}. We will now show that the use of this proxy results in higher $MSEs$ of the volatility forecasts, since e^2_{t+s} is a noisy estimator of the actual conditional volatility. For example, Lopez (2001) states that "...even if one is willing to accept a proxy that is up to 50% different from h^2_{t+s}, e^2_{t+s} would fulfill this condition only 25% of the time". To be more specific, he shows that $\Pr(e^2_{t+s} \in [\frac{1}{2}h^2_{t+s}, \frac{3}{2}h^2_{t+s}]) = 0.2588$.[7]

First of all, we calculate the feasible conditional MSE, $MSE^F_{G,t}(s) := E_t[(e^2_{t+s} - \widehat{h}^2_{t+s})^2]$. We

[6] We use the following equality: $E_t(e^4_{t+s}) = k_2 E_t(h^4_{t+s})$ for $s > 0$.

[7] Lopez (2001) assumes normal innovations, that is $z_t \sim N(0,1)$.

end up with the following formula:

$$\begin{aligned} MSE^F_{G,t}(s) &= k_2\{\pi_2^{s-1}h^4_{t+1} + w^2\sum_{i=1}^{s-1}\pi_2^{i-1} + \\ &\quad +2\pi_1 w\sum_{i=1}^{s-1}\pi_2^{i-1}[\sigma^2 + (a+b)^{s-i-1}(h^2_{t+1} - \sigma^2)]\} - \\ &\quad \{[\sigma^2 + (a+b)^{s-1}(h^2_{t+1} - \sigma^2)]^2\} \end{aligned}$$

The feasible unconditional MSE, defined as $MSE^F_G(s) := E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2] = E\{E_t[(e^2_{t+s} - \widehat{h}^2_{t+s})^2]\}$, is as follows:[8]

$$MSE^F_G(s) = \sigma^4[(a+b)^{2(s-1)} - 1] + 3cw\sigma^2\frac{1+a+b}{1-3ca^2-2ab-b^2}[1 - \frac{(a+b)^{2(s-1)}}{k_2}]$$

It is easy to see that the MSE-inflation of the volatility forecasts of CARCH(1,1), $MSEI_G := MSE^F_G(s) - MSE^{NF}_G(s)$, which corresponds to the increase in the MSE due to the use of e^2_{t+s} instead of h^2_{t+s}, is given by:

$$MSEI_G = 3cw\sigma^2\frac{1+a+b}{1-3ca^2-2ab-b^2}(1 - \frac{1}{k_2})$$

Remark (1): $MSEI_G$ is independent of s and an increasing function of the GARCH parameters (w, a and b). Finally, $MSEI_G$ increases with k_2, suggesting that the effect of approximating h^2_{t+s} with e^2_{t+s} is more pronounced when z_t follows a leptokurtic distribution. This is illustrated in Figure 1, which shows $MSEI_G$ for the Gaussian and t(5)-distributions respectively.

2.2 Non-Feasible and Feasible MSEs of the Forecasts of the Homoscedastic Model

In the previous section, we showed that squared residuals e^2_{t+s} are noisy estimators of the unobserved conditional variance h^2_{t+s}. Therefore, the evaluation of the forecast accuracy of GARCH models based on the feasible unconditional MSE, $MSE^F_G(s)$, can be misleading. In this section, we move one step forward and examine whether the poor volatility forecasting performance of GARCH models reported in empirical studies is the result of the utilization of a noisy proxy for the unobserved actual volatility. The utilization of a GARCH model in empirical studies would be justified if this model could at least produce more accurate forecasts

[8]Once again, we assume finiteness of the fourth moment of e_t. That is, we assume that $E(e^4_t) < \infty$, $\forall t$.

than a model that assumes constant conditional volatility (homoscedasticity). For this reason, many empirical studies compare the volatility forecasting accuracy of GARCH(1,1) to that of an homoscedastic model that serves as a benchmark. In order to reveal the effect of the utilization of the squared shocks as a proxy for the unobserved conditional volatility on the comparison of the relative forecasting performance of the GARCH(1,1) and the homoscedastic models, we need to derive analytical formulas for the MSE of the forecasts of an homoscedastic model.

Assume that a researcher ignores the existence of time heterogeneity in the conditional variance and she chooses to estimate a model with constant variance. That is, she erroneously assumes that $h_t^2 = \sigma^2$, $\forall t$ and she utilizes σ^2 instead of $\widehat{h}_{t+s}^2$ to forecast the volatility. We will calculate the feasible and non-feasible $MSEs$ of the volatility forecasts in this case. First, we derive the non-feasible MSE, $MSE_H^{NF}(s) := E[(h_{t+s}^2 - \sigma^2)^2]$, which is based on the unobserved conditional variance h_{t+s}^2. We end up with the following formula:

$$MSE_H^{NF}(s) = \sigma^2[\frac{3cw}{k_2}\frac{1+a+b}{1-3ca^2-2ab-b^2} - \sigma^2]$$

Similarly, we derive the feasible MSE, $MSE_H^F(s) := E[(e_{t+s}^2 - \sigma^2)^2]$:

$$MSE_H^F(s) = \sigma^2[3cw\frac{1+a+b}{1-3ca^2-2ab-b^2} - \sigma^2]$$

Remark (2): Interestingly, the employment of e_{t+s}^2 as a proxy for h_{t+s}^2 in the definition of MSE produces a MSE-inflation for the homoscedastic model, denoted as $MSEI_H := MSE_H^F(s) - MSE_H^{NF}(s)$, that is equal to the $MSEI_G$ produced by the GARCH predictor $\widehat{h}_{t+s}^2$. That is,

$$MSEI_H = 3cw\sigma^2\frac{1+a+b}{1-3ca^2-2ab-b^2}(1-\frac{1}{k_2}) = MSEI_G$$

2.3 The Relative Forecasting Performance of the GARCH(1,1) Model and the Homoscedastic Model

We now compare the volatility forecasts of the GARCH(1,1) model to those of the naive homoscedastic model. The following proposition is rather trivial stating that the forecasts of the conditional variance based on the GARCH model are more accurate (in terms of the MSE criterion) than those based on an homoscedastic model.

Proposition 1 *Let* $R1(s) := \frac{E[(h_{t+s}^2 - \widehat{h}_{t+s}^2)^2]}{E[(h_{t+s}^2-\sigma^2)^2]}$. *Then,* $R1 < 1$ *for every finite* s.

Proof. See Appendix. ■

The following proposition is far less trivial, stating that the $MSE-$superiority of the GARCH over the homoscedastic forecasts is maintained even if the $MSEs$ are defined with respect to e_{t+s}^2 instead of the unobserved h_{t+s}^2.

Proposition 2 *Let* $R2(s) := \frac{E[(e_{t+s}^2 - \widehat{h}_{t+s}^2)^2]}{E[(e_{t+s}^2 - \sigma^2)^2]}$. *Then,* $R2 < 1$ *for every finite* s.

Proof. See Appendix. ■

The following proposition is very crucial for the comparison of the two models in empirical studies. It states that the difference between the forecasting performance of the GARCH(1,1) model and the homoscedastic model when the true conditional volatility h_{t+s}^2 is used for the calculation of the MSEs equals the difference between the forecasting performance of the two models when e_{t+s}^2 is used as a proxy for h_{t+s}^2 for the calculation of the MSEs.

Proposition 3 *Let* $D^F(s)$ *be the difference between the feasible MSE of the naive forecast and the feasible MSE of the GARCH forecast for horizon* s, *i.e.* $D^F(s) := MSE_H^F(s) - MSE_G^F(s)$. *Let* $D^{NF}(s)$ *be the difference between the non-feasible MSE of the naive forecast and the non-feasible MSE of the GARCH forecast for horizon* s, *i.e.* $D^{NF}(s) := MSE_H^{NF}(s) - MSE_G^{NF}(s)$. *Then* $D^F(s) = D^{NF}(s)$, $\forall s$.

Proposition (3) states a very interesting result, since it shows that no information is lost concerning the relative forecasting performance of the two models when e_{t+s}^2 is used as a proxy for h_{t+s}^2 for the calculation of the MSEs of the volatility forecasts. This result suggests that the MSE can be used in empirical studies to compare the relative forecast performance of such models.

Remark (3): It is easy to show that the difference, D, between the MSE of the naive (homoscedastic) forecasts and the MSE of the GARCH-based forecasts is as follows:

$$D := (a+b)^{2(s-1)}[\frac{3cw\sigma^2}{k_2}\frac{1+a+b}{1-3ca^2-2ab-b^2} - \sigma^4] = (a+b)^{2(s-1)} MSE_H^{NF}(s)$$

This is the difference for both the feasible and the non-feasible case. Obviously, D is a decreasing function of s and it is maximized when $s=1$. Figure 2 illustrates D for $s=1$ and $s=5$. D is also an increasing function of a, b and w. Finally, D is an increasing function of k_2.

