8-2004

Essays on the Econometrics of Financial Volatility

Yasemin Bardakci
Western Michigan University

ESSAYS ON THE ECONOMETRICS OF FINANCIAL VOLATILITY

by

Yasemin Bardakci

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Economics

ADVISOR: M. L. HIGGINS, Ph.D.

Western Michigan University
Kalamazoo, Michigan
August 2004

UMI Number: 3138765

ACKNOWLEDGMENTS

I sincerely thank my chair Prof. Matthew L. Higgins for his continuous support and supervision through out this dissertation. Without his help this work would have never been completed. I am also grateful to my other committee members, Prof. Debasri Mukherjee and Prof. Onur Arugaslan, who supported this project. I would also like to thank to my closest friends and my family for their emotional support.

Yasemin Bardakci

TABLE OF CONTENTS

CHAPTER

APPENDIX

LIST OF TABLES

LIST OF FIGURES

ESSAYS ON THE ECONOMETRICS OF FINANCIAL VOLATILITY

Yasemin Bardakci, Ph.D.

Western Michigan University, 2004

My dissertation consists of three essays on the econometric analysis of financial volatility.

My first essay is titled "The Runs Test for Volatility Forecastibility: Extensions and Comparisons with Tests for GARCH." Recently, Diebold and Christoffersen (2000) introduced a test for forecastable volatility. In this paper, I compare the size and the power of the runs test and the optimal LM test for GARCH by Monte Carlo simulation. For high frequency returns the LM test has superior power to the runs test. For low frequency returns however, the tests have very similar power. I also propose a switching variance model. For this process, I find that the runs test has greater power than the LM test.

The second essay of the dissertation I drive the population moments of criteria commonly used to evaluate accuracy of volatility forecasts from GARCH models. I state the existence conditions for the population moments. The criteria include the mean squared error (MSE), the mean absolute error (MAE) and a heteroscedasticity adjusted mean square error (HMSE). Using Monte Carlo simulation, I analyze the sampling properties of these criteria and the sampling properties of the R^2's and t-statistics from the Mincer and Zarnowitz (1969) regression. When volatility is highly

persistent, I find that the majority of the sampling distribution of the R^2 lies below the population R^2. Also, the t-statistics for testing forecast efficiency are unreliable. For a logarithmic version of the Mincer and Zarnowitz regression, I find that R^2's tend to be smaller, but inference concerning forecast efficiency are valid. Among the accuracy criteria I find that the HMSE is preferable.

My third essay considers situations when the loss function is asymmetric. Most of the forecast criteria used in the literature consider symmetric loss functions like MSE because of mathematical convenience. However, forecast evaluation results are very sensitive to the proper specification of the loss function. In this paper, I use both parametric and nonparametric estimation techniques for the optimal predictor of volatility when the loss function is asymmetric. The results suggest that a constant /time-varying bias term is important.

CHAPTER I

A MONTE CARLO COMPARISON OF THE RUNS TEST FOR VOLATILITY FORECASTABILITYAND THE LM TEST FOR GARCH USING AGGREGATED RETURNS

Introduction

Recently, Christoffersen and Diebold (2000) (henceforth abbreviated CD) presented a test for variance forecastability in long-horizon returns based on the number of runs in a hit indicator for a time invariant prediction interval for the returns. CD proposed their runs test as an omnibus test for volatility forecastability in long-horizon returns. CD argued that the G/ARCH model of Engle (1982) and Bollerslev (1986), which is widely used for daily asset returns, may not be appropriate for returns at longer horizons. In spite of the result of Drost and Nijman (1993), who showed that a weak GARCH process is closed under temporal aggregation[1]. CD asserted that the GARCH model is only an approximation for daily returns and may provide a poor approximation for returns at longer horizons. The advantage of their test is that it is not constructed against a specific alternative and there is an exact sampling theory for the distribution. Although CD presented a Monte Carlo simulation that showed their test appeared to have good power with aggregated returns when the daily return

[1] Drost and Nijman (1993) defined a weak GARCH process as a process where the time-varying volatility h_t is not necessarily the conditional variance of the return, but the projection of the squared return on lagged squared returns and lagged volatilities.

process is GARCH, they did not directly compare the power of their test to the conventionally used LM test for GARCH. To be a useful omnibus test, the runs test should have reasonably good power when the true volatility process is GARCH and should have superior power for some empirically relevant class of alternatives to the standard GARCH model.

In this paper we compare the size and power properties of the runs test and the LM test for returns at different horizons using Monte Carlo simulation. We first consider the tests when the daily return process is GARCH. We find that the LM test can have greater power than the runs test for returns up to a 30 day horizon. For longer horizons, the runs test performs as well as the LM test. We also consider a qualitative threshold GARCH model which produces a first order Markov hit indicator. For the threshold model, the runs test has superior power to the LM test. In the next section, we summarize the test procedure advocated by CD. In the third section, we describe the qualitative threshold ARCH model for which the runs test should be optimal for daily returns. The Monte Carlo results are presented in the fourth section. The last section is a conclusion.

Runs test for variance forecastability

To define the CD test, let $[-d,d]$ be a symmetric around the origin time invariant forecast interval for the observed sequence of returns y_t, $t=1,\dots,T$. Assume that the returns have been centered on their unconditional mean. Define the sequence of hit indicators $I_t = I(y_t \in [-d,d])$, where $I(\cdot)$ is the indicator function which takes the value one when the designated event occurs and zero otherwise. If the sequence of

returns are independent, the sequence of hit indicators will also be independent. CD suggested testing for dependence in the binary sequence I_t using the classical runs test. Their test is a general test for dependence in the sequence y_t. Because y_t represents a financial return, in accordance with efficient market theory, y_t should be serially uncorrelated in levels. Therefore, the test should be useful for detecting dependence in second or higher order moments. The value of d for which the test is computed should be linked to the dispersion of the returns. In their Monte Carlo study, CD found that the test has the greatest power when $d = 1.5 s_y$, where s_y is the unconditional standard deviation of the returns. They recommended, however, taking $d = 2.0 s_y$ to give unconditional hit probabilities in the range of .9 to .95. These probabilities correspond to those used in typical value-at-risk calculations.

CD stated that p-values for the runs test can be computed using the standard formula for the probability function for the number of runs. Computing the probability function for the number of runs, however, requires computing factorials of the magnitude $T!$. This is not practical for the type of return data for which the test is envisioned because T can be in the thousands. To obtain critical values for the runs test, we use the normal approximation to the probability function for the number of runs. Let n_0 and n_1 be the number of zeros and ones. Let r be the number of runs. Then the mean and variance of the number of runs is $\mu_r = 2n_0 n_1 / (n_0 + n_1) + 1$ and $\sigma_r^2 = (\mu_r - 1)(\mu_r - 2) / (n_0 + n_1 - 1)$. It can then be shown that $R = (r + .5 - \mu_r) / \sigma_r \xrightarrow{D} N(0,1)$ [see Campbell, Lo and MacKinlay (1997), p. 41]. The .5 that appears in the test statistic is the standard continuity correction that is used

when the probability function of a discrete random variable is approximated by the continuous standard normal density function.

Although CD presented their test in the guise of forecast evaluation, their test is obviously equivalent to the runs test for dependence applied to the squared returns y_t^2 using the threshold d^2. Therefore, when considering the performance of their test, it should be judged relative to other tests for dependence in second moments. The most commonly used model for time-varying conditional variances in economics is the GARCH model. Assuming a zero conditional mean, a GARCH(p,q) model for returns is

$$y_t = \sqrt{h_t} \cdot \varepsilon_t, \quad t = 1,\ldots,T,$$
$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i},$$

where $\varepsilon_t \sim IID(0,1)$, $\alpha_0 > 0$, $\alpha_i \geq 0,\ i = 1,\ldots,q$ and $\beta_i \geq 0,\ \mathrm{i} = 1,\ldots,\mathrm{p}$. The standard test for GARCH is the LM test. The LM test is computed as $T \cdot R^2$, where R^2 is the coefficient of determination from the regression of y_t^2 on an intercept and $y_{t-1}^2,\ldots, y_{t-q}^2$ [see Lee (1991)]. The statistic has an asymptotic χ_q^2 distribution. The LM test is a test for correlation between squared returns, where as the runs test is a test for dependence between level crossing of d^2 by the squared returns. Note that if all the β_i coefficients are equal to zero we have the ARCH model of Engle (1982).

A qualitative threshold ARCH model

To be a useful omnibus test for variance forecastability in long-horizon returns, the runs test should not only have reasonably good power relative to the LM test when

the true daily return generating process is GARCH, but should also have superior power for some empirically relevant alternative to the standard GARCH model. We focus on a process with forecastable variance for which the runs test is optimal. Lehmann (1986, p. 177) showed that the runs test is uniformly most powerful for detecting dependence in a binary sequence when the sequence is a first order Markov process. Suppose the return y_t is generated by

$$y_t = \sqrt{h_t} \cdot \varepsilon_t \quad t = 1,\dots,T,$$
$$h_t = \gamma_0 + \gamma_1 I^*_{t-1}$$
$$I^*_t = I(y_t \in [-\delta, \delta])$$

where $\varepsilon_t \sim IID(0,1)$. This process is an example of the qualitative threshold ARCH (QTARCH) model of Gourieroux and Monfort (1992) and is similar to the threshold models of Rabemananjara and Zakoian (1993) and Glosten, Jaganathan and Runkle (1993). The parameters in the conditional variance function are assumed to satisfy $\gamma_0 > 0$, $\gamma_1 < 0$ and $|\gamma_1| < \gamma_0$. The first inequality insures that $h_t > 0$ when $I_{t-1} = 0$. If the second inequality is satisfied, y_t will display positive volatility clustering. The third inequality insures $h_t > 0$ when $I_{t-1} = 1$.

The switch indicator I^*_t is easily seen to be a first order Markov process. Let $F(\varepsilon)$ be the CDF of ε_t. The transition probabilities for the indicator are

$$\pi_{11} \equiv P(I^*_t = 1 \mid I^*_{t-1} = 1) = 2F(\delta / \sqrt{\gamma_0 + \gamma_1}) - 1 \tag{1}$$
$$\pi_{01} \equiv P(I^*_t = 1 \mid I^*_{t-1} = 0) = 2F(\delta / \sqrt{\gamma_0}) - 1. \tag{2}$$

The interval bound δ that governs the variance switch is an unobserved parameter. If the bound was known, the runs test could be computed directly with I^*_t. In practice,

the bound δ and the latent indicator I_t^* must be estimated. Using the law of iterated expectations, the unconditional variance is seen to be $\mathrm{var}(y_t) = \sigma_y^2 = \gamma_0 + \gamma_1 p$, where

$$p \equiv P(I_t = 1) = \frac{\pi_{01}}{1 + \pi_{01} - \pi_{11}} \tag{3}$$

is the unconditional probability of the latent indicator recording a hit. Therefore, we may write

$$\delta = m \cdot \sigma_y = m\sqrt{\gamma_0 + \gamma_1 p}\ . \tag{4}$$

For the computation of the runs test, we assume the user believes they know the relevant multiple m. The standard deviation σ_y is estimated with the sample standard deviation s_y. The test is then computed with the bound $d = m \cdot s_y$. In the Monte Carlo, we will consider the performance of the runs test when m is correctly and incorrectly specified.

Monte Carlo

In this section we compare the size and power properties of the GARCH LM test and the runs test. We consider LM tests using one and five lagged squared returns. The two tests are denoted by LM(1) and LM(5). Two runs test are also considered. The first uses $d = 1.5 \cdot s_y$, which CD found had the greatest power. The second uses $d = 2 \cdot s_y$, which CD claimed is the most relevant for financial forecasting. These two tests are denoted by R(1.5) and R(2.0).

The finite sample sizes of the tests are estimated under the null hypothesis of IID normal returns. We generate an initial sample of returns and conduct the tests using five percent nominal critical values. We then add adjacent returns and recompute the tests. The aggregation and computation of the tests are repeated until the returns have been aggregated across 60 observations. In their Monte Carlo study, CD only considered aggregation up to 20 observations. Our results may be interpreted as sequentially aggregating daily returns up to quarterly returns assuming a five day business week. To make our results directly comparable to CD, we use their initial sample size of 6350. The procedure is repeated 25,000 times. The finite sample size is estimated at a given level of aggregation as the proportion of rejections among the Monte Carlo replications.

We consider different variance models for power comparisons including GARCH(1,1)-n, GARCH(1,1)-t, ARCH(1)-n, ARCH(1)-t, and the threshold ARCH model. Weak GARCH models are closed under temporal aggregation for flow variables while the weak ARCH models are not (Drost and Nijman, 1993). This gives us the opportunity to compare the power of two tests against different commonly used variance models under aggregation.

For the power comparisons, we first consider daily GARCH(1,1)-n and GARCH(1,1)-t models. To once again make our results comparable to those of CD, we use an initial GARCH process similar to theirs. For the Student-t innovation sequence, we use $\varepsilon_t = \sqrt{3/5} \cdot t(5)$, where $t(5)$ is a Student's t random variable with five degrees of freedom. For the conditional variance parameters, we use $\alpha_0 = 1.0$,

$\alpha_1 = .05$ and $\beta_1 = .93$[2]. This is a highly persistent GARCH process with heavy tails and is representative of processes typically found for daily asset returns. As the returns are aggregated, the power of the tests will be reduced by two phenomena. First, for a fixed initial daily return sample, the available long-horizon sample declines as the length of the horizon increases. And second, the persistence in volatility declines in magnitude. To illustrate this second phenomenon, we compute the sequence of GARCH(1,1) coefficients for the aggregated returns implied by our initial model using the Drost and Nijman (1993) temporal aggregation formulas for flows. The sequence of coefficients and their sum are shown in Figure 1. The value of α_1 initially rises, peaks at the 16 day return and then very gradually declines. The aggregated value of β_1 monotonically declines with the length of the horizon. The sum of the two coefficients is an overall measure of volatility persistence. This sum also monotonically declines, indicating that variance forecastability declines with increasing return horizon. For completeness we also consider ARCH(1)-n and ARCH(1)-t processes.

As a second alternative, we consider the QTARCH process described in the previous section. To select a particular data generating process, we would like to specify the multiple m and choose plausible transition probabilities π_{11} and π_{01}. Then, in principle, (1)-(4) could be used to solve for the parameters γ_0, γ_1 and δ.

---

[2] CD actually used $\alpha_1 = .06$. At their parameter values, the unconditional kurtosis of the return y_t does not exist by formula (18) of Drost and Nijman (1993). Subsequently, this would imply that the coefficients in the GARCH model for the aggregated returns do not exist. Because we would like to compute these GARCH coefficients, we take $\alpha_1 = .05$ to have a finite return kurtosis.

Unfortunately, for an arbitrary choice of m, π_{11} and π_{01}, a solution for γ_0, γ_1 and δ may not exist. To demonstrate this, let

$$a_{11} = F^{-1}\left(\frac{\pi_{11}+1}{2}\right) \text{ and } a_{01} = F^{-1}\left(\frac{\pi_{01}+1}{2}\right).$$

Then for a given m and a choice of π_{11} and π_{01}, which determine a_{11}, a_{01} and p, (1)-(4) produce the homogeneous linear system of equations

$$\begin{bmatrix} a_{11}^2 & a_{11}^2 & -1 \\ a_{01}^2 & 0 & -1 \\ m^2 & m^2 p & -1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \delta^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5}$$

In general, the coefficient matrix of this system is not singular. The multiples m and the pairs of transition probabilities π_{11} and π_{01} that allow for solutions to (5) satisfy the determinantal equation

$$(a_{11}^2 - a_{01}^2)m^2 p + a_{11}^2(a_{01}^2 - m^2) = 0. \tag{6}$$

To proceed, for both $m = 1.5$ and $m = 2.0$, we step over values for π_{11} in increments of .01 and numerically solve (6) for π_{01}. For $m = 1.5$ we choose the pair of transition probabilities that produced p closest to .87. For $m = 2.0$ we choose the pair of transition probabilities that produced p closest to .95. The probabilities .87 and .95 are the approximate unconditional coverage probabilities for the forecast intervals $[-1.5\sigma_y, 1.5\sigma_y]$ and $[-2.0\sigma_y, 2.0\sigma_y]$. For $m = 1.5$, the algorithm produces $\pi_{11} = .93$, $\pi_{01} = .473$, and $p = .871$. While for $m = 2.0$, it produces $\pi_{11} = .98$ $\pi_{01} = .39$ and $p = .951$. The resulting transition probabilities seem quite reasonable. The wider forecast interval, with $m = 2.0$, requires more persistence in the volatility to insure that (4) holds.

In Figure 2, we present the estimated sizes for the tests without the continuity correction for the runs test. The nominal size of R(1.5) begins around 0.05, and is estimated to be 0.067 at the 60 day horizon. While the nominal size of R(2) is estimated to be around .05 at one day horizon and jumps to 0.07 after 10 days of aggregation and reaches to .1 after 60 days of aggregation. For the initial sample, the sizes of both LM(1) and LM(5) are close to their nominal sizes of .05. As the order of aggregation increases, the sizes of the LM tests very gradually decline. At a 60 day horizon, the sizes of LM(1) and LM(5) are both estimated to be .036. Size properties of the LM tests are quite reliable even for highly aggregated returns. However, the size properties of the two runs test become highly oversized as the order of aggregation increases. This shows that due to size distortions the power comparisons will be misleading. The results on the estimated size of the runs tests suggest that there is need for size correction.

In Figure 3, we present the estimated sizes for the tests with the continuity correction. For the initial sample, R(1.5) and R(2.0) both also have estimated sizes very close to the nominal five percent. As the order of aggregation increases, the size of R(1.5) falls modestly below the nominal size. The size is estimated to be .045 at the 60 day horizon. The size of R(2.0) also initially declines below the nominal .05 level, but abruptly begins to increase at the 47 day horizon. At the 60 day horizon, the size of R(2.0) is estimated to be .057[3]. The results indicate that the size properties of all four of the tests are quite reliable even for highly aggregated returns. LM(1), LM(5)

[3] The reversal in the size of R(2.0) in Figure 2 is likely due to the failure of the continuity correction to maintain the normal approximation in the extreme tail of the runs distribution for the small samples.

and R(1.5) are all slightly undersized and R(2.0) is slightly oversized. The estimated size of R(1.5) appears to be closest to the nominal size.

In Figure 4, we present the estimated power of the tests for the GARCH(1,1)-n process. At short horizons, LM(5) has superior power. The power of LM(5) begins to decrease as the volatility persistence falls with increasing aggregation horizon. R(1.5) out performs R(2.0) up to 20 day return horizon. After 25 return horizons, all test have almost similar powers. The results suggest the use of LM test for high order GARCH processes for moderate-horizon returns. LM tests clearly have superior power for moderate horizons if the true daily process is GARCH(1,1)-n. Recall, CD only considered the performance of their test at a maximum horizon of 20 days. For the horizon CD considers their test seem to have less power compared to the LM tests when the DGP is GARCH(1,1)-n. Our results suggest that when the true daily process is GARCH, for long horizons, the power of the two runs tests are at least as good as the LM tests.

In Figure 5, we present the estimated power of the tests for the GARCH(1,1)-t(5) process. At short horizons, LM(5) clearly has superior power. As the return horizon increases, and the volatility persistence falls, the power of LM(1) increases relative to LM(5) and has higher power for horizons beyond 31 days. Consistent with the findings of CD, R(1.5) has higher power than R(2.0) at all horizons. Strikingly, the power of R(1.5) is almost identical to that of LM(1) over all horizons. The results present two practical conclusions for using the tests. First, the LM test for high order GARCH should still be used for moderate-horizon returns because it has superior power for moderate horizons if the true daily process is GARCH. Second, for longer horizon returns, the runs tests perform as well as the LM tests when the true daily

process is GARCH, suggesting that the runs test may provide a good omnibus test for variance forecastabilty at very long horizons. Recall, CD only considered the performance of their test at a maximum horizon of 20 days. Our results suggest that when the true daily process is heavy tailed GARCH, the advantages of their runs test may be realized at much longer horizons.

In Figure 6 and 7 we present the estimated power of the tests for the ARCH-n and ARCH-t processes. For the conditional variance parameters, we use $\alpha_0 = 1.0$, $\alpha_1 = .8$. It is evident from Figure 6 that the power of all the tests decline very rapidly as the order of aggregation increases. After 10 day horizon all the tests have power less then .2. All the tests have similar power less then .1 after 20 day of aggregation. Up to 10 days of aggregation two runs test have superior power to the LM tests. LM(5) has the lowest power as expected. Between 10-20 days of return horizon LM(1), and the R(1.5) runs have very similar power while R(2.0) slightly out performs them. The results suggest that when the daily DGP is ARCH-n, for moderate horizon of aggregation the two runs test have superior power to the LM tests. This result is consistent with the Drost and Nijman (1993) which denotes that the weak ARCH processes are not closed under temporal aggregation for flow variables. As the order of aggregation increases, the LM tests lose their advantage of optimality against the ARCH model since the aggregated model is likely not to be an ARCH. In Figure 7, we present the estimated power of the tests for the ARCH-t process. For the innovation sequence, we use $\varepsilon_t = \sqrt{3/5} \cdot t(5)$, where $t(5)$ is a Student's t random variable with five degrees of freedom. The results are very similar to that of the normal case except for the fact that the decline in power of the tests is faster.

In Figures 8 and 9, we present the estimated power of the tests when the data generating process is the QTARCH model with the threshold δ equal to $1.5\sigma_y$ and $2.0\sigma_y$. For the QTARCH models, the power of all of the tests falls much more rapidly than they do for the GARCH process. Therefore, we only present the powers out to the 20 day return at which point the power is approximately equal to the size of the test. It is evident in both figures that the power of the runs tests can be considerably higher than that of the LM tests when the return horizon is in the range of two to ten days. The results indicate that there are plausible time-varying conditional variance models for which the runs test has superior power to the LM test for GARCH. In Figure 4, when the threshold is $1.5\sigma_y$, the powers of R(1.5) and R(2.0) are very similar. In Figure 5, when the threshold is $2.0\sigma_y$, the power of R(2.0) is considerably higher than that of R(1.5). Therefore, we cannot conclude that R(1.5) uniformly has the highest power. The ranking of the power of the runs tests based on different multiples m depends on the particular alternative being considered. In practical applications, this suggests that the runs test should be computed for a variety of multiples of the return standard deviation.

Conclusions

Our Monte Carlo study provides evidence that the runs test is useful for detecting volatility forecastability in long-horizon returns. It should not be used, however, to the total exclusion of the widely used LM test for GARCH. When the true daily return process is a persistent GARCH process, a high order LM test for GARCH has better

power than the runs test for return horizons out to about six weeks. Beyond a six week horizon, the runs test basically proxies a low order LM test. We have also demonstrated that there do exist processes, ARCH-n, ARCH-t and in particular a QTARCH process, for which the runs test does have better power than the LM test for GARCH when returns are considered at a one to two week horizon. Therefore, the runs test does appear to be a good omnibus test for volatility forecastability in aggregated returns. It has reasonably good power when the true process is GARCH, but superior power for other volatility processes.

References

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307-327.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay, 1997, The Econometrics of Financial Markets, (Princeton University Press, New Jersey).

Christoffersen, P. F. and F. X. Diebold, 2000, How relevant is volatility forecasting for financial risk management?, The Review of Economics and Statistics 82, 12-22.

Drost, F. C. and T. E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica 61, 909-927.

Engle, R. F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987-1008.

Glosten, L. R., R. Jaganathan and D. Runkle, 1993, On the relation between the expected value and the volatility of the normal excess return on stocks, Journal of Finance 48, 1779-1801.

Gourieroux, C. and A. Monfort, 1992, Qualitative threshold ARCH models, Journal of Econometrics 52, 159-199.

Lee, J. H., 1991, A Lagrange multiplier test for GARCH models, Economic Letters 37, 265-271.

Lehmann, E. L., 1986, Testing Statistical Hypothese, 2nd edition (John Wiley, New York).

Rabemananjara, R. and J. M. Zakoian, 1993, Threshold ARCH models and asymmetries in volatility, Journal of Applied Econometric 8, 31-49.

Figure 1. Aggregated GARCH(1,1)-t coefficients.