The previous remark has important implications for empirical studies. It shows that the difference in the MSEs of the two models is maximized for one-step ahead forecasts. This suggests that when comparing the forecasting performance of alternative econometric models based on a formal statistical test, the researcher should use sequences of one-step ahead forecasts, rather than sequences of $s-$step ahead forecasts ($s > 1$), in order to increase the power of the tests to indicate the optimal model for forecasting volatility. In the following section, we will investigate the power of a variety of statistical procedures to select the GARCH(1,1) model over the misspecified homoscedastic model by means of Monte Carlo simulations.

3 The Comparison of the Forecasting Performance of the two Models

Before presenting the results of the simulation, we first describe briefly the statistical tests used in the simulation to compare the forecasting accuracy of the two alternative models. In the literature, there is a variety of statistical procedures to compare the relative forecasting performance of econometric models. These procedures are separated into two distinctive groups. The first group contains the statistical tests that are properly designed to compare non-nested models, while the second group contains all the tests designed to compare nested models.

The standard procedure used by practitioners to compare the relative forecasting performance of two models is as follows. Assuming that a sample of $T = R + P$ observations of a variable Y_t is available, the first R observations are used to estimate the parameters of the models, while the last P observations are used for the out-of-sample forecast exercise. There are three alternative ways to generate the sequence of the forecasts, namely the recursive, rolling and fixed schemes described in more details in Section 4. Given the two sequences of P forecasts (one for each model), namely $f_{i,t}$, $i = 1, 2$ and $t = 1, 2, ..., P$, the forecast errors are calculated ($\epsilon_{i,t} = Y_{t+t} - f_{i,t}$, $i = 1, 2$ and $t = 1, 2, ..., P$). Finally, the predictive ability of the two models is compared based on the researcher's preferred loss function $L(\cdot)$ and statistical test. In the case of the MSE, we have that $L(\epsilon_{i,t}) := \epsilon^2_{i,t}$, $i = 1, 2$ and $t = 1, 2, ..., P$. We now describe briefly some of the statistical procedures available in the literature for comparing the forecasting accuracy of two alternative models.

3.1 Tests for Comparing Forecast Accuracy

First of all, we present the popular Diebold-Mariano statistic, introduced by Diebold and Mariano (1995), which was designed to compare the forecasting performance of two non-nested models. Under the null hypothesis, the test assumes equal forecast accuracy, that is H_0 : $E[L(\epsilon_{1,t})] = E[L(\epsilon_{2,t})]$ or equivalently $H_0 : E[d_t] = 0$ where $d_t = L(\epsilon_{1,t}) - L(\epsilon_{2,t})$. Given the sequence of loss differential $\{d_t\}_{t=1}^{P}$, Diebold and Mariano (1995) show that $\sqrt{P}(\overline{d}-\mu) \xrightarrow{d} N(0,\Omega)$ where $\overline{d} := P^{-1}\sum_{t=1}^{P}[L(\epsilon_{1,t}) - L(\epsilon_{2,t})]$. The test statistic they propose is the following:

$$DM - statistic = S_1 := \left(\sqrt{P}\right)^{-1}\left(\sqrt{\widehat{\Omega}}\right)^{-1}\sum_{t=1}^{P}[L(\epsilon_{1,t}) - L(\epsilon_{2,t})] \tag{3}$$

where $\widehat{\Omega}$ is a consistent estimator of the asymptotic variance Ω. Under the null, S_1 follows a $N(0,1)$ distribution under specific assumptions.

A key assumption for deriving the asymptotic distribution of S_1 is that the two models under examination are non-nested. In the case of two nested models, $\Omega = 0$ invalidating the procedure proposed by Diebold and Mariano (1995), since the limiting distribution of S_1 is non-standard. However, McCracken (2004) derives numerically estimates of the asymptotic critical values for S_1. He also suggests two alternative statistical tests to compare the forecast accuracy of two nested models. The first one is based on a method introduced by Granger and Newbold (1977) and used by Ashley *et al.* (1980). This test is based on the t-statistic of the parameter γ of the following regression: $(\epsilon_{1,t} - \epsilon_{2,t}) = \gamma(\epsilon_{1,t} + \epsilon_{2,t}) + error\ term$.[9] The proposed statistic is as follows:

$$OOS - t := \sqrt{P-1}\frac{P^{-1}\sum_{t=1}^{P}[\epsilon_{1,t}^2 - \epsilon_{2,t}^2]}{\sqrt{([P^{-1}\sum_{t=1}^{P}[\epsilon_{1,t} + \epsilon_{2,t}]^2][P^{-1}\sum_{t=1}^{P}[\epsilon_{1,t} - \epsilon_{2,t}]^2] - \overline{d}^2)}} \tag{4}$$

The second statistic that McCracken (2004) proposes for the comparison of the forecast accuracy of two nested models based on the MSE criterion is as follows:

$$OOS - F := \frac{\sum_{t=1}^{P}[\epsilon_{1,t}^2 - \epsilon_{2,t}^2]}{P^{-1}\sum_{t=1}^{P}\epsilon_{2,t}^2} \tag{5}$$

McCracken (2004) shows that the limiting distributions of S_1, $OOS - t$ and $OOS - F$ are non-standard[10] and he provides numerical estimates of the asymptotic critical values for valid inference. It is important to note that in the case of nested model, the tests are one-sided. More specifically, in the case of nested models the null hypothesis is $H_0 : E[L(\epsilon_{1,t})] = E[L(\epsilon_{2,t})]$, while

[9]It is important to note that henceforth when we consider the case of nested models, $i = 1$ stands for the restricted model, while $i = 2$ corresponds to the unrestricted model.

[10]However, McCracken (2004) shows that the three test statistics are asymptotically pivotal.

the alternative hypothesis is $H_A : E[L(\epsilon_{1,t})] > E[L(\epsilon_{2,t})]$.

4 Monte Carlo Simulations

In Section 2, we derive analytical formulas for the feasible and non-feasible $MSEs$ of the volatility forecasts for both the GARCH(1,1) model and the homoscedastic one. We proved that the utilization of the squared shocks e_{t+s}^2 as a proxy for the unobserved conditional variance inflates equally the MSE of both models. Thus, the MSE criterion retains its ability to evaluate properly the relative forecast accuracy of the two models. However, it is interesting to examine whether applied researchers are able to distinguish between the two models based on standard statistical procedures such as those presented in the previous section. In other words, we want to investigate the power of the statistical tests to select the GARCH(1,1) model over the homoscedastic model when the evaluation is based on the MSE criterion. We first investigate the size properties of the test statistics to bring to light any size distortions that we should take into consideration when examining the power of these statistics.

In our first experiment, we generate random realizations of the error term e_t assuming constant variance, that is a and b are set equal to zero in (1). Two different cases are considered for the distribution of the innovation term z_t. That is, we allow the innovations to follow either a normal distribution or a t-distribution with 5 degrees of freedom. The estimation of the parameters of the model is based on the Quasi-Maximum Likelihood method, which is robust to non-Gaussian innovations. We use the Berndt-Hall-Hall-Hausman (BHHH) algorithm for the maximization of the likelihood function.

We generate $T = R+P+200$ observations. To account for the effect of the initial conditions, the first 200 observations are discarded. The estimation of the models is based on R observations $\{e_t\}_{t=201}^{200+R}$, while the last P observations are used for the out-of-sample exercise. We consider the case of generating one-step ahead forecasts. As we have already mentioned previously, there are three alternative methods to generate a sequence of P one-step ahead forecasts, namely the recursive, rolling and fixed schemes.

The recursive scheme generates the first forecast based on a model estimated by using observations $\{e_t\}_{t=201}^{200+R}$. Afterwards, the model is re-estimated based on the sample $\{e_t\}_{t=201}^{200+R+1}$ and the second one-step ahead forecast is generated. This procedure is repeated until the last one-step ahead forecast is produced based on a model estimated by using observations

$\{e_t\}_{t=201}^{200+R+P-1}$. Thus, in each step the estimation sample increases by one observation.

Under the rolling scheme, the first forecast is generated with the same way as in the case of the recursive scheme. However, the second forecast is based on a model estimated by using observations $\{e_t\}_{t=201+1}^{200+R+1}$ and so on until the last forecast is generated based on a model estimated by using $\{e_t\}_{t=201+P-1}^{200+R+P-1}$. In other words, the estimation is based on a fixed window of R observations. For each observation added at the end of the estimation sample, an observation in the distant past is discarded.

Finally, the fixed scheme is based on a single estimation of the model based on observations $\{e_t\}_{t=201}^{200+R}$. This method is only used when the computation burden is large.