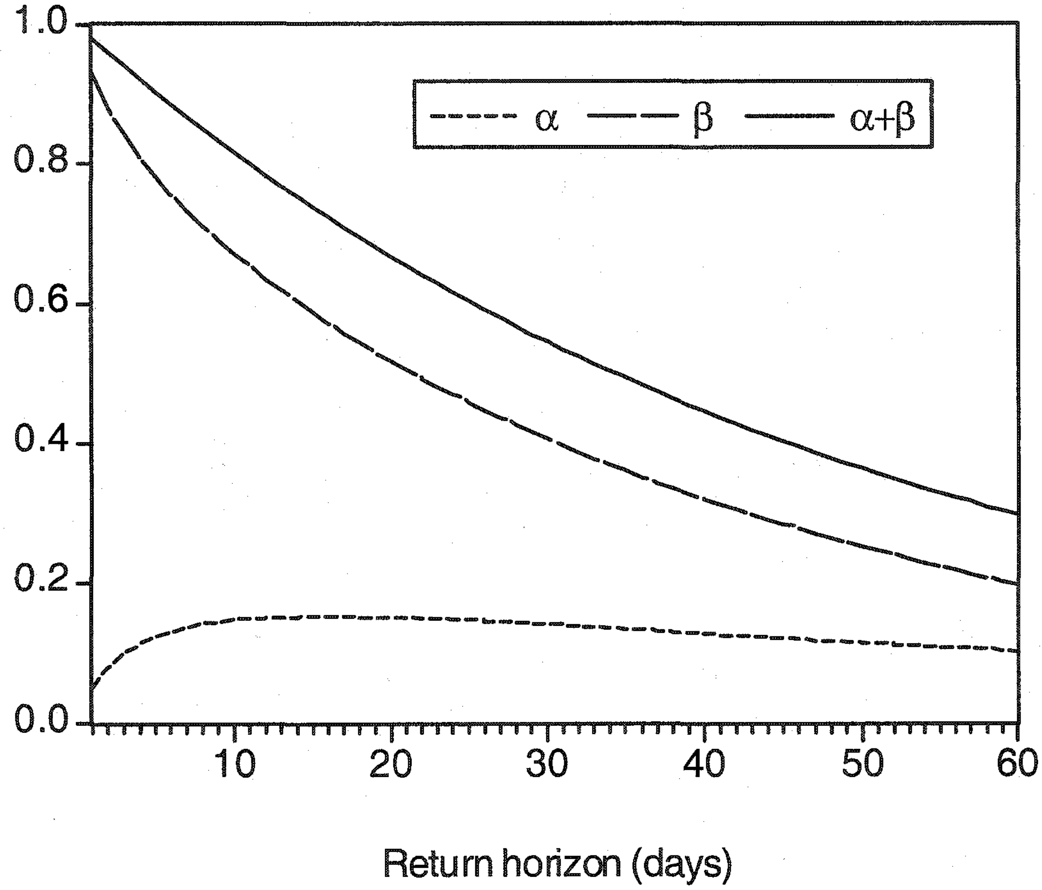


Figure 2. Estimated size of the tests, no continuity correction.

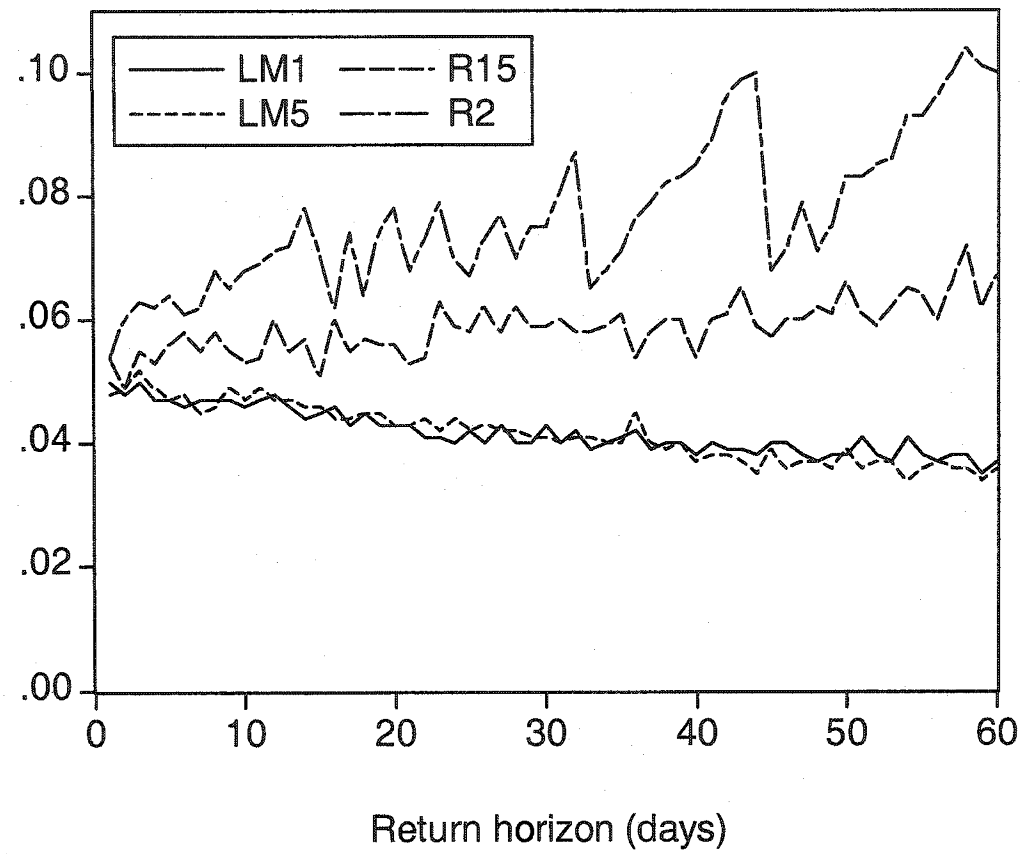


Figure 3. Estimated size of the tests, with continuity correction.

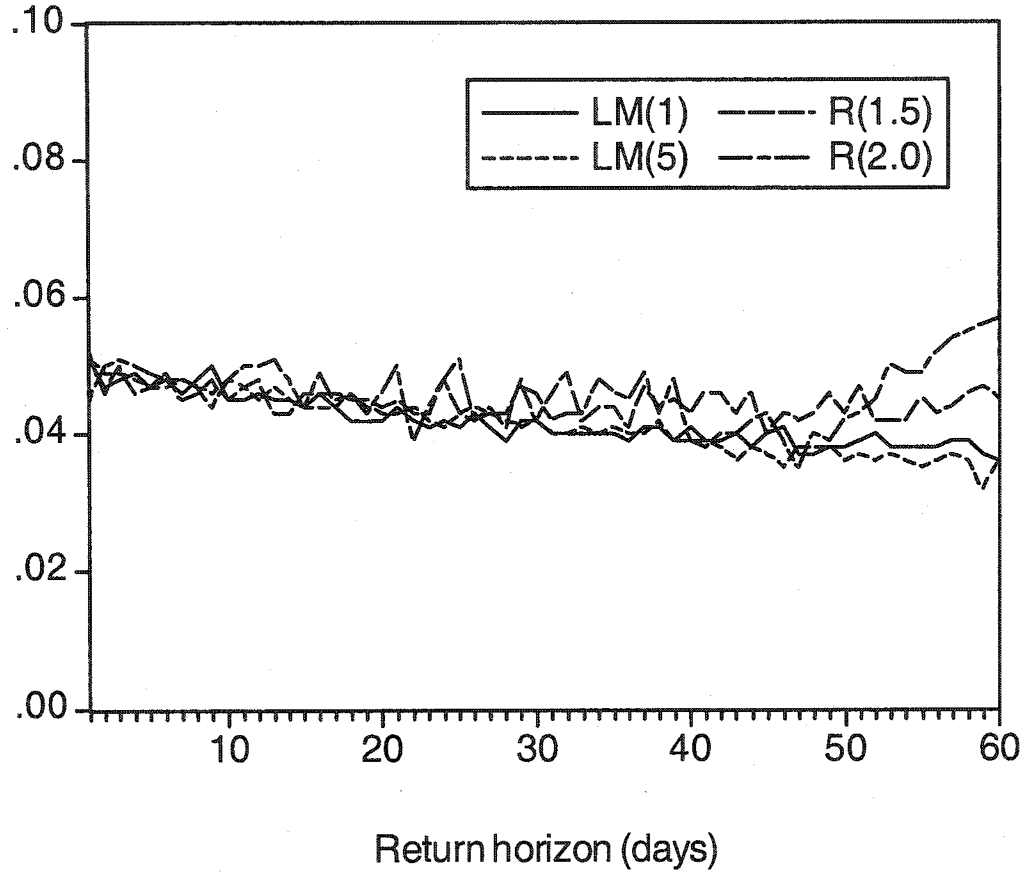


Figure 4. Estimated power against GARCH(1,1)-n.

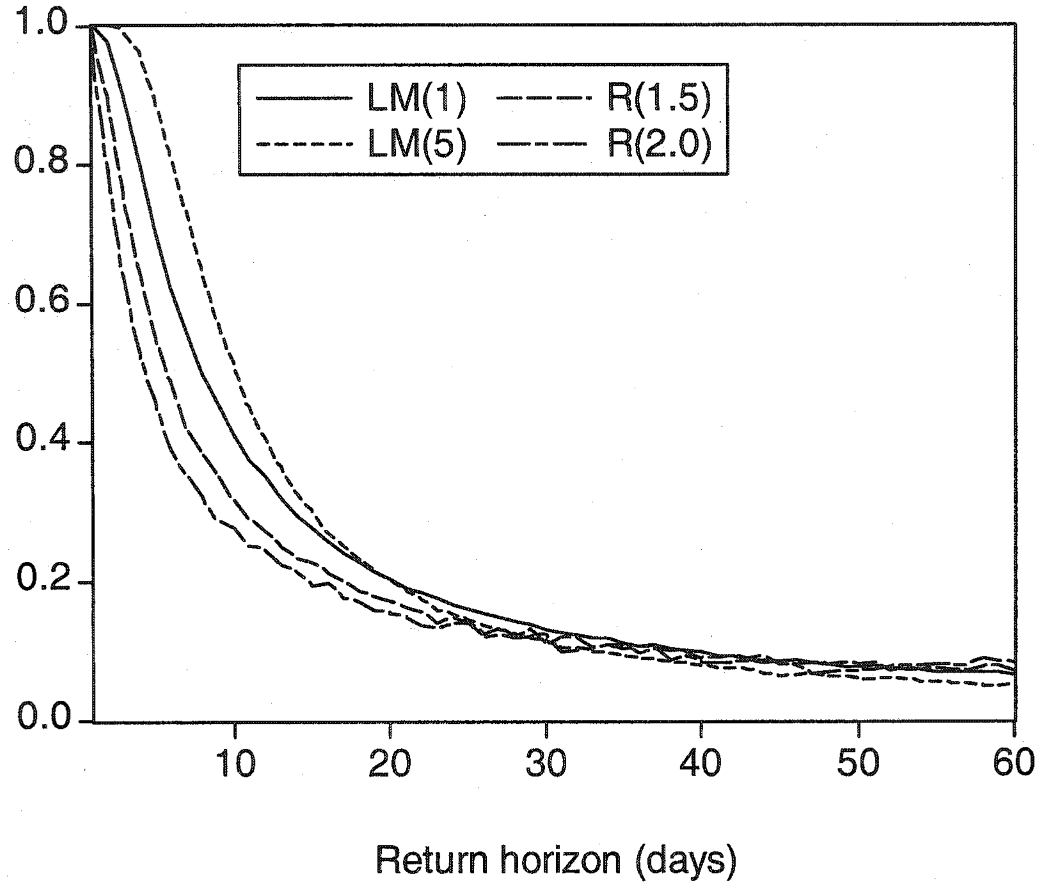


Figure 5. Estimated power against GARCH(1,1)-t.