Among the three alternative schemes, the recursive scheme seems to be the most efficient one since it takes advantage of all the available information (i.e. observations) to generate the forecasts. In our study we consider all three alternative schemes but for brevity we present the results for the recursive scheme only. We mention in the text any cases where the rolling or fixed schemes generate different results. We examine four different estimation sample sizes ($R = 300$, 600, 900 and 1200). For each value of R, we consider three different values for the number of the out-of-sample forecasts P, that is $P = 0.1 * R$, $P = 0.2 * R$ and $P = 0.4 * R$. We examine a variety of statistical procedures to test the relative forecast accuracy of the GARCH(1,1) model and the homoscedastic model. In our case, we compare two nested models. Therefore, the three statistics given in (3), (4) and (5) are suitable for comparing the forecast accuracy of the two models. However, we also consider a number of alternative tests for comparison reasons. More specifically, we consider the Diebold-Mariano statistic based on critical values from the $N(0, 1)$ distribution (we call this statistic $S_1 - DM$), the sign test (a finite-sample version, S_2, and an asymptotic version, S_2^*), the Wilcoxon's signed-rank test (S_3), the Morgan-Granger-Newbold test (MGN), the Meese-Rogoff test (MR) and the Van der Waerden test (VDW). More details about these statistics can be found in Lehmann (1975), Diebold and Mariano (1995) and Sheskin (2000).

The results concerning the empirical size of the test statistics for the Normal and the t(5) distributions are given in Tables 1 and 2 respectively. A nominal size of 5 percent is assumed. The main findings can be summarized as follows:

1. In the case of Gaussian innovations, $OOS - F$ and S_1 are slightly oversized remaining very close to the nominal size of 5 percent in all cases. These can be considered as minor finite sample distortions, since both statistics are based on asymptotic critical values. In all cases considered,

the empirical size of $OOS - F$ and S_1 does not go over 8 percent. On the other hand, $OOS - t$ is heavily undersized in most cases. However, the empirical size of the statistic approaches the nominal size as the number of out-of-sample forecasts P increases. In the case of leptokurtic innovations (t(5)-distribution), the behavior of the three statistics deteriorates slightly but they all become well-sized as P increases.

2. $S_1 - DM$ and MGN are undersized when normal innovations are assumed, while they both improve their size properties in the case of leptokurtic innovations. Surprisingly, the behavior of the two statistics deteriorates as P increases, becoming more conservative. On the other hand, VDW is slightly oversized in the Gaussian case. However, leptokurtic innovations seem to have a negative effect on the size properties of the statistic. Finally, all the other statistics considered are heavily oversized, especially in the case of leptokurtic innovations. For example, in the case of the t(5)-distribution and for $R = 600$ and $\pi := \frac{P}{R} = 0.1$, the empirical size of S_2 is 42 percent. In some extreme cases, the empirical size of some statistics is higher than 60 percent.

In summary, the simulation reveals that among all the tests considered, $OOS - F$ and S_1 are the statistics with the best size properties. Furthermore, the good behavior of $OOS - F$ and S_1 is not affected by leptokurtic innovations. On the other hand, the size properties of $OOS - t$ crucially depend on the number of out-of-sample forecasts, P, used in the evaluation procedure. More specifically, $OOS - t$ is undersized, unless P is large enough ($P > 450$) to ensure proper size properties of the statistic. Meanwhile, the size distortions observed in the case of $S_1 - DM$ and MGN are not negligible. Finally, the rest of the statistics suffer from severe size distortions and therefore they are excluded from the simulations that follow.

We now examine the power of the test statistics to select the GARCH(1,1) model over the homoscedastic model when heteroscedasticity is present. The evaluation of the tests is based on the MSE criterion. Given the size distortions revealed by our first Monte Carlo experiment, the comparison of the statistics is based on their size-adjusted power. To be more specific, in the following simulation the rejection or not of the null hypothesis of equal forecast accuracy of the two models is based on the comparison between the estimated value of each statistic and the 95th percentile of the empirical distribution of the statistic, as calculated in our first simulation. That is, standard critical values of the statistics are replaced by the empirical critical values from the corresponding size simulation.

Our second simulation aiming at examining the power properties of the test statistics is organized as follows. We assume that the data are generated by the GARCH(1,1) model given in (1). Once again, the innovation term z_t is assumed to follow either a Normal distribution or a t-distribution with 5 degrees of freedom.

It is well-known that the moment characteristics of e_t are determined by the value of the parameters a and b in (1). We consider the following three cases:

Case I: Fourth-order stationarity

(C1): $3ca^2 + 2ab + b^2 < 1$ where $c := \frac{E(z_t^4)}{3}$. Among others, Hafner (2003) proves that condition (C1) is necessary and sufficient for the finiteness of $E(e_t^4)$. Davidson (2001) demonstrates that condition (C1) is necessary and nearly sufficient for e_t to be near-epoch dependent on z_{1t} in L_2-norm ($L_2 - NED$).[11]

Case II: Second-order stationarity

(C2): $a + b < 1$. Condition (C2) is necessary and sufficient for the existence of the unconditional variance $\sigma^2 := Var(e_t)$ of e_t. If condition (C2) holds, while condition (C1) fails, e_t possesses unconditional second (but not fourth) moments. Davidson (2001) proves that under the second-order stationarity conditions, e_t is $L_1 - NED$ on z_t. More recently, Carrasco and Chen (2002) prove that the same condition is necessary and sufficient for e_t to be $b-$mixing.

Case III: First-order stationarity (IGARCH)

(C3) $a + b = 1$, In this case e_t is not a covariance stationary process.[12]

In the simulation, we consider a set a various combinations for the values of a and b for all the cases reported above in order to examine the effect of the moment characteristics of e_t on the power of the statistical tests under examination. More specifically, we provide the results for nine different Data Generating Processes (DGPs) reported in the first two columns of Tables 3-10. Under normality, the processes considered in the first, fourth and seventh DGP possess finite fourth moments, while the second, fifth and eighth DGP belong to Case II. Finally, the third, sixth and ninth DGP correspond to IGARCH processes. However, we should note that the formulas for the MSE of the volatility forecasts derived in Section 2 are valid only when fourth-order stationarity holds (Case I). Otherwise, the MSEs of the volatility forecasts become infinite.

[11]Davidson (2001) shows that the fourth moment condition is necessary and nearly sufficient for the L_2-NED property, irrespectively of the distribution of the innovations process. The L_2-NED property of a GARCH(1,1) process was first proved by Hansen (1991) under the additional assumption of normality of z_t.

[12]Nelson (1990) demonstrates that an IGARCH or even a mildly explosive process may still be strictly stationary and ergodic.

Throughout the simulation we keep the constant term w equal to 0.1. All the other settings of the simulation are identical to those of our first experiment. That is, we generate $T = P+R+200$ observations and the first 200 observations are discarded. The initial estimation of the models is based on R observations, while the last P observations are used for the out-of-sample exercise, which is based on one-step ahead forecasts in order to increase the power of the tests (see Remark 3). Finally, for each set of values for a and b, we examine all the different combinations for the values of R and P considered in the previous experiment. The results of the simulation for the two alternative distributions of z_t are presented below.

Case A: Normal innovations

As we have already mentioned, we only consider the following five statistical procedures: S_1, $OOS-t$, $OOS-F$, MGN and S_1-DM. However, there is no need to distinguish between S_1 and S_1-DM, since the size-adjusted power of the two statistics is calculated based on the same empirical critical values. Thus, the results are identical for the two statistics. The results for the four tests, displayed in Tables 3-6, can be summarized as follows.[13]

1. The $OOS-F$ statistic is the more powerful procedure, followed by MGN and S_1.[14] The size-adjusted power of $OOS-F$ reaches 100 percent when the estimation sample size, R, and the number of out-of-sample forecasts, P, are high. The size-adjusted power of $OOS-t$ is very low, unless P is greater than 450. For example, in the case of $DGP-3$ and for $R = 1200$, the size-adjusted power of $OOS-t$ is 25 percent and 90 percent for $\pi = 0.1$ and 0.4 respectively (see Table 5).

2. The size-adjusted power of all the tests considered increases with the persistence of the process under examination. In other words, the size-adjusted power of the tests is maximized in the case of an IGARCH process. For example, when $R = 1200$ and $\pi = 0.1$, the size-adjusted power of the S_1 test increases from 38 to 72 percent for $DGP-1$ and $DGP-3$ respectively. On the other hand, the effect of the relative values of a and b is not straightforward. Initially, the size-adjusted power of all the statistics considered increases with the value of a. However, an extreme value of a that is higher than that of b (this is the case in $DGPs$ 7, 8 and 9) has a negative effect on the behavior of most statistics. However, such extreme values of the $ARCH$ parameter are rarely observed in empirical studies where the estimate of the $GARCH$

[13]In almost all the cases, the power of the statistics examined is identical for the three alternative schemes. We only observe minor differences of less than 5 percent.

[14]The superiority of $OOS-F$ over S_1 and $OOS-t$ when the number of the additional parameters of the unrestricted model and $\pi := \frac{P}{R}$ are small is also reported by McCracken (2004).

parameter is usually much higher than that of the $ARCH$ parameter.

3. The estimation sample size R has minor effects on the size-adjusted power of all tests as long as the number of observations is sufficient to ensure reliability of the parameter estimators. More specifically, as the estimation sample size increases, we observe small increases in the size-adjusted power of the tests. For example, in the case of $DGP-1$ and for $P = 240$ (that is, when $\pi = 0.4$ and 0.2 for $R = 600$ and 1200 respectively) the size-adjusted power of the $OOS-F$ statistic (Table 3) increases by only 2 percent (from 83 percent to 85 percent), while in the case of $DGP-6$ and for $P = 240$ the size-adjusted power of $OOS-F$ remains unchanged to 98 percent.

4. On the other hand, the number of out-of-sample forecasts P is crucial for the power of all tests considered in the simulation. The size-adjusted power of the tests increases substantially with P, reaching 100 percent in some cases. For example, for $R = 600$ and $DGP-1$ the size-adjusted power of the $OOS-F$ test increases from 58 to 83 percent for $\pi = 0.1$ and 0.4 respectively. We should note that the $OOS-t$ statistic is very sensitive to the value of R. At least 240 out-of-sample forecasts are necessary so that the size-adjusted power of $OOS-t$ surpasses 50 percent.

Case B: Leptokurtic Innovations

The assumption of leptokurtic innovations has two important implications with opposite effect on the power of the statistics considered. The formulas for the MSE of the volatility forecasts derived in Section 2 imply that the superiority of the volatility forecasts of the GARCH(1,1) model over the homoscedastic model is more evident in the case of leptokurtic distributions. This is expected to increase the power of the statistics examined in this paper. On the other hand, leptokurtic innovations may have a negative effect on the Maximum Likelihood estimates of the GARCH parameters, increasing the finite sample bias. This can lead to the deterioration of the forecast accuracy of the GARCH(1,1) model, resulting in smaller differences between the MSEs of the forecasts of the two models. Finally, we should note that when innovations follow a t-distribution with 5 degrees of freedom, $DGP-1$ is the only case from the nine DGPs reported that satisfies the fourth order stationarity condition C(1).

The main results, reported in Tables 7-10, can be summarized as follows.

1. In general, the size-adjusted power of most of the test statistics considered in the simulation decreases in the case of leptokurtic innovations compared to the case of Gaussian innovations.

For example, for $DGP-1$, $R=1200$ and $\pi=0.1$, the size-adjusted power of $OOS-F$ is 73 and 65 percent for the Gaussian and t(5)-distribution respectively.

2. Similarly to the Gaussian case, the size-adjusted power of all the tests increases with the number of out-of-sample forecasts, P, and the persistence of the process under examination.[15] For example, when $R=600$ and $\pi=0.1$, the size-adjusted power of the MGN test (Table 10) is 22 and 42 percent for $DGP-1$ and $DGP-3$ respectively. The size-adjusted power of MGN further increases to 74 percent for $DGP-3$ and $\pi=0.4$. Once again, the estimation sample size, R, has minor effects on the behavior of all tests.

3. In the case of leptokurtic innovations, the size-adjusted power of the three statistics proposed by McCracken (2004) declines compared to the Gaussian case. $OOS-F$ still outperforms the other two statistics but a large number of out-of-sample forecasts (P) is now necessary to achieve a power over 80 percent. For example, in the case of $DGP-1$ and for $R=1200$, the size-adjusted power of $OOS-F$ is 65 and 86 percent for $\pi=0.1$ and 0.4 respectively.

In summary, the simulation reveals the superiority of $OOS-F$ over all the other test statistics considered. $OOS-F$ has good size properties (only minor finite sample size distortions are observed), while it has substantial higher power than the other statistics. Moreover, it seems robust to leptokurtic innovations, since its size and power properties are only marginally affected by the violation of the normality of the error term. S_1 has similar size properties but it has substantially lower size-adjusted power than $OOS-F$. The performance of $OOS-t$ is very poor, unless the researcher uses a large number of observations for the out-of-sample forecast evaluation. The other statistics examined suffer from substantial size distortions and low power.

The results show that the MSE criterion remains capable to evaluate properly the relative forecast accuracy of the models considered here, even if the squared residuals are used as a proxy for the unobserved conditional variance. However, the researcher has to use proper statistical procedures (that is, tests that are suitable to compare nested models) for the relative forecast evaluation of the two models. Furthermore, the simulation shows that the power of all the statistics increases substantially with the number of out-of-sample forecasts, reaching 100 percent in some cases. On the other hand, the estimation sample size has minor effect on the power of the tests. Thus, applied researchers should use a large proportion of the available sample for the out-of-sample exercise in order to increase the power of the selected statistic.

[15] In few cases, the size-adjusted power of $OOS-t$ slightly decreases with the persistence of the process.

5 An Empirical Illustration

In this section, we examine how the test statistics perform in practical settings with an application to exchange rates. We first generate forecasts for the variance of the exchange rates based on a GARCH(1,1) model and an homoscedastic model. We then compare the relative forecast accuracy of the two alternative models. Given the simulation results, presented in the previous section, among all the statistics considered in this study the three statistics proposed by McCracken (2004), that is S_1, $OOS-t$ and $OOS-F$, are the most reliable ones. These statistics are designed to compare nested models and our simulation revealed that they have good size and power properties. Therefore, our empirical analysis is based on these three statistics. We consider the data used by West and Cho (1995), that is we examine five bilateral exchange rates for the US dollar. More specifically, we use weekly spot exchange rates for the US dollar versus the currencies of Canada, France, Germany, Japan and the UK. The data are provided by the Federal Reserve Bank of New York and the sample covers a period of 16 years, from March 7, 1973, to September 20, 1989.[16] The exchange rates are measured as dollars per unit of foreign currency.

The analysis is based on the logarithmic differences of the series. To be more specific, the variables examined are:

$$x_t = 100 * [\ln(\epsilon_t) - \ln(\epsilon_{t-1})]$$

where ϵ_t is the spot exchange rate at period t. Thus, an initial observation is lost due to differencing, resulting in a sample of 863 observations. Both levels, ϵ_t, and first differences, x_t, of the series are displayed in Figure 3. Volatility clustering is obvious in the first differences of the series, suggesting that the variables should be described by a model that allows for time heterogeneity in the conditional variance. Some useful descriptive statistics of the first differences of the series under examination are provided in Table 11.

Based on the MSE criterion, West and Cho (1995) find that for a one week horizon, the GARCH(1,1) model produces more accurate forecasts for the conditional volatility than the homoscedastic model in three out of the five currencies examined. The GARCH(1,1) model outperforms the homoscedastic model for the currencies of Canada, Japan and the UK. On the other hand, the superiority of the GARCH(1,1) model is lost for longer forecast horizons. However, when West and Cho (1995) compare the relative forecast accuracy the two models

[16] Similarly to West and Cho (1985), the data are Wednesday, New York spot rates. When Wednesday was a holiday, we use the Thursday spot rate; when Thursday was a holiday as well, we used the Tuesday spot rate.

based on statistical testing, they fail to reject the null hypothesis of equal forecast accuracy in all cases. In other words, they conclude that the GARCH(1,1) model does not provide more accurate volatility forecasts than the homoscedastic model for all the currencies and all the forecast horizons examined.

We now repeat the comparison of the two models based on S_1, $OOS-t$ and $OOS-F$. Similarly to West and Cho (1985), we begin the out-of-sample forecast exercise at the midpoint of the sample. The estimation results for the GARCH(1,1) model for the first estimation sample (i.e. for the first 432 observations) are presented in Table 12.[17] The estimates suggest a fourth order stationary process for the Canada and the UK. On the other hand, the exchange rates of France, Germany and Japan are highly persistent, corresponding to second order stationary processes.

We generate sequences of 1−step ahead volatility forecasts for each of the two models based on either the recursive or rolling scheme. We test the null hypothesis of equal forecast accuracy of the GARCH(1,1) model and the homoscedastic model, against the alternative of superiority (i.e. smaller MSE) of the GARCH(1,1) model. Thus, we have an one-sided test. In our case, the number of in-sample observations equals the number of out-of-sample observations, that is $R = P$.

The results for the three test statistics for both the recursive and the rolling schemes are presented in Tables 13. We consider the $OOS-t$ test to be reliable because the number of out-of-sample forecasts used in the experiment is high enough to assure good size and power properties. However, among the three statistics, $OOS-F$ has the higher power and thus it is preferable to the other statistics.

In the case of the rolling scheme, the most powerful test, $OOS-F$, indicates that the GACRH(1,1) model outperforms the homoscedastic model in the cases of Canada, Japan and the UK. The null hypothesis of equal forecast accuracy is rejected at an 1 percent level. Similarly, the other two statistics, S_1 and $OOS-t$, indicate superiority of the GARCH(1,1) model at either 5 or 10 percent level. In the case of France and Germany, the MSE of the homoscedastic model turns out to be lower than that of the GARCH(1,1) model. As a result, all the tests fail to reject the null hypothesis of equal forecast accuracy, since we have an one-sided test. The results for the recursive scheme are similar. However, in the case of Japan the superiority of

[17] We also examined the Italian data. However, we excluded Italy from the experiment because the GARCH estimates corresponded to an explosive process.

the GARCH(1,1) model is only supported by $OOS-F$.