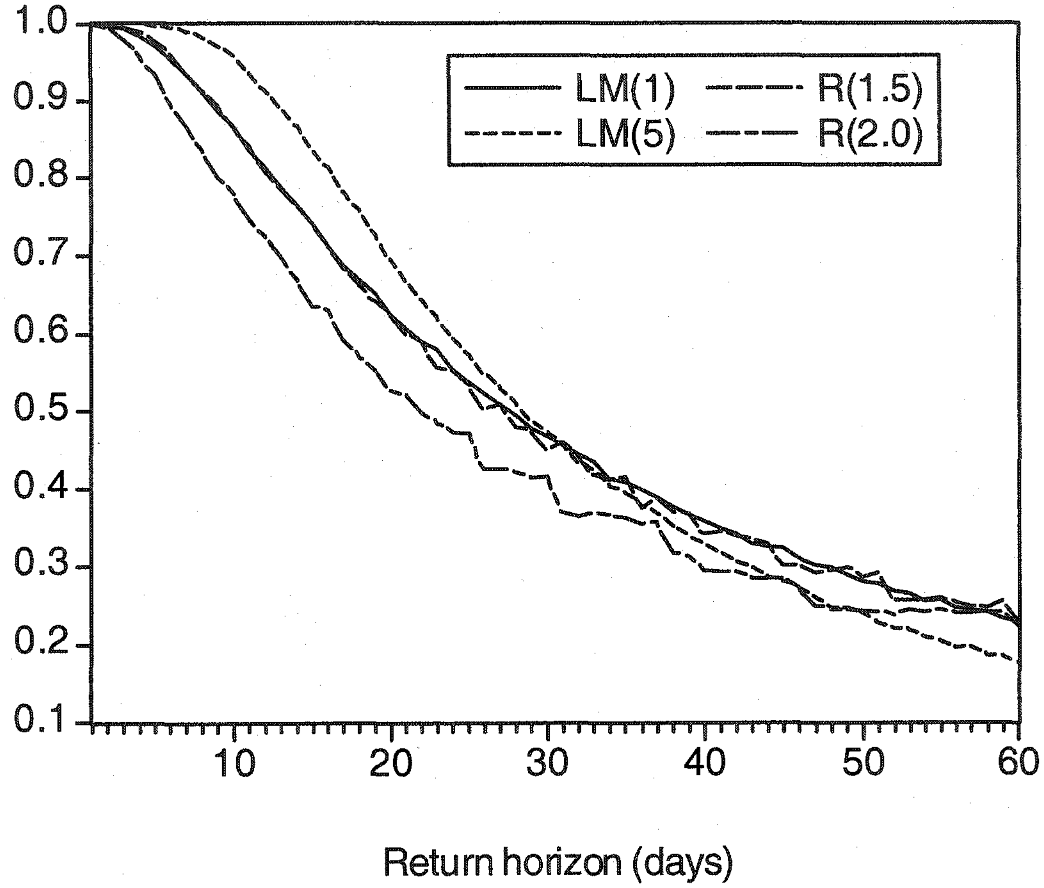


Figure 6. Estimated power against ARCH(1)-n.

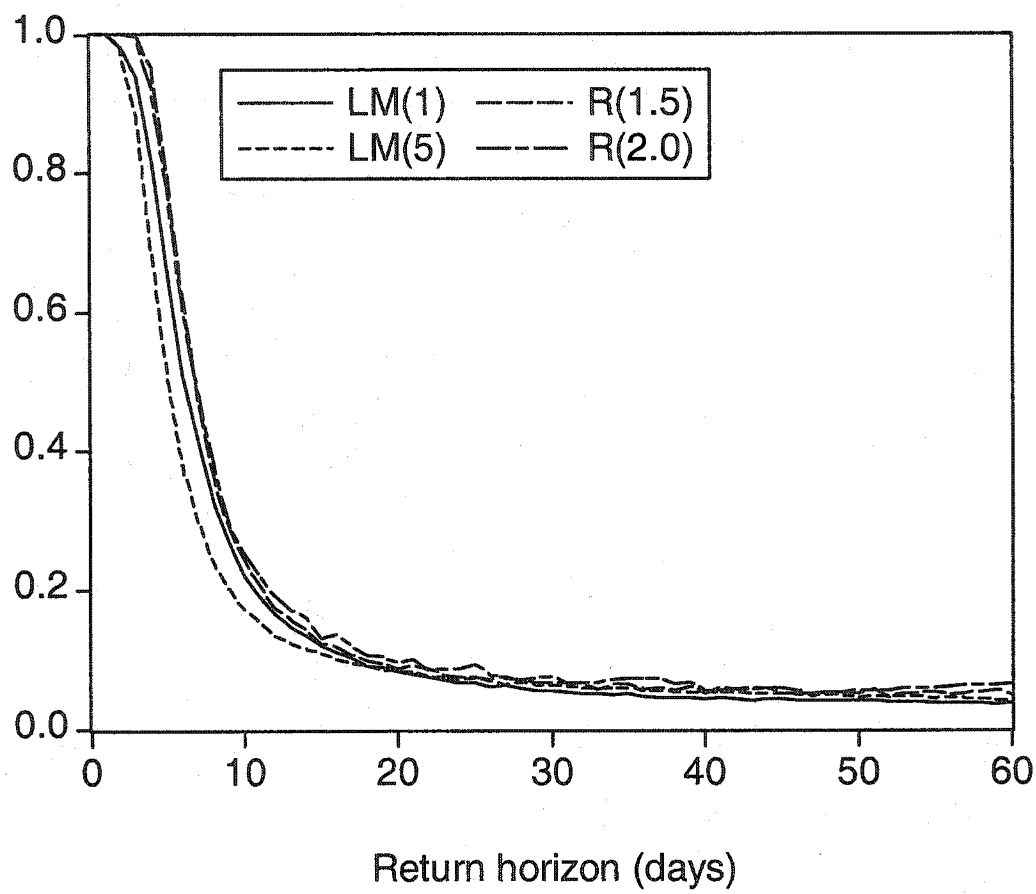


Figure 7. Estimated power against ARCH(1)-t.

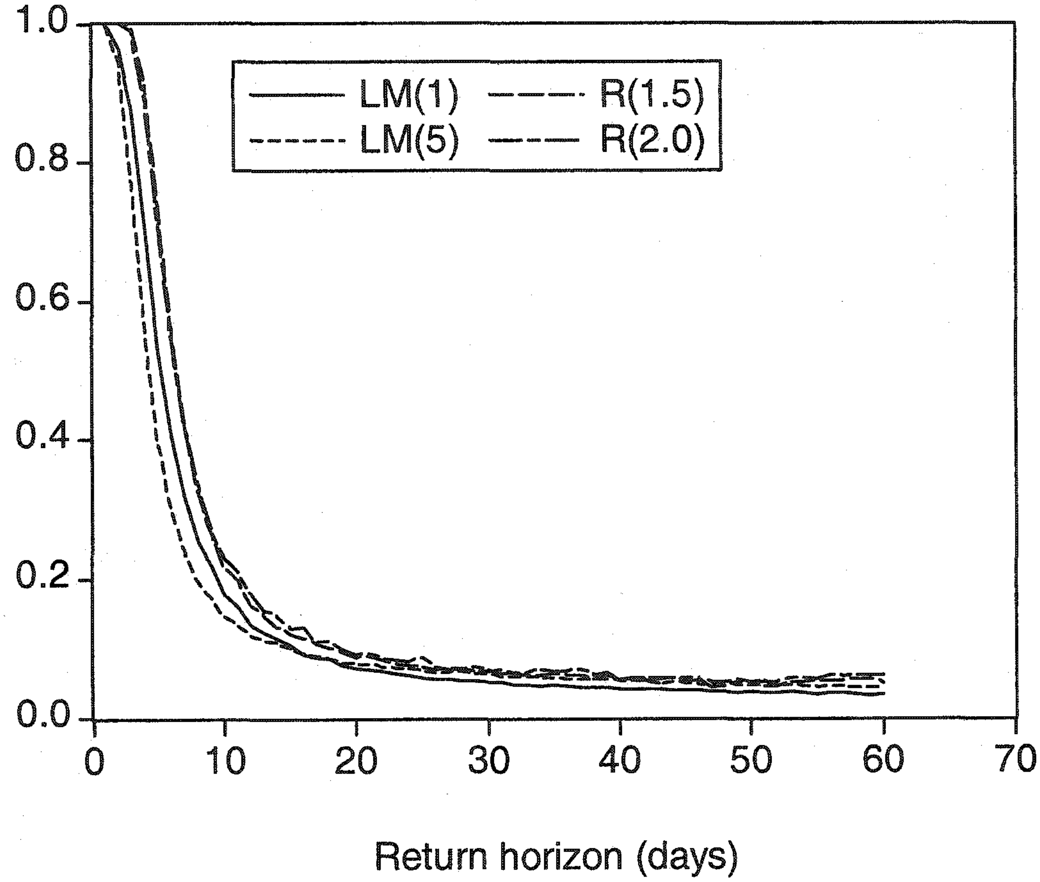


Figure 8. Estimated power against QTARCH, $\delta = 1.5\sigma$.

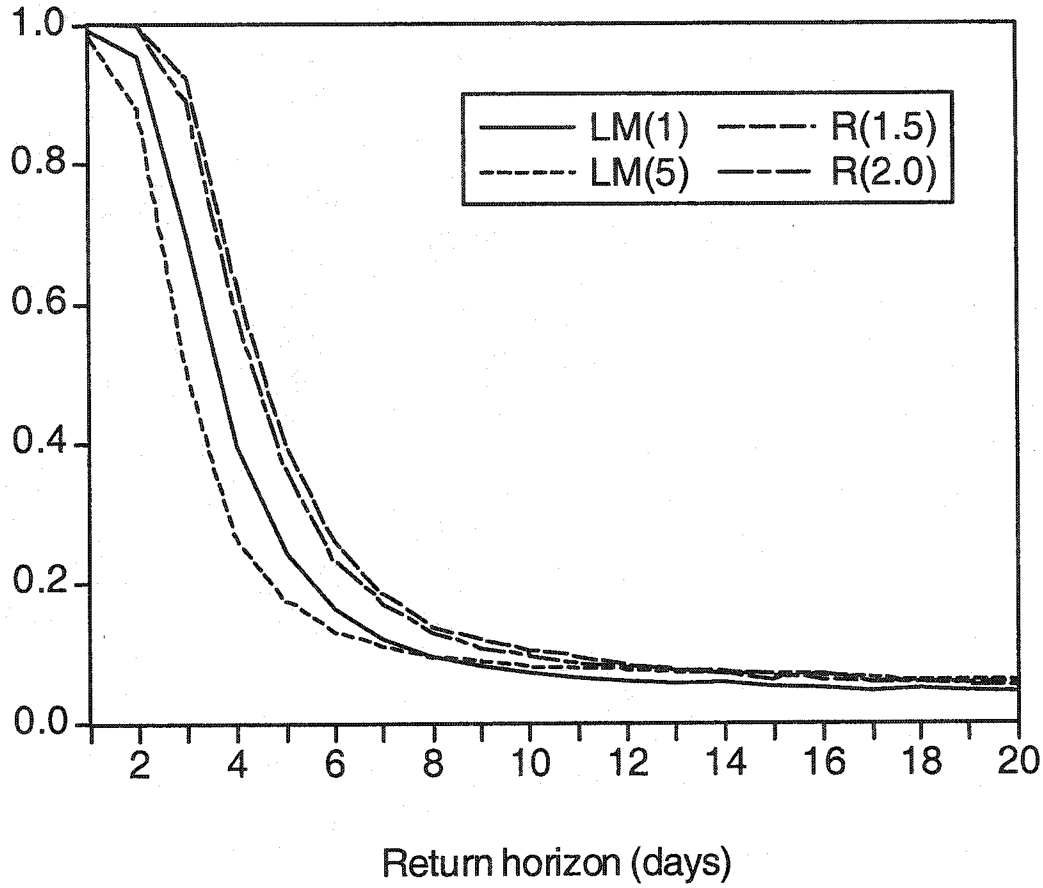
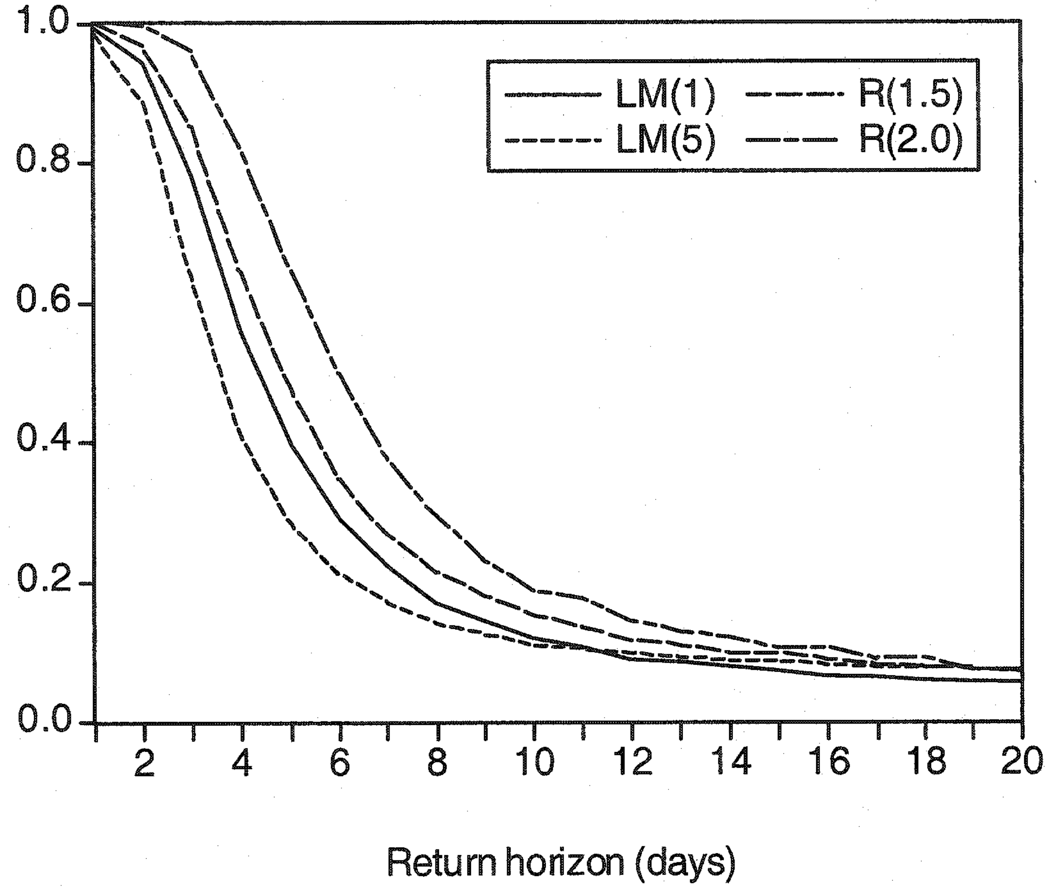


Figure 9. Estimated power against QTARCH, $\delta = 2.0\sigma$.