Contrary to the findings of West and Cho (1985), our experiment suggests that the GARCH(1,1) model generates more accurate volatility forecasts than the homoscedastic model in the case of Canada, Japan and the UK. It is therefore obvious that statistical testing supports the superiority of the GARCH(1,1) model over the homoscedastic model in many cases, given that (i) the comparison of the two models is based on proper statistics for nested models and (ii) the evaluation is based on sequences of $1-$step ahead forecasts, rather than sequences of $s-$step ahead forecasts ($s>1$), in order to increase the power of the tests. Furthermore, the results indicate that the MSE criterion remains capable to evaluate properly the relative forecast performance of the two models.

Finally, we repeat the comparison of the two models for forecasting the volatility of the same five exchange rates for a different period. More specifically, we use daily data from August 31, 1971 to December 31, 1998. Once again, the analysis is based on the first differences of the series, x_t. We now have 7128 observations and we set $\pi=\frac{P}{R}$ equal to 0.4. The results for the recursive scheme, reported in Table 14, support the superiority of the GARCH(1,1) model over the homoscedastic model in all cases but France. The preferable $OOS-F$ statistic and the $OOS-t$ statistic (for Canada, Germany and the UK) indicate that the GARCH(1,1) model outperforms the homoscedastic model.

6 Conclusions

A number of previous empirical studies compare the conditional volatility forecasts of the GARCH(1,1) model to those of a naive model that assumes constant conditional variance (homoscedasticity) and conclude that the GARCH(1,1) model does not outperform the naive model. Applied researchers usually compare the relative forecasting accuracy of the two models by means of a selected formal statistical test which is based on the MSEs of the forecasts. In this study, we investigated the reasons for the poor forecasting performance of the GARCH(1,1) model reported in empirical studies.

Andersen and Bollerslev (1998) and Christodoulakis and Satchell (1998, 2003) have argued that the poor forecasting performance of the GARCH models in empirical studies is deceptive, resulting from the utilization of the squared shocks as a proxy for the unobserved true conditional volatility. This approximation results in an increase in the MSE of the forecasts (MSE-inflation), delivering a misleading evaluation of the GARCH models. However, in this study we show that the poor forecasting performance of the GARCH(1,1) model compared to that of the naive homoscedastic model cannot be attributed to the utilization of the squared errors as a proxy for the unobserved true conditional variance, since the utilization of this proxy "inflates" equally the MSEs of the forecasts of both models. Therefore, our analytical results suggest that if we assume that heteroscedasticity is present in real data, then the seemingly poor forecasting performance of GARCH(1,1) in empirical studies can only result from the low power of the statistical tests used to compare the relative forecasting accuracy of the GARCH(1,1) and the homoscedastic models.

We go one step forward and show that the difference between the MSEs of the forecasts of the two models (i.e. the GARCH(1,1) and homoscedastic models) is maximized for one-step ahead forecasts, while the two models tend to become equivalent as the forecast horizon increases. This finding suggests that applied researchers should base the statistical testing of the relative forecast accuracy of the two models on sequences of one-step ahead forecasts in order to increase the power of the tests. More importantly, researchers should note that the two models are nested and thus the comparison of the relative forecasting performance of the models should be based on statistical tests which are proper for nested model. Otherwise, the results of the tests are not reliable. The results of a Monte Carlo simulation that examines the size and power properties of a variety of statistical procedures to select the GARCH(1,1) model over the misspecified homoscedastic model show that the tests that are designed to compare non-nested models suffer from severe size distortions when applied to nested models. On the other hand, tests designed to compare the forecasting accuracy of nested models have good size properties and high power to indicate the superiority of the GARCH(1,1) model over the homoscedastic model. Finally, we illustrate the proper procedure for testing the relative forecast accuracy of the GARCH(1,1) model and the homoscedastic model with an application to five bilateral exchange rates for the US dollar. The results indicate the superiority of the GARCH(1,1) model over the homoscedastic model for forecasting the conditional variance in three out of the five exchange rates considered.

References

[1]Andersen, T.G. and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39, 885-905.

[2]Ashley, R., C.W.J. Granger and R. Schmalensee (1980). Advertising and Aggregate Consumption: An Analysis of Causality. *Econometrica* 48, 1149-1167.

[3]Baillie, R.T and T. Bollerslev (1992). Prediction in Dynamic Models with Time Dependent Conditional Variances. *Journal of Econometrics*, Vol.52, No.1, pp.91-113; Reprinted in The International Library of Critical Writings in Economics: *Economic Forecasting* (ed. Terence C. Mills), London: Edward Elgar Publishing Limited, (1998).

[4]Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307-327; Reprinted in *The International Library of Critical Writings in Econometrics: Time Series* (ed. Andrew Harvey), London: Edward Elgar Publishing Limited, (1994); Reprinted in *ARCH: Selected Readings* (ed. Robert F. Engle), Oxford: Oxford University Press, (1995); Reprinted in *Foundations of Probability, Econometrics and Economic Games* (eds. O.F. Hamouda and J.C.R. Rowley), London: Edward Elgar Publishing Limited, (1996); Reprinted in *Journal of Econometrics, 100th Anniversary Commemorative Issue*, Vol.100, No.1, (2001).

[5]Carrasco, M. and X. Chen (2002). Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18, 17-39.

[6]Christodoulakis, G.A. and S.E. Satchell (1998). Hashing Garch. In Knight, J. and S.E. Satchell (eds.), *Forecasting Volatility in Financial Markets*, Butterworth-Heinemann, Oxford.

[7]Christodoulakis, G.A. and S.E. Satchell (2003). Forecast Evaluation in the Presence of Unobserved Volatility. *City University*, Cass Business School, Working Paper.

[8]Davidson, J. (2001). Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes. *Journal of Econometrics* 106, 243-269.

[9]Diebold, F.X. and R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-265; Reprinted in *Economic Forecasting* (T.C. Mills,

ed.), part of the International Library of Critical Writings in Economics (M. Blaug, series ed.), Edward Elgar Publishing, (1998); Reprinted in E. Ghysels and A. Hall (eds.), *Special Twentieth Anniversary Commemorative Issue of Journal of Business and Economic Statistics* 20, 134-144, (2002), containing the ten most influential papers published in the journal's first twenty years.

[10]Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica* 50, 987-1007.

[11]Franses, P.H. and D. van Dijk (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.

[12]Granger, C.W.J. and P. Newbold (1977). *Forecasting Economic Time Series*. New York: Academic Press.

[13]Hafner, C.M. (2003). Fourth Moment Structure of Multivariate GARCH Models. *Journal of Financial Econometrics* 1, 26-54.

[14]Hansen, B.E. (1991). Strong Laws for Dependent Heterogeneous Processes. *Econometric Theory* 7, 213-221.

[15]Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day, Inc.

[16]Lopez, J.A. (2001). Evaluating the Predictive Accuracy of Volatility Models. *Journal of Forecasting* 20, 87-109.

[17]McCracken, M.W. (2004). Asymptotics for Out-of-Sample Tests of Granger Causality. Department of Economics, University of Missouri-Columbia, Working Paper.

[18]Nelson, D.B. (1990). Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory* 6, 318-334.

[19]Sheskin, D. J. (2000). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall, Boca Raton, Florida, USA.

[20]West, K.D. and D. Cho (1995). The Predictive Ability of Several Models of Exchange Rate Volatility. *Journal of Econometrics* 69, 367-391; Reprinted in 505-529 in Volume 2 of *Economic Forecasting*, T. Mills (ed.), Cheltenham: Edward Elgar Publishing Ltd.

[21]West, K.D., H.J. Edison and D. Cho (1993). A Utility Based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics* 35, 23-45.

Appendix

Proof of Proposition 1

First, we observe that $E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2] \overset{s\to\infty}{\to} E[(h^2_{t+s} - \sigma^2)^2]$. Thus, $R1 \overset{s\to\infty}{\to} 1$. Given that both $E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2]$ and $E[(h^2_{t+s} - \sigma^2)^2]$ are positive by definition and that $E[(h^2_{t+s} - \sigma^2)^2]$ is independent of s, the proposition will have been proved if we show that $E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2]$ is an increasing function of s.

$E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2] - E[(h^2_{t+s-1} - \widehat{h}^2_{t+s-1})^2] = (a+b)^{2(s-2)}\{\frac{k_2-1}{k_2}a^2[3cw\sigma^2\frac{1+a+b}{1-3ca^2-2ab-b^2}]\} = (a+b)^{2(s-2)}MSE^{NF}_G(2) > 0$.

Thus, $E[(h^2_{t+s} - \widehat{h}^2_{t+s})^2]$ is an increasing function of s and the proposition has been proved.■

Proof of Proposition 2

First, we observe that $E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2] \overset{s\to\infty}{\to} E[(e^2_{t+s} - \sigma^2)^2]$. Thus, $R2 \overset{s\to\infty}{\to} 1$. Given that both $E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2]$ and $E[(e^2_{t+s} - \sigma^2)^2]$ are positive by definition and that $E[(h^2_{t+s} - \sigma^2)^2]$ is independent of s, the proposition will have been proved if we show that $E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2]$ is an increasing function of s.

$E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2] - E[(e^2_{t+s-1} - \widehat{h}^2_{t+s-1})^2] = (a+b)^{2(s-2)}[1-(a+b)^2][\frac{3c}{k_2}\frac{1+a+b}{1-3ca^2-2ab-b^2} - \frac{1}{1-a-b}] > 0$, since (i) $(a+b)^{2(s-2)} > 0$ because $a, b > 0$, (ii) $[1-(a+b)^2] > 0$ because $a+b < 1$ (the process is assumed to be fourth order stationary) and (iii) $[\frac{3c}{k_2}\frac{1+a+b}{1-3ca^2-2ab-b^2} - \frac{1}{1-a-b}] = \frac{1}{w\sigma^2}MSE^F_H(s) > 0$.

Thus, $E[(e^2_{t+s} - \widehat{h}^2_{t+s})^2]$ is an increasing function of s and the proposition has been proved.■

Figure 1: $MSEI_G$ for w=1, $\alpha \in [0 , 0.2]$ and $b \in [0 , 0.5]$

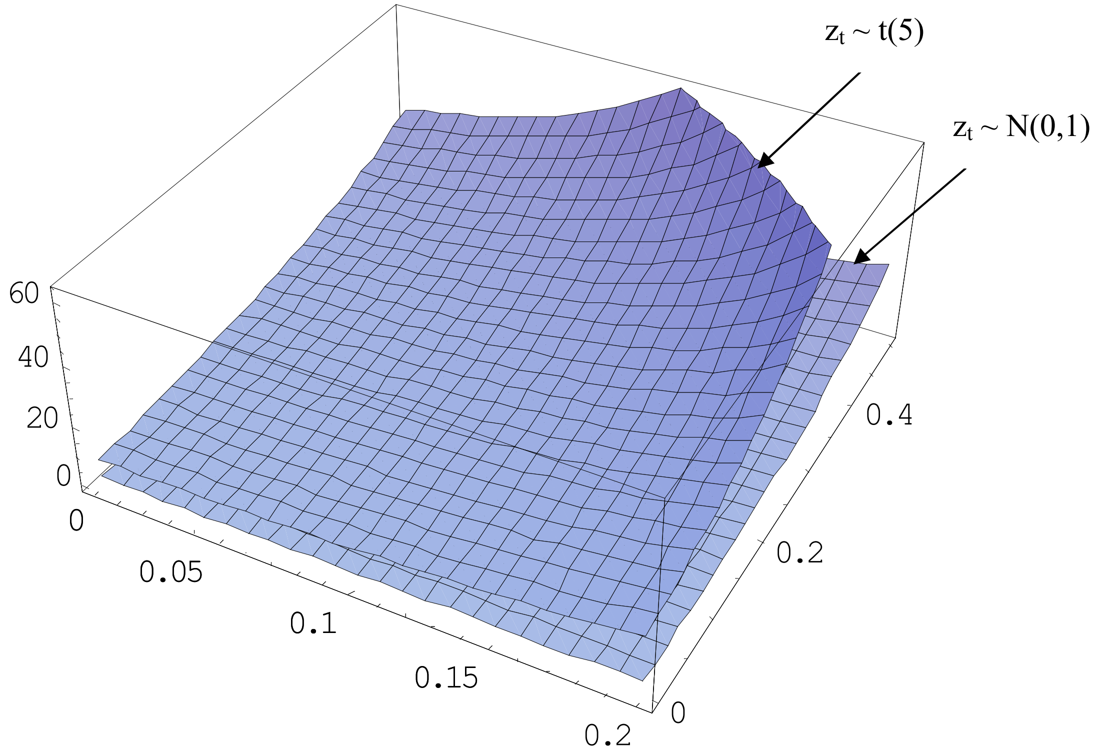

Figure 2: $MSE_H - MSE_G$ for s=1, 5 {$z_t \sim N(0,1)$, w=1, $\alpha \in [0 , 0.2]$ and $b \in [0 , 0.5]$}