CHAPTER II

SAMPLING PROPERTIES OF CRITERIA FOR EVALUATING GARCH FORECASTS OF ASSET RETURN VOLATILITY

Introduction

Understanding the relationship between risk and expected returns is essential in finance. Most of the fundamental theories in finance, like the CAPM and option pricing, use this relationship. Financial decisions about hedging and trading strategies also require return volatility. This makes return volatility forecasting an important part of financial decisions. However, modeling and forecasting volatility has became a popular topic in financial economics only recently. After the studies of Mandelbrot (1963) and Fama (1965), it became a stylized fact that financial returns exhibit volatility clustering in which large movements in prices tend to be followed by large movements, producing serial correlation in squared returns. This gives the idea that current and past volatility can be used to predict future volatility. The most commonly used models to capture volatility clustering in financial asset returns are the ARCH model developed by Engle (1982) and the GARCH model by Bollerslev (1986).

Although GARCH accurately characterize volatility clustering by giving very significant in sample parameter estimates, they have been frequently found to poorly forecast volatility out of sample. Using the forecast evaluation procedure of Mincer and Zarnowitz (1969), a regression of realized volatility on the estimated conditional

variance often produces a very small R^2. For example, West and Cho (1995) found R^2's ranging from 0.001 to 0.045 for GARCH models for 5 different U.S. dollar exchange rates. Day and Lewis (1992) found a R^2 value of 0.039 using a GARCH (1,1) model for weekly returns on a U.S. stock index. Pagan and Schwert (1990) used historical U.S. monthly stock returns to construct out-of-sample forecasts over two different periods and found R^2's of .075 and .078 using a GARCH (1,2) model. Similar results can be found in more recent papers cited in Granger and Poon's (2003) comprehensive survey.

The fact that ARCH/GARCH models give very significant in sample parameter estimates, but explain little variability in ex-post squared returns suggested that these models might suffer from misspecification. From finance perspective, the fact that volatility models are useful if they are able to predict volatility, gave rise to the idea that maybe GARCH models are not of practical use in forecasting volatility. However, as pointed out by Andersen and Bollerslev (1998), unless you have a theoretically suggested population value for the evaluation criteria under the correctly specified model, evaluations based on the empirical results can be misleading. In fact, Andersen and Bollorslev (1998) showed that low R^2 values from the Mincer and Zarnowitz regression are not in contradiction with a corrrectly specified GARCH (1,1) model.

In this paper we derive population moments of criteria commonly used point forecast evaluate the accuracy of volatility forecasts when the true data generating process is GARCH(1,1). We also analyze the properties of R^2's and coefficients from regressions used for volatility forecast evaluation. We then study their sampling properties by Monte Carlo simulation. We consider scenarios when the returns are

conditionally normally and conditionally t-distributed. Our results provide several practical recommendations for forecasting volatility.

The moments of criteria for evaluating volatility forecast accuracy

In this section, we derive the population moments of criteria that are commonly used to evaluate GARCH volatility forecasts. We state conditions for the existence of the population moments. The moment existence conditions provide an indication for when the criterion may perform poorly. We also establish relationships between the magnitude of the population moments. We focus exclusively on the GARCH(1,1) model because it is used in the majority of GARCH forecasting applications.

A GARCH (1,1) model for the return on a financial asset, r_t, can be written

$$r_t = \sqrt{h_t} \cdot z_t, \; z_t \sim IID(0,1) \qquad (1)$$
$$h_t = \gamma + \alpha r_{t-1}^2 + \beta h_{t-1}$$

where $\gamma > 0$, $\alpha \geq 0$ and $\beta \geq 0$. We assume z_t is an *IID* innovation with finite first and second moments. To simplify the derivations, we assume that the returns have a conditional mean of zero. This assumption does not alter the expressions for the population moments, nor their existence conditions. The return r_t is weakly stationary if its variance is finite. This will be the case if $\alpha + \beta < 1$. Since $E(r_t^2 \mid \Phi_{t-1}) = h_t$, where Φ_{t-1} is the information set available at time t-1, the conditional variance h_t is the minimum mean square error predictor of the realized volatility r_t^2.

The most widely used criteria to evaluate the accuracy of forecasts is mean squared error (MSE). MSE assumes the forecaster faces a quadratic loss. The MSE for a GARCH volatility forecast is

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(r_t^2 - h_t\right)^2.$$

Obviously, MSE decreases as the forecast accuracy increases. In an actual forecasting application, the unobserved h_t would be replaced by a sample estimate $\hat{h}_t$ as in Andersen and Bollerslev (1998), we abstract from parameter uncertainty, we analyze the criteria assuming the parameters in (1) are known. Criterion such as the MSE are usually reported simply as descriptive statistics that summarize the accuracy of forecasts in a given data set. They can also be viewed as sample estimates of a population moment that characterizes the inherent forecastablility of the data generating process. We now give the population moment for the MSE of a GARCH(1,1) process.

Theorem 1: For the GARCH (1,1) process given in (1), the necessary and sufficient conditions for the existence of the population MSE are that $E(z_t^4) = \kappa < \infty$ *and* $\alpha^2\kappa + \beta^2 + 2\alpha\beta < 1$. *When it exists, the population MSE is*

$$MSE_p \equiv E[(r_t^2 - h_t)^2] = (\kappa - 1)\frac{\psi^2(1+\alpha+\beta)}{(1-\alpha-\beta)[1-(\alpha^2\kappa+\beta^2+2\alpha\beta)]}.$$

Proof: See appendix.

The MSE_P increases with the magnitude of ψ and κ, the parameters which respectively determine the unconditional variance and kurtosis of returns. Theorem 1 indicates that the condition $\alpha^2\kappa + \beta^2 + 2\alpha\beta < 1$ that must hold for MSE_p to be finite

is the same as the existance condition for the kurtosis of the returns [see Andersen and Bollerslev (1998)] . This condition is frequently violated for actual returns. When the MSE_P is not defined, interpretation of the sample *MSE* is difficult. The sample *MSE* provides a measure of forecast accuracy for the given realized sample, but conveys little information about how well the model will forecast in the future.

The *MSE* is frequently criticized as a measure of forecast accuracy when applied to volatility forecasts. For a process in which the conditional variance is time-varying, the accuracy of a particular forecast should be measured relative to the inherent uncertainty in predicting that particular observation. Bollerslev, Diebold and Engle (1994) and Bollerslev and Ghysels (1996) suggested that the accuracy of volatility forecasts should instead be evaluated using a heteroscedasticity adjusted *MSE* (*HMSE*). The *HMSE* is defined as

$$HMSE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_t^2 - h_t}{h_t}\right)^2 .$$

We now state the population value of HMSE.

Theorem 2: For the GARCH (1,1) process given by (1), the necessary and sufficient conditions for existence of the population HMSE, is $E(z_t^4) = \kappa < \infty$. *When it exists*

$$HMSE_p \equiv E\left[\left(\frac{r_t^2 - h_t}{h_t}\right)^2\right] = \kappa - 1 .$$

Proof: See appendix.

Unlike MSE_p, the existence of $HMSE_p$ is not constrained by the necessary condition for the fourth moment of returns to exist. In fact, it does not even require the

variance to be finite. It is defined for an IGARCH model where $\alpha+\beta=1$. Many empirical studies find IGARCH characteristics in financial returns. This suggests that $HMSE$ might be a better choice than MSE, to evaluate GARCH forecasts for of return volatility.

From Theorems (1) and (2), an immediate relation between MSE_p and $HMSE_p$ is

$$MSE_p = \frac{\psi^2(1+\alpha+\beta)}{(1-\alpha-\beta)[1-(\alpha^2\kappa+\beta^2+2\alpha\beta)]} \cdot HMSE_p .$$

This equation reveals that for a GARCH(1,1) process it is impossible to make any statement about the relative magnitudes of MSE_p and $HMSE_p$ for uniformly all parameter values. The parameter ψ, which determines the magnitude of the unconditional variance of the return, is only constrained by $\psi > 0$. Since $HMSE_p$ does not depend on ψ, for a sufficiently large ψ we will have $MSE_p > HMSE_p$. Similarly, for sufficiently small ψ, $MSE_p < HMSE_p$. An inequality between $HMSE_p$ and MSE_p can be derived as

$$MSE_p > \frac{\mathrm{var}(r_t^2)}{\kappa} \cdot HMSE_p ,$$

suggesting that the magnitude of the two measures depend on the ratio between the variance of the square of the returns and the fourth moment of the return innovation. If this ratio is greater than one, MSE_p will be larger than $HMSE_p$.

Another popular criterion for evaluating the accuracy of volatility forecasts forecast is the mean absolute error (MAE). The MAE of a GARCH volatility forecast is

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|r_t^2 - h_t\right|.$$

The population value of MAE is given in the next theorem.

Theorem 3: For the GARCH (1,1) process given by (1), the necessary and sufficient conditions for the existence of the population value for MAE is $\alpha + \beta < 1$ *and* $E(|z_t^2 - 1|) < \infty$. *When it exists,*

$$MAE_p \equiv E(|r_t^2 - h_t|) = E(|z_t^2 - 1|)\frac{\psi}{1-\alpha-\beta}.$$

Proof: See appendix.

Theorem 3 indicates that MAE_P requires the first absolute moment of the squared innovation to be finite and that the unconditional variance of the return to exist. Since $\mathrm{var}(r_t) = \psi(1-\alpha-\beta)^{-1}$ [see Bollerslev (1986, Theorem 2)], MAE_p is simply the return variance rescaled by this absolute moment. It does not require to fourth moment of the innovation to exist. If the innovation is normal, $E(|z_t^2 - 1|) = 2\sqrt{2/e\pi}$. We are not aware of an analytic expression for this moment for the t-distribution.

A commonly used forecast evaluation criteria in many studies is the regression procedure proposed by Mincer and Zarnowitz (1969). The Mincer-Zarnowitz (MZ) regression for the return volatility can be written as:

$$r_t^2 = a + bh_t + v_t, \qquad (2)$$

where v_t is a zero mean error term. If the GARCH model is correctly specified, the population value of a is zero and the population value of b is one. The regression can be estimated and t-tests of the hypotheses that $a=0$ and $b=1$ provide a test that h_t is an "efficient" predictor of r_t^2. The R^2 from this regression measures how well the model forecasts. The regression in (2) is typically estimated using a small portion of the data retained exclusively for forecast evaluation with h_t replaced by estimates.

In spite of the previously cited papers, which find that the GARCH poorly predicts volatility, Andersen and Bollerslev (1998) argue that correctly specified GARCH models and low R^2 values from the MZ regression are not in contradiction. If $\kappa\alpha^2+\beta^2+2\alpha\beta<1$, the fourth moment of the return r_t is finite. Assuming this condition is met, Andersen and Bollerslev showed that the population R^2 from this regression is

$$R_p^2=\frac{\operatorname{var}(h_t)}{\operatorname{var}(r_t^2)}=\frac{\alpha^2}{1-\beta^2-2\alpha\beta}<\frac{1}{\kappa} \tag{3}$$

when the true data generating process is GARCH (1,1). This bound can be quite small. For example, if the innovation is normal, $\kappa=3$ and the population R^2 is bounded by 1/3. They explained the reason for getting low R^2 values based on the fact that, although r_t^2 is an unbiased estimator for h_t, it might give noisy results due to the innovation z_t.

As an alternative to the MZ regression, several authors have suggested a log version of the regression. Schwert and Pagan (1990) and Engle and Patton (2001) recommend using the R^2 from a regression of $\log(r_t^2)$ on $\log(h_t)$ because it measures

the relative forecast accuracy. The log MZ regression is similar to (2), however, the population value of a is not zero. From (1), $r_t^2 = h_t \cdot z_t^2$ and taking logs we can write

$$\begin{aligned} \log(r_t^2) &= E[\log(z_t^2)] + \log(h_t) + \log(z_t^2) - E[\log(z_t^2)] \\ &= a + b\log(h_t) + u_t, \end{aligned} \tag{4}$$

where $a = E[\log(z_t^2)]$, $b = 1$ and $u_t = \log(z_t^2) - E[\log(z_t^2)]$. The value of a does not depend on the parameters in h_t, but only the parameters in the density of z_t. The log MZ regression can be used to test the efficiency of the forecast by testing whether a is equal to an estimate of $E[\log(z_t^2)]$.

Illustrative GARCH models

To provide empirical examples of the above criteria, we estimate GARCH models for six series that are representative of the type of returns for which volatility forecast are often constructed. We estimate GARCH models for the exchange rate between the U.S. dollar and the Canadian Dollar, the Japanese Yen and the British Pound. We also estimate models for three major U.S. market indexes: the Dow Jones Industrial Average, the NASDAQ Composite and the S&P 500. The data are daily from July 3, 1995 to July 3, 2001. Data from July 3, 1995 to 30 June 2000 is used for estimation and the last year is retained for evaluating the out of sample volatility forecasts.

We estimate GARCH(1,1) models for each return series. Although the efficient market hypothesis suggests that returns should be serially uncorrelated, we check for possible serial correlation in the returns. With the exception of Japanese Yen, which

we identify as an AR(3) process, we find that all the return series are serially uncorrelated. Table 1 shows the estimated GARCH(1,1) models and diagnostic statistics for the standardized residuals. The estimates for the conditional variance parameters are all significant. The skewness coefficients of the standardized residuals show that they are slightly negatively skewed. The kurtosis coefficients are between four and six, suggesting that the data do not have normal distributions. The Jarque-Bera statistics are all highly significant. The Ljung-Box Portmanteau tests for the serial correlation in the standardized and squared standardized residuals up to 10 lags indicate that the residuals are white noise. This suggests that we can characterize the inter-daily volatility dependencies by a GARCH (1,1) model. When we look at the volatility persistence as indicated by $\alpha+\beta$, the parameter estimates sum close to unity for Japan, the Dow Jones, and the NASDAQ, and almost precisely to unity for Canada and the S&P 500. This indicates that the return volatility may be the IGARCH(1,1) model of Engle and Bollerslev (1986). This high persistence suggests that market volatility is predictable.

We use daily data from July 3, 2000 to July 3, 2001, for one day ahead out of sample forecasts. Table 2 presents the calculated accuracy criteria. We also report the theoretical population moments evaluated at the estimated parameters. It is important to see the performance of the criteria with respect to its population value, which is derived under the null that the true data generating process is GARCH (1,1). *MSE* results show that for Canadian Dollar, and Japanese Yen, the empirical values are less then those of the population values. For the three stock indexes, MSE_p is not defined due to the fact that the conditions for the existence of the fourth moment of the return

does not hold. The results for *HMSE* show that the *HMSE* values are close in relative magnitude to their population values. The sample *HMSE* are, with the exception of the Japanese Yen, consistently greater than their population values. There is no systematic pattern between the *MAE*'s and the MAE_p.

The estimated coefficients and the R^2's from the *MZ* and *log MZ* are also reported in Table 2. The reported standard errors are White robust standard errors. If the forecasts are efficient, $a = 0$ and $b = 1$. For two of the six series, the Japanese Yen and the NASDAQ, the hypothesis $a = 0$ is rejected at the five percent level using a two sided test. At the ten percent level, the hypothesis is rejected for the Canadian Dollar and is very close to being rejected for the U.K. Pound and the S&P 500. The hypothesis $b = 1$ is rejected for the Japanese Yen and the NASDAQ at the 5 percent level. For the three exchange rates, the sample R^2's a all very small and substantially less than the estimated population values. The stock indexes also have small sample R^2's and distributions are sufficiently heavy tailed that the R_p^2 are not finite. Overall, the MZ regression suggest that the forecasts are likely not to be efficient and have very low predictive power.

The *log MZ* regressions give some what different results. To test efficiency, we test $a = -1.287$ and $b = 1$. Now the test for the intercept a is significant at the five percent level only for the NASDAQ and is significant at the 10 percent level only for the Japanese Yen. The hypothesis $b = 1$ is rejected at the 10 percent level only for the Canadian Dollar. With the exception of the slope coefficient on the Canadian Dollar, the magnitude of the t-tests for efficiency are uniformly smaller for the *log MZ* regression than for the *MZ* regression. The sample R^2's from the *log MZ* regressions

tend to be smaller than the sample R^2's from the MZ regressions. Over all, the *log* MZ regressions suggest that the volatility forecasts are more efficient but have smaller predictive ability.

In the next section we analyze the sampling properties of the forecast evaluation criteria by a Monte-Carlo study for a conditionally normally distributed and conditionally t-distributed GARCH(1,1) model. This will help us to compare the simulated results under a correctly specified GARCH (1,1) model with the results obtained from real data.

Monte Carlo results

In this section we study the properties of the accuracy criteria sampling properties of point forecast evaluation criteria and the statistics from the efficiency regressions by means of a Monte-Carlo simulation. We consider scenarios when the returns are conditionally normally and conditionally t-distributed. It is well known that financial returns have thicker tails then a normal distribution, and therefore, returns may better be characterized by a t-distribution. We use a conditional t-distribution with seven degrees of freedom, which corresponds to a kurtosis coefficient of five for the return innovation. This kurtosis is representative of the magnitude typically found in empirical work. We simulate data from a GARCH (1,1) model with α equal to 0.05 and β equal to 0.93. These parameter values are representative values that are found in empirical work and similar to the estimated coefficients in the previous section. These values also satisfy the condition required for the existence of

the MSE_p and the R_p^2 when the distribution is normal. This is important because part of the purpose of our analysis is to investigate the relationship between the sampling distribution of MSE and population moment MSE_p given in Theorem 1 and the sampling distribution of the R^2 and the population moment R_p^2 given in (3). To do this, we want to insure that the MSE_p and R_p^2 are defined. When the innovations have a t-distribution with seven degrees of freedom, the condition for the existence of the MSE_p and the R_p^2 is not satisfied. The condition will be violated for any realistic choice of the innovation kurtosis. The results from this case are likely to reflect the performance of the criteria in actual practice.

We use sample sizes 50, 250, and 500. The first two sample sizes would represent a situation in which one is forecasting weekly data or daily data using an out-of-sample period of one year. This is typical in empirical work. A sample of 500 provides a better contrast for how the properties of the sampling distribution are affected by the sample size. In the Monte Carlo study, when we estimate the regressions in (2) and (4), we use the actual value of h_t rather than an estimated value. In empirical work, often the majority of the data is used to estimate the parameters in the conditional variance, and perhaps ten percent of the data is used for the out-of-sample forecasting exercise. It is likely that the variability in the forecast criteria is dominated by the variability in the innovation rather than the variability due to the estimation of parameters. We present kernel density estimates of the sampling distributions of each of the criteria we consider. The density estimates are based on 10,000 Monte Carlo replications.

The first two plots in Figure 10 show the density estimates of the sampling distributions of the *MSE*'s from the GARCH (1,1) model. The population value calculated for the normal case using Theorem 1 is approximately 0.57. As the sample size increases it is evident from the first graph in Figure 1 that the sampling distribution converges very slowly to the population value. For all sample sizes the sampling distributions of the *MSE*'s are skewed to the right. It is evident from the graph that a large portion of the sampling density lies well below the population value. This implies that a forecaster is very likely to find a sample *MSE* that is much smaller than the population value with very high probability. For the case when z_t are t-distributed, the *MSE* values are larger then the normal case. In this case the existence condition for the fourth moment for r_t is violated, thus the population value for *MSE* is not defined. We do observe that the sampling distributions of the *MSE* are even more highly skewed than compared to the conditional normal case.

The estimates of the sampling distributions of the *HMSE*'s from the GARCH(1,1) model are shown in the third and fourth plots in Figure 10. The population value calculated for the normal case using Theorem 2 is two. Although, the sampling distribution for the smallest sample size is skewed to the right, the estimated sampling distributions quickly converge to normality and are centered on the calculated population value. For the case when z_t are t-distributed with seven degrees of freedom and have a kurtosis of five, the population value for *HMSE* using Theorem 2 is four. The results for the conditional t case indicate that the sampling distribution of *HMSE* are slightly skewed. Nevertheless, it converges quickly to the population value. The results suggest that the *HMSE* criterion is more reliable than the

MSE criterion because it is less likely to systematically understate the forecast accuracy.

The estimates of the sampling distributions of the *MAE*'s from the GARCH(1,1) model are shown in the last two pplots in Figure 10. For the normal case $E(|z_t^2 - 1|) = 2\sqrt{2/e\pi}$. Thus the population value for the *MAE* is 0.49. As sample size increases the sampling distribution converges to the population value. However, even for the largest sample size the sampling distribution is slightly skewed to the right. When the return innovation is t-distributed with 7 degrees of freedom, the *MAE* values are larger then the normal case. We simulated $E(|z_t^2 - 1|)$ for the case when z_t is t-distributed with 7 degrees of freedom and found it to be equal to 1.34. Thus the population value for MAE_p is 0.67 for the parameter values that we are using. The sampling performance of the *MAE* is similar to that of the *HMSE*. It, too, should be more reliable than *MSE*.

The results for the sampling distribution of R^2's in the MZ regressions are shown in the first two plots of Figure 11. When the innovations are normal, the R^2's from the MZ regression are much smaller than would be indicated by the population R^2. The R_p^2 from (3) is 0.06. For all three sample sizes, the distribution of R^2's are skewed to the right, with the majority of the density well below R_p^2. For sample size 500, the R^2's have a median of only 0.025. Notice that the Andersen and Bollerslev (1998) bound of 1/3 is drastically overly optimistic for the magnitude of the sample R^2. Similarly, the distribution of the R^2's from the *MZ* regression with t-distributed innovations are shifted to the left as in the normal case. Recall, for this case, the

existence condition for the fourth moment of r_t is violated, and hence, R_p^2 is not defined.

The Monte Carlo estimates of the sampling distributions of the t-statistics for $a=0$ and $b=1$ are shown in the final four plots in Figure 11. Under correct specification of the model, valid inference requires the t-statistic to be centered on zero. We see that as the sample size increases the density functions for the t-statistics appear to converge very slowly, if they converge at all. The distribution of the t-statistics of a converges from above and the t-statistics of b converges from below. For the t-distributed case, convergence to the given values is faster. The apparent inconsistency and non-normality of the least squares estimators of a and b in (2) is likely due to the non-existence of moments. Since h_t is serially correlated and $v_t = h_t(z_t^2 - 1)$ is serially dependent through higher order moments, standard consistency and normality proofs of the least squares estimator would require the fourth moments of h_t and v_t to exist. This is equivalent to requiring the eight moment of r_t to exist. This moment does not exist for the parameter values at which we have conducted the Monte Carlo. The results imply that for a correctly specified GARCH model with high volatility persistence, inference based on the coefficients in the MZ regression are not reliable and will likely lead to a rejection of unbiased and efficient volatility forecasts.

The estimates of the sampling distributions of the R^2's from the log MZ regression are shown in the first two plots in Figure 12. The distribution of the R^2's from this regression is skewed to right as well. For the case when z_t are t-distributed, the R^2's are slightly larger than the normal case. As we do not have a population

value for a correctly specified model for log MZ regression, it is not possible to compare the population R_p^2 with the estimated sampling distribution of the R^2. We do observe that the distributions of the R^2's from the log MZ regression lie further to the left than the distributions of the R^2's from the standard MZ regression. Estimating the *MZ* regression in log form will tend to reduce the measure of forecast accuracy.