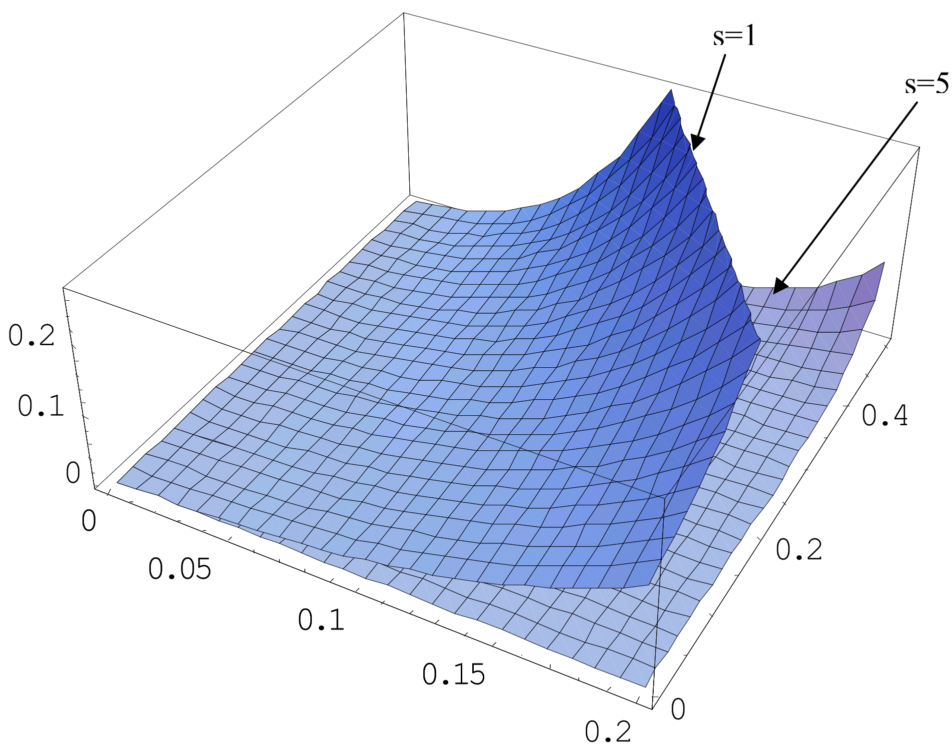

Figure 3: Levels and First Differences of the Exchange Rates

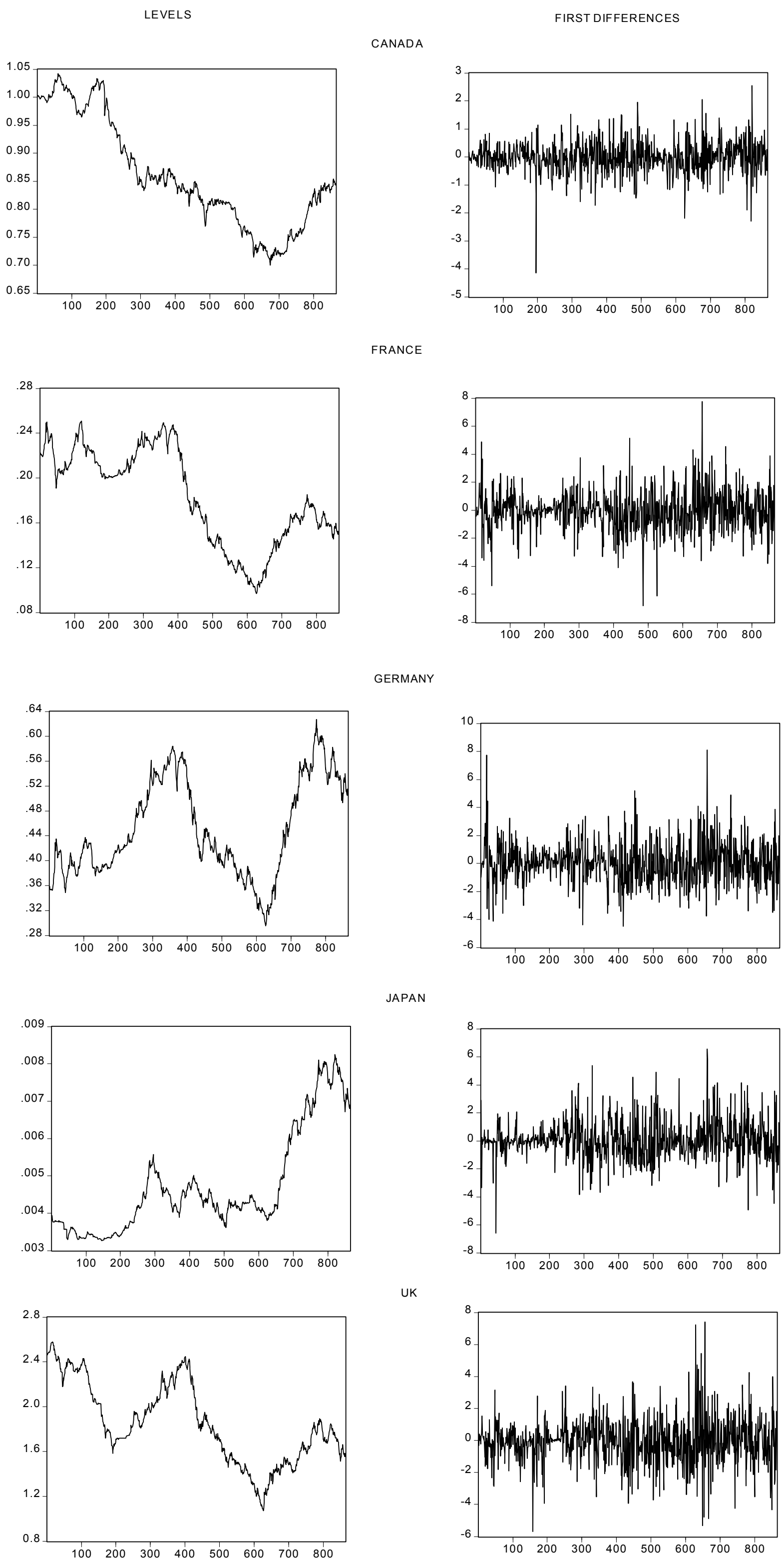

Table 1:
Empirical Sizes (Static Forecasts, Normal Distribution). Nominal Size 5%

	T=300			T=600			T=900			T=1200		
Test	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
OOS-F	0.08	0.07	0.05	0.05	0.06	0.08	0.07	0.05	0.06	0.05	0.07	0.06
S_1	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.04	0.05	0.04	0.05	0.06
OOS-t	0.01	0.01	0.02	0.01	0.02	0.03	0.02	0.03	0.02	0.02	0.02	0.05
F-test	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MGN	0.04	0.02	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.02
MR	0.13	0.14	0.14	0.12	0.14	0.12	0.13	0.13	0.11	0.13	0.12	0.10
S_1-DM	0.06	0.02	0.02	0.03	0.02	0.01	0.02	0.02	0.01	0.02	0.01	0.02
S_2	0.24	0.25	0.27	0.25	0.26	0.34	0.25	0.30	0.38	0.26	0.30	0.25
S_2*	0.17	0.20	0.23	0.21	0.23	0.34	0.25	0.27	0.38	0.22	0.30	0.25
S_3	0.21	0.25	0.26	0.25	0.29	0.33	0.31	0.30	0.32	0.27	0.33	0.34
VDW	0.08	0.07	0.08	0.08	0.08	0.10	0.09	0.09	0.08	0.07	0.09	0.06

Table 2:
Empirical Sizes (Static Forecasts, t(5)-Distribution). Nominal Size 5%

	T=300			T=600			T=900			T=1200		
Test	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
OOS-F	0.05	0.06	0.05	0.11	0.11	0.06	0.11	0.07	0.07	0.07	0.07	0.05
S_1	0.02	0.09	0.08	0.10	0.09	0.05	0.09	0.05	0.04	0.06	0.05	0.06
OOS-t	0.56	0.07	0.06	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.06
F-test	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MGN	0.00	0.02	0.01	0.06	0.05	0.02	0.05	0.03	0.01	0.03	0.03	0.02
MR	0.00	0.11	0.17	0.13	0.12	0.15	0.14	0.13	0.13	0.10	0.10	0.16
S_1-DM	0.01	0.05	0.02	0.06	0.04	0.01	0.06	0.03	0.01	0.04	0.02	0.01
S_2	0.39	0.26	0.28	0.42	0.54	0.60	0.44	0.58	0.70	0.53	0.65	0.39
S_2*	0.26	0.23	0.26	0.38	0.50	0.60	0.44	0.55	0.70	0.48	0.65	0.38
S_3	0.14	0.43	0.49	0.38	0.46	0.50	0.43	0.50	0.58	0.45	0.51	0.69
VDW	0.03	0.11	0.13	0.18	0.22	0.21	0.20	0.21	0.25	0.20	0.26	0.13

Table 3:
Size-Adjusted Power of OOS-F in Percentages (Normal Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	40.0	51.0	68.0	58.0	71.0	83.0	68.0	81.0	93.0	73.0	85.0	97.0
	0.895	55.0	67.0	81.0	73.0	85.0	95.0	82.0	92.0	98.0	89.0	97.0	100.0
	0.9	56.0	66.0	81.0	76.0	85.0	96.0	85.0	93.0	98.0	90.0	97.0	100.0
0.2	**0.72**	54.0	64.0	79.0	70.0	82.0	92.0	80.0	90.0	98.0	85.0	93.0	99.0
	0.77	63.0	72.0	87.0	77.0	89.0	96.0	88.0	96.0	99.0	91.0	97.0	100.0
	0.8	68.0	78.0	90.0	84.0	93.0	98.0	91.0	97.0	100.0	94.0	98.0	100.0
0.5	**0.15**	54.0	62.0	73.0	64.0	73.0	84.0	71.0	80.0	90.0	76.0	84.0	93.0
	0.45	68.0	73.0	81.0	76.0	83.0	91.0	83.0	89.0	95.0	88.0	92.0	95.0
	0.5	72.0	74.0	81.0	79.0	86.0	92.0	86.0	90.0	95.0	89.0	93.0	96.0

Table 4:
Size-Adjusted Power of S_1 in Percentages (Normal Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	21.0	31.0	46.0	25.0	41.0	61.0	31.0	53.0	79.0	38.0	57.0	81.0
	0.895	33.0	46.0	65.0	48.0	63.0	83.0	58.0	78.0	94.0	69.0	84.0	96.0
	0.9	32.0	46.0	62.0	49.0	65.0	84.0	62.0	81.0	95.0	72.0	86.0	97.0
0.2	**0.72**	24.0	37.0	55.0	30.0	48.0	73.0	40.0	64.0	86.0	46.0	67.0	91.0
	0.77	34.0	45.0	65.0	44.0	61.0	82.0	53.0	77.0	94.0	61.0	80.0	94.0
	0.8	43.0	55.0	71.0	57.0	69.0	88.0	67.0	83.0	96.0	76.0	88.0	95.0
0.5	**0.15**	22.0	31.0	46.0	20.0	28.0	52.0	25.0	43.0	70.0	28.0	45.0	68.0
	0.45	37.0	44.0	55.0	41.0	48.0	66.0	48.0	62.0	76.0	54.0	61.0	72.0
	0.5	44.0	47.0	56.0	48.0	54.0	67.0	55.0	65.0	77.0	61.0	65.0	71.0

Table 5:
Size-Adjusted Power of OOS-t in Percentages (Normal Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	5.0	15.0	32.0	16.0	28.0	54.0	19.0	41.0	69.0	25.0	49.0	76.0
	0.895	4.0	13.0	32.0	11.0	27.0	62.0	16.0	45.0	81.0	24.0	54.0	88.0
	0.9	3.0	12.0	31.0	11.0	27.0	63.0	16.0	45.0	83.0	25.0	56.0	90.0
0.2	**0.72**	8.0	26.0	47.0	24.0	41.0	68.0	33.0	57.0	80.0	41.0	65.0	87.0
	0.77	7.0	24.0	48.0	21.0	42.0	71.0	29.0	60.0	86.0	41.0	68.0	92.0
	0.8	7.0	24.0	52.0	20.0	44.0	75.0	30.0	64.0	89.0	41.0	72.0	95.0
0.5	**0.15**	21.0	38.0	52.0	34.0	46.0	67.0	42.0	55.0	75.0	48.0	65.0	80.0
	0.45	13.0	32.0	46.0	26.0	43.0	64.0	34.0	53.0	74.0	42.0	61.0	80.0
	0.5	12.0	29.0	47.0	24.0	41.0	63.0	32.0	52.0	74.0	38.0	59.0	78.0

Table 6:
Size-Adjusted Power of MGN in Percentages (Normal Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	23.0	36.0	52.0	33.0	49.0	67.0	40.0	59.0	83.0	44.0	64.0	87.0
	0.895	35.0	51.0	68.0	55.0	69.0	86.0	63.0	82.0	96.0	73.0	89.0	98.0
	0.9	33.0	51.0	68.0	55.0	71.0	88.0	66.0	84.0	96.0	76.0	90.0	98.0
0.2	**0.72**	29.0	45.0	64.0	42.0	59.0	79.0	51.0	72.0	90.0	58.0	77.0	95.0
	0.77	38.0	56.0	73.0	54.0	72.0	88.0	64.0	84.0	97.0	71.0	88.0	98.0
	0.8	45.0	62.0	79.0	64.0	79.0	92.0	76.0	89.0	98.0	82.0	94.0	99.0
0.5	**0.15**	27.0	43.0	56.0	36.0	51.0	70.0	44.0	58.0	79.0	51.0	68.0	84.0
	0.45	43.0	54.0	64.0	54.0	65.0	79.0	61.0	74.0	87.0	69.0	79.0	89.0
	0.5	49.0	56.0	66.0	58.0	70.0	81.0	66.0	77.0	88.0	72.0	83.0	90.0