Monte Carlo estimates of the sampling distributions of the t-statistics based on the parameter estimators from the log *MZ* regression can be seen in the last four plots in Figure 12. For the log *MZ* regression, the population value for a is given as in equation (3). We estimated these moments by simulation for $z_t \sim N(0,1)$ and for $z_t \sim \sqrt{5/7}\,t(7)$, and found them to be approximately equal to –1.287 and –1.482. Under correct specification of the model when $z_t \sim N(0,1)$, we have that $a = -1.287$ and $b = 1$. As the sample size increases, the distributions of the t-statistics tend to center on zero and approach the normal distribution. Even when returns are t-distributed, with $a = -1.482$ and $b = 1$, the convergence of the sampling distribution of the t-statistics of a appears to be very fast. The distributions of the t-statistics for $b = 1$ are also very close to the standard normal for all three sample sizes and both normal and t-distributed innovations. The log transformation apparently relaxes the condition required for the existence of higher order moments in the regression in (4) to the point where the least squares estimator becomes consistent and asymptotically normal. These results suggest that even when higher order moments of the return r_t are infinite, reliable inference about the unbiasedness and efficiency of the volatility forecast may be conducted in the log form of the MZ regression.

Conclusions

Our study provides several immediate conclusions concerning the evaluation of GARCH volatility forecasts. Among the accuracy evaluation criteria, we recommend the use of *HMSE* based on the fact that it has a population value that can be used as a benchmark when the condition for the existence of the fourth moment of the return is violated, and even when conditional variance has IGARCH properties. The results indicate that the Andersen and Bollerslev (1998) bound are actually optimistic about likely realized values for the R^2 from the MZ regression. The R^2 from both the standard *MZ* regression and the log *MZ* regression are probably not useful descriptive statistics for measuring volatility forecast accuracy. Due to the lack of convergence, t-tests for forecast efficiency should not be used from the *MZ* regression. Inference concerning forecast efficiency can be conducted with the log *MZ* regression.

References

Andersen, T. G., Bollerslew, T., (1998), "Answering the Skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39, 885-904.

Bollerslev,T.,(1986),"Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev T, Ghysels E., (1996), "Periodic autoregressive conditional heteroscedasticity," Journal of Business and Economic Statistics 14: 139-157.

Day, T. E. and C.M. Lewis, (1992) "Stock Market Volatility and the information content of Stock Index Options," *Journal of Econometrics*, 52, 267-287.

Engle, R., (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

Engle, R., (2002), "New Frontiers for ARCH models," *Journal of Applied Econometrics*, 17, 425-446.

Engle, R., Andrew Patton (2001), "What good is a volatility model?" *Quantitative Finance*, pp. 237-245.

Engle, R.F., T. Bollerslev, "Modelling the Persistence of Conditional Variances," *Econometric Reviews 5* (1986), 1-50.

Fama, E., "The Behaviour of Stock Market Prices," *Journal of Business 38* (1965), 34-105.

Granger, W. J., Poon, Ser-Huang, (2003), "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature*, Vol. 41, No. 2, June 2003.

Kenneth D. West, Dongchul Cho, (1995), "Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69 , 367-391.

Mandelbrot, B.B., " The variation of certain speculative prices," *Journal of Business* 36 (1963), 394-419.

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

Pagan, A.R. and Schwert, G.W., (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267-290.

Table 1. Estimated GARCH(1,1) model for returns on exchange rates and market indices.

	FX Canada	FX Japan	FX U.K	Dow Jones	Nasdaq	S&P 500
ψ	.0007 (.000)	0.017 (.008)	.056 (.003)	.012 (.006)	.06 (.019)	.008 (.004)
α	.055 (.007)	.066 (.023)	.030 (.012)	.084 (.028)	.163 (.034)	.071 (.020)
β	.938 (.008)	.908 (.029)	.943 (.023)	.91 (.025)	.819 (.031)	.926 (.017)
Skewness	-.021	-.518	-.08	-.59	-.58	-.70
kurtosis	4.14	5.92	5.21	4.82	4.06	5.67
Jarque-Bera	67.02	502.57	256.78	247.13	130.60	479
Q (10)	26.8	5.01	13.73	22.13	9.76	22.61
$Q^2(10)$	13.4	8	4.95	14.16	11.24	6.96
Existence condition*	.99	.96	.95	1.004	1.021	1.0074

* The reported number is the left hand side of $3\alpha^2 + \beta^2 + 2\alpha\beta < 1$, evaluated at the estimated parameters. If the reported number is larger than one, it indicates that the fourth moment of the return is infinite, and therefore, the population MSE and R^2 do not exist.

Table 2. Criteria for evaluating the accuracy and efficiency of volatility forecasts.

		FX U.K.	FX Japan	Dow Jones Industrial	Nasdaq Composite	S&P 500
MSE	.05	.223	.25	6.23	290.57	.03
MSE_p	.06	.10	1.16	∞	∞	∞
HMSE	2.63	3.45	1.47	3.09	2.78	2.64
$HMSE_p$	2	2	2	2	2	2
MAE	.14	.288	.38	1.48	9.44	1.88
MAE_p	.12	.21	.67	2.26	3.66	5.54
MZ Regression						
a	.08 (.048)	.22 (.136)	.30 (.137)	.38 (.299)	4.81 (1.539)	0.54 (.334)
t-ratio a=0	1.67*	1.62	2.19**	1.27	3.12**	1.62
b	.45 (.417)	.2 (.522)	.04 (.131)	.69 (.232)	.514 (.224)	.66 (.211)
t-ratio b=1	1.32	1.53	7.33**	1.34	2.17**	1.61
R^2	.009	.001	.0	.05	.04	.04
R_p^2	.21	.017	.082	∞	∞	∞
Log MZ Regression						
a	-2.54 (.791)	-1.67 (.993)	-2.36 (.590)	-1.23 (.147)	-0.48 (.390)	-1.21 (.179)
t-ratio a=-1.287	-1.58	-.38	-1.81	.388	2.07**	.43
b	.32 (.372)	.627 (.73)	.01 (.664)	.87 (.206)	.70 (.181)	.74 (.241)

Table 2—Continued

		FX U.K.	FX Japan	Dow Jones Industrial	Nasdaq Composite	S&P 500
t-ratio b=1	1.83*	.511	1.49	.631	1.66	1.07
R^2	.003	.003	0	.046	.04	.032

Robust standard errors are in parenthesis, * denotes significant at the 10 percent

Figure 10. Sampling distribution of criteria for evaluation of volatility forecasts.

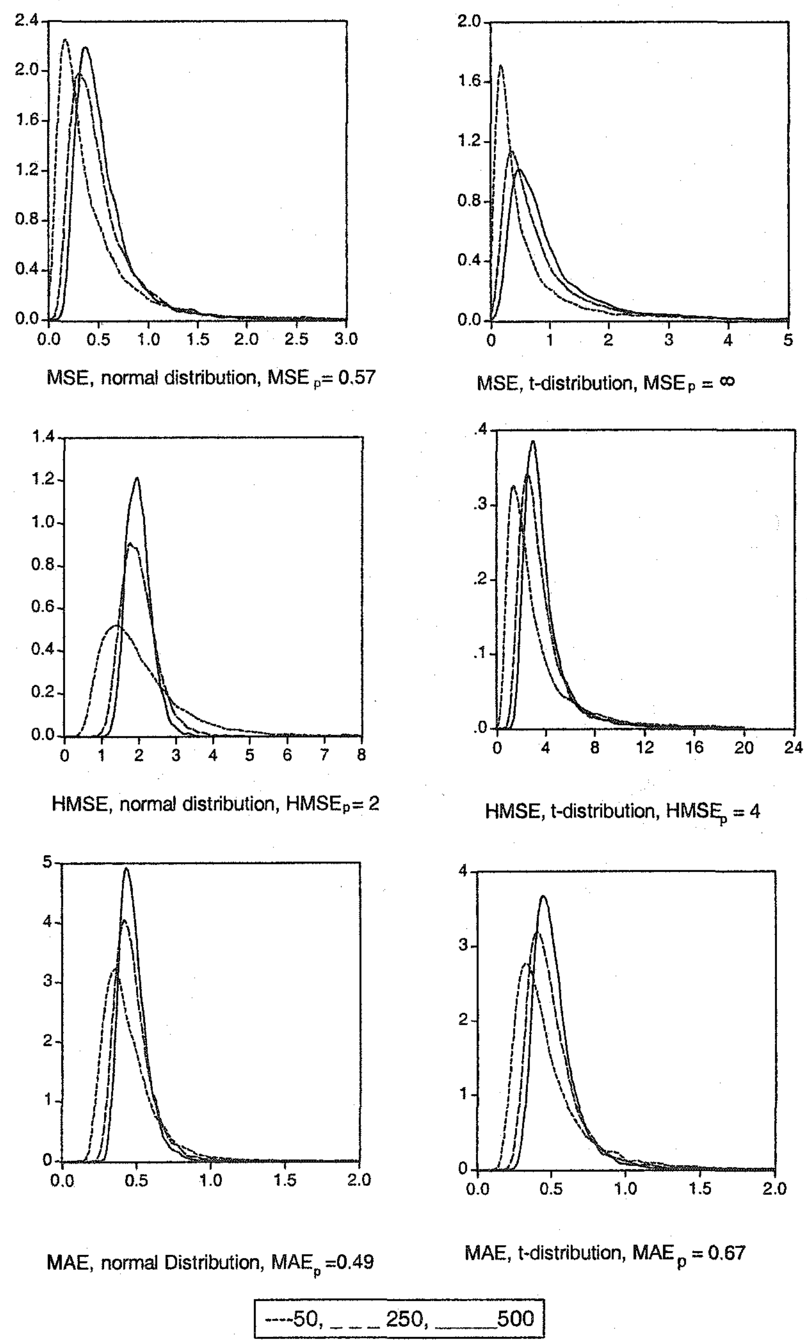


Figure 11. Sampling distributions of statistics from Mincer-Zarnowitz regression.