Table 7:
Size-Adjusted Power of OOS-F in Percentages (t(5)-Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	40.0	47.0	56.0	44.0	54.0	70.0	56.0	68.0	80.0	65.0	76.0	86.0
	0.895	50.0	57.0	70.0	64.0	72.0	86.0	73.0	84.0	93.0	80.0	90.0	96.0
	0.9	50.0	58.0	71.0	64.0	74.0	88.0	75.0	85.0	95.0	81.0	91.0	96.0
0.2	**0.72**	48.0	54.0	62.0	54.0	61.0	74.0	64.0	74.0	82.0	70.0	79.0	86.0
	0.77	54.0	61.0	70.0	64.0	71.0	82.0	74.0	83.0	88.0	78.0	86.0	92.0
	0.8	60.0	65.0	74.0	71.0	77.0	87.0	81.0	87.0	92.0	85.0	90.0	94.0
0.5	**0.15**	48.0	49.0	53.0	47.0	50.0	58.0	52.0	59.0	62.0	55.0	60.0	65.0
	0.45	57.0	59.0	60.0	60.0	63.0	66.0	67.0	67.0	70.0	68.0	70.0	73.0
	0.5	62.0	61.0	62.0	65.0	67.0	68.0	70.0	71.0	71.0	72.0	73.0	74.0

Table 8:
Size-Adjusted Power of S_1 in Percentages (t(5)-Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	51.0	22.0	28.0	21.0	27.0	50.0	24.0	44.0	64.0	31.0	45.0	63.0
	0.895	60.0	32.0	44.0	38.0	46.0	68.0	46.0	64.0	81.0	54.0	69.0	83.0
	0.9	59.0	33.0	45.0	41.0	48.0	70.0	49.0	67.0	84.0	58.0	71.0	86.0
0.2	**0.72**	57.0	24.0	30.0	22.0	26.0	51.0	26.0	46.0	66.0	34.0	46.0	61.0
	0.77	61.0	30.0	37.0	31.0	36.0	60.0	39.0	55.0	72.0	45.0	57.0	72.0
	0.8	64.0	37.0	43.0	45.0	48.0	66.0	50.0	63.0	77.0	55.0	67.0	77.0
0.5	**0.15**	51.0	18.0	21.0	14.0	14.0	33.0	15.0	29.0	45.0	17.0	25.0	34.0
	0.45	60.0	28.0	28.0	28.0	28.0	43.0	32.0	42.0	53.0	34.0	38.0	43.0
	0.5	63.0	33.0	32.0	35.0	34.0	46.0	38.0	45.0	53.0	40.0	43.0	45.0

Table 9:
Size-Adjusted Power of OOS-t in Percentages (t(5)-Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	0.0	2.0	8.0	8.0	18.0	43.0	13.0	35.0	54.0	24.0	37.0	48.0
	0.895	0.0	3.0	10.0	7.0	20.0	52.0	12.0	38.0	68.0	23.0	42.0	61.0
	0.9	0.0	3.0	11.0	7.0	20.0	52.0	12.0	41.0	69.0	24.0	44.0	66.0
0.2	**0.72**	1.0	4.0	15.0	14.0	25.0	51.0	21.0	45.0	62.0	32.0	43.0	55.0
	0.77	1.0	4.0	17.0	12.0	26.0	55.0	18.0	45.0	66.0	32.0	45.0	60.0
	0.8	1.0	6.0	19.0	11.0	29.0	58.0	18.0	46.0	70.0	32.0	48.0	65.0
0.5	**0.15**	1.0	12.0	24.0	22.0	30.0	46.0	27.0	43.0	48.0	34.0	40.0	47.0
	0.45	1.0	9.0	20.0	16.0	25.0	42.0	20.0	37.0	47.0	28.0	35.0	43.0
	0.5	1.0	8.0	20.0	14.0	24.0	41.0	18.0	35.0	46.0	27.0	36.0	42.0

Table 10:
Size-Adjusted Power of MGN in Percentages (t(5)-Distribution, Static Forecasts)

		T=300			T=600			T=900			T=1200		
a	**b**	**P/R=0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**	**0.1**	**0.2**	**0.4**
0.1	**0.85**	53.0	30.0	34.0	22.0	26.0	55.0	28.0	44.0	64.0	35.0	46.0	68.0
	0.895	61.0	39.0	49.0	40.0	46.0	72.0	48.0	64.0	83.0	55.0	71.0	88.0
	0.9	61.0	39.0	49.0	42.0	48.0	74.0	50.0	66.0	85.0	58.0	73.0	91.0
0.2	**0.72**	58.0	33.0	39.0	25.0	31.0	59.0	32.0	50.0	68.0	43.0	53.0	72.0
	0.77	61.0	40.0	46.0	34.0	40.0	68.0	45.0	59.0	78.0	52.0	65.0	83.0
	0.8	65.0	45.0	52.0	47.0	51.0	73.0	54.0	68.0	83.0	60.0	74.0	88.0
0.5	**0.15**	54.0	29.0	34.0	18.0	26.0	44.0	26.0	41.0	50.0	34.0	41.0	55.0
	0.45	60.0	41.0	42.0	34.0	38.0	52.0	39.0	50.0	58.0	45.0	51.0	62.0
	0.5	64.0	43.0	46.0	39.0	43.0	56.0	46.0	52.0	59.0	49.0	54.0	63.0

Table 11:
Descriptive Statistics of the First Differences of the Exchange Rates

	CANADA	FRANCE	GERMANY	JAPAN	UK
Mean	-0.0197	-0.0438	0.0420	0.0685	-0.0518
Median	-0.0297	0.0000	0.0226	-0.0263	-0.0216
Maximum	2.5505	7.7413	8.1133	6.5461	7.3974
Minimum	-4.1551	-6.8252	-4.4877	-6.5869	-5.6911
St. Deviation	0.5523	1.4083	1.4656	1.3607	1.4063
Skewness	-0.4210	0.1031	0.4801	0.3847	0.2609
Kurtosis	7.9541	5.4907	5.1336	5.5867	6.0920
Observations	863	863	863	863	863

Table 12:
Estimation Results for the GARCH(1,1) Model (from March 14, 1973 to June 17, 1981)

	w	a	b
CANADA	0.0557 (0.0133)	0.2597 (0.0243)	0.5442 (0.0622)
FRANCE	0.1372 (0.0262)	0.3470 (0.05447)	0.6137 (0.0437)
GERMANY	0.2007 (0.0446)	0.2973 (0.0492)	0.6149 (0.0533)
JAPAN	0.0076 (0.0035)	0.0530 (0.0076)	0.9443 (0.0069)
UK	0.1883 (0.0725)	0.1160 (0.0466)	0.7323 (0.0985)

Estimation Standard Errors are reported in parentheses.
All estimates are statistically significant at a 5 percent level.

Table 13:
Test Statistics, Weekly Data

Recursive	CANADA	FRANCE	GERMANY	JAPAN	UK
	Statistic	**Statistic**	**Statistic**	**Statistic**	**Statistic**
OOS-F	12.276***	-39.178	-18.195	6.750***	27.756***
S_1	0.412*	-1.134	-0.758	0.292	0.762**
OOS-t	0.411*	-3.137	-1.860	-0.327	1.187**

Rolling	CANADA	FRANCE	GERMANY	JAPAN	UK
	Statistic	**Statistic**	**Statistic**	**Statistic**	**Statistic**
OOS-F	16.482***	-28.198	-13.683	11.476***	15.131***
S_1	0.616*	-1.091	-0.686	0.532*	0.440*
OOS-t	1.057**	-2.277	-1.485	0.810**	0.510*

*** indicate rejection of the null hypothesis of equal forecast accuracy at 1 percent level.
** indicate rejection of the null hypothesis of equal forecast accuracy at 5 percent level.
* indicate rejection of the null hypothesis of equal forecast accuracy at 10 percent level.

Table 14:
Test Statistics. Daily data. Recursive Scheme.

Recursive	CANADA	FRANCE	GERMANY	JAPAN	UK
	Statistic	**Statistic**	**Statistic**	**Statistic**	**Statistic**
OOS-F	153.671***	-77.076	145.071***	12.856***	111.595***
S_1	0.441	-0.479	0.674*	0.049	0.580
OOS-t	2.852***	-0.735	4.291***	0.188	2.980***

*** indicate rejection of the null hypothesis of equal forecast accuracy at 1 percent level.
** indicate rejection of the null hypothesis of equal forecast accuracy at 5 percent level.
* indicate rejection of the null hypothesis of equal forecast accuracy at 10 percent level.