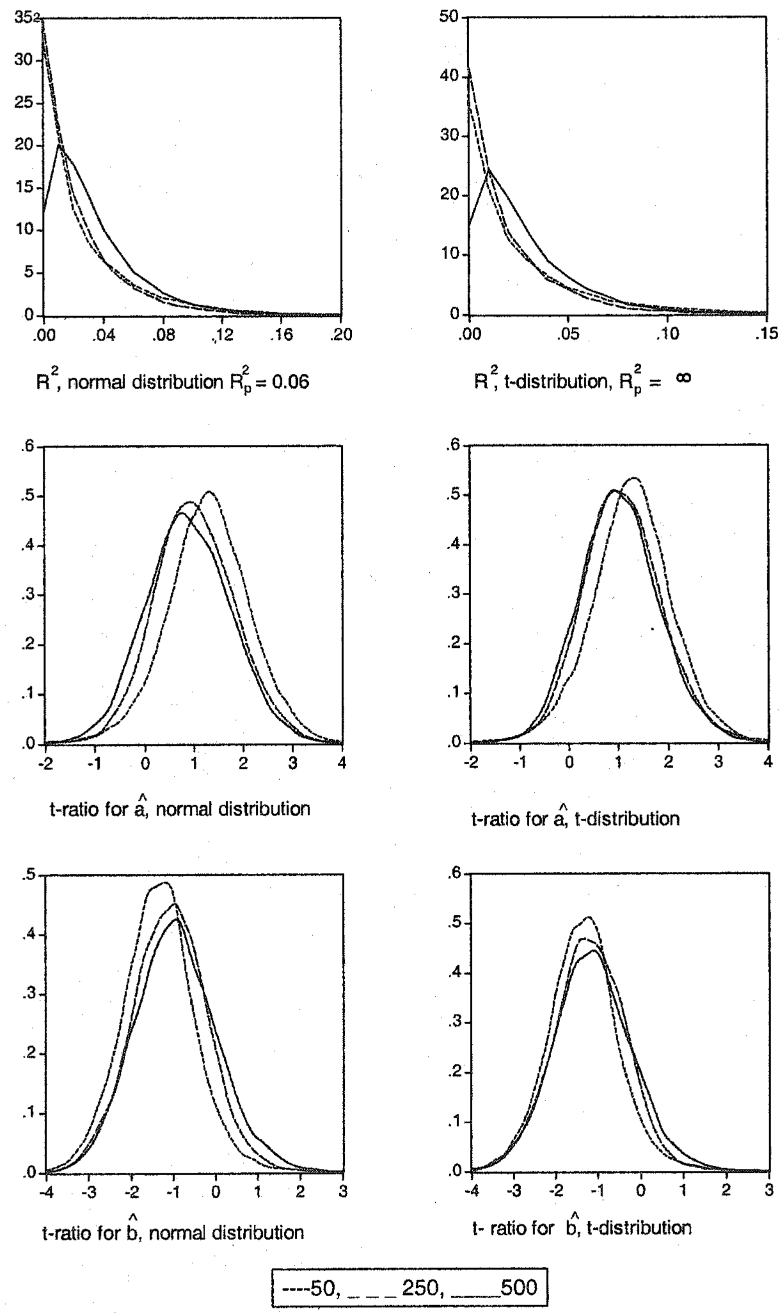
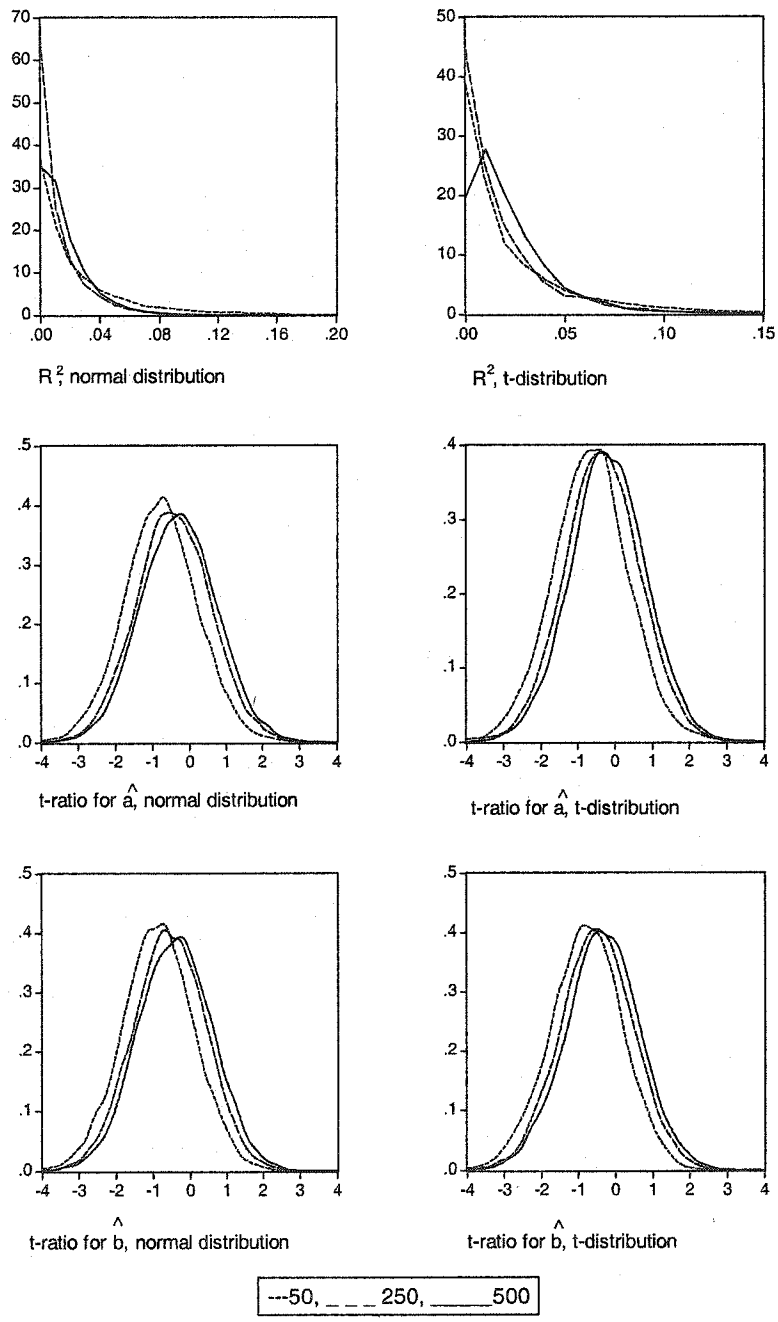


Figure 12. Sampling distributions of statistics, log Mincer-Zarnowitz regression.



CHAPTER III

OPTIMAL PREDICTION UNDER ASYMMETRIC LOSS

Introduction

In the literature, a widely used forecast evaluation criteria is the MSE, which is a symmetric quadratic loss function. MSE penalizes the positive errors and negative errors of the same magnitude equally. We know in finance, however, that positive errors can be more costly then negative errors. Despite this fact, MSE has become a very popular loss function mostly because of mathematical convenience. Realistically, forecasters do not necessarily have a quadratic cost nor a symmetric loss function[4].

Symmetric loss functions are convenient because the analytical expression for the optimal predictor is straightforward: it is the conditional mean. However, if we use a general loss function, deriving the closed form for the optimal predictor analytically is very demanding, and some times impossible. Studies have avoided using general asymmetric loss functions mainly because most of the time the closed form for the optimal predictor does not exist. Under asymmetric loss, the optimal predictor is no longer the conditional mean. Granger (1969) showed that the optimal predictor under asymmetric loss is the conditional mean plus a constant bias term. Granger (1969) assumed a constant conditional prediction error variance. Christoffersen and Diebold

[4] See Granger (1969), Granger and Newbold (1986, p.125) and Stockman (1987).

(1997, 1996) considered the same problem, but generalized Granger's result. They showed that for conditionally Gausian processes if an agent has an asymmetric loss function, adding a constant term is not sufficient and that time varying second order moments become relevant for optimal prediction. They derived the analytical expression for the optimal predictor for two specific asymmetric loss functions. These are the LINLIN, first used by Granger (1969), and LINEX loss function introduced by Varian (1974) and used by Zellner (1986). For more general loss functions they showed how to approximate the optimal predictor numerically.

Although Christoffersen and Diebold (1996, 1997) have important practical implications, there is no empirical study that illustrates the gain in adding a bias to the conditional mean as the optimal predictor when agents have asymmetric loss. In this paper we bridge this gap by illustrating the loss associated with using the optimal, the pseudo optimal and the conditional mean predictor under asymmetric loss. We consider returns on three representative exchange rates and returns on five representative market indices that are commonly used in empirical work. We consider different predictors of the variance since there is no unanimous agreement on the best variance model in the literature. We choose the most popular models that are believed to characterize conditional volatility in asset returns. The models we consider are normal-GARCH(1,1), t-GARCH(1,1) and a nonparametric model. We empirically illustrate that the loss associated by using the optimal predictor which incorporates time varying second order moments is much smaller than the loss associated with using the conditional mean predictor when agents have asymmetric loss function. However, when we compare the difference in loss between the optimal predictor and

the pseudo optimal one, the results are mixed and very sensitive to the degree of asymmetry and the conditional variance parameters being used.

In section 2, we introduce the LINLIN asymmetric loss function and the variance models. Section 3 describes the data. Section 4 presents the results. Section 5 is a conclusion.

Loss function and the variance models

LINLIN loss function

The LINLIN loss function was used by Granger (1969). It is linear on each side of the origin, however, positive errors are penalized differently than the negative errors in that the lines have different slopes in each side of the origin. The LINLIN loss function is

$$L(y_{t+h}-\hat{y}_{t+h})=\begin{cases} a\,|\,y_{t+h}-\hat{y}_{t+h}\,| & if\,(y_{t+h}-\hat{y}_{t+h})>0 \\ b\,|\,y_{t+h}-\hat{y}_{t+h}\,| & if\,(y_{t+h}-\hat{y}_{t+h})<=0 \end{cases},$$

where $L(.)$ is a loss function defined on h-step a-head prediction error, y_{t+h} is the realized value of y, $t+h$ periods a head and $\hat{y}_{t+h}$ is the predicted value of y_{t+h}. The ratio a/b measures the cost of over predicting relative to the cost of under predicting. If a/b=2, it means that the loss associated with a positive error is twice as much the loss associated with negative error of the same magnitude.

Christoffersen and Diebold (1997) derived the optimal predictor, a pseudo-optimal predictor and the expected losses associated with each for the LINLIN asymmetric loss function under the assumption of conditional normality. Given $y_{t+h} \mid \Omega_t \sim N(\mu_{t+h|t}, \sigma_{t+h|t})$, they showed the optimal predictor to be

$$y_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t}\Phi^{-1}(\frac{a}{a+b}),$$

and the pseudo-optimal predictor to be

$$y_{t+h} = \mu_{t+h|t} + \sigma_h\Phi^{-1}(\frac{a}{a+b}),$$

where σ_h^2 is the h-step a head homoscedastic prediction error variance and $\Phi(z)$ is the N-(0,1) c.d.f[5]. Under asymmetric loss, if the conditional heteroscedasticity is ignored, the associated conditionally expected loss will be greater then the case when optimal predictor is used. The pseudo-optimal predictor coincides with the optimal predictor when $\sigma_h^2 = \sigma_{t+h|t}^2$. Notice that for a given series, and for certain degree of asymmetry the difference in mean losses associated with using the optimal versus the pseudo optimal predictor, depend on the deviation of the square root of conditional variance from its mean $(\sigma_{t+h|t} - \sigma_h)$. Thus the difference between the mean losses is proportional to this variation. If this variation is large, then incorporating the time varying second order moments can be very crucial in terms of reducing the losses.

Throughout the paper we consider $h = 1$, which corresponds to one-step a-head prediction. For the series we consider, the conditional means are constant. We obtain the conditional variance using different models. We do not limit ourselves to

[5] However, as CD (1997) point out, the conditionally Gaussian assumption can be relaxed. The optimal predictor is obtained by substituting the appropriate conditional CDF.

conditionally Gaussian processes. Although CD assumed conditional normality it is a common fact that financial return series have excessive kurtosis. To take this fact into account we estimate a GARCH (1,1) model with t-distributed innovations as well. In order to prevent any possible misspecification problems due to parameterization, we also use a nonparametric model. We calculate the value of the optimal, the pseudo-optimal and the conditional mean predictors and the average losses associated with them in the out of sample prediction period. We consider different degrees of asymmetry to compare the loss associated with using different predictors. Specifically, we fix $b=1$ and change the values of a. We consider cases up to where $a/b=20$. This asymmetric penalization scheme is plausible in finance.

Variance models

GARCH models

The most commonly used model for time-varying volatility is the G/ARCH model of Engle (1982) and Bollerslev (1986). A GARCH(1,1) model for the return on a financial asset, r_t, can be written

$$\begin{aligned} r_t &= \sigma_t \cdot z_t,\ z_t \sim IID(0,1) \\ \sigma_t^2 &= \gamma + \alpha\, r_{t-1}^2 + \beta\, \sigma_{t-1}^2 \end{aligned} \tag{1}$$

where $\gamma > 0$, $\alpha \geq 0$ and $\beta \geq 0$. We assume z_t has finite first and second moments. For a normal GARCH(1,1) model, denoted n-GARCH(1,1), we assume an independent normal innovation. The return r_t is weakly stationary if its variance is finite. This will be the case if $\alpha + \beta < 1$. Since $E(r_t^2 \mid \mathrm{I}_{t-1}) = \sigma_t^2$, where I_{t-1} is the

information set available at time t-1, the conditional variance σ_t^2 is the minimum mean square error predictor of the realized volatility r_t^2.

Financial returns are known to be heavy tailed, and often the conditional normality assumption is inappropriate. We also estimate GARCH(1,1) models with conditional Student t distributions, denoted by t-GARCH (1,1), for each return series. The one-step a head optimal and the pseudo optimal predictors for the GARCH (1,1) model under conditional-t- distribution assumption are:

$$y_{t+1} = \mu_{t+1|t} + \sigma_t F^{-1}(\frac{a}{a+1}, \upsilon)$$

and

$$y_{t+1} = \mu_{t+1|t} + \sigma\ F^{-1}(\frac{a}{a+1}, \upsilon),$$

where υ is the data specific degrees of freedom and F^{-1} is the inverse of conditional t cumulative distribution function.

Non-parametric variance model

Non-parametric models are popular because they do not introduce any parametric assumption for the underlying distribution and thus prevent the bias problem due to misspecification. Predicting exchange rate returns non- parametrically goes back to Diebold and Nason (1990). The idea behind using nonparametric techniques to estimate and predict exchange rates is to exploit any non-linearities may be present in the financial return data. Pagan and Schewert (1990) also use this technique for in sample prediction as well.

Let X_i be a set of conditioning random variables. Let $m(x_i)$ be the mean of the conditional density $f(y_i \mid x_i) = f(Y \mid X = x_i)$. By definition of the conditional mean,

$m(x) = \int_{-\infty}^{+\infty} (yf(y,x)/f_1(x)dy,$ where $f_1(x)$ is the marginal density of X at x. The densities $f(y,x)$ and $f_1(x)$ can be placed by kernel density estimators $\hat{f}(y,x)$ and $\hat{f}_1(x)$ [see Nadaraya (1964) and Watson (1964)]. The estimator for $m(x)$ at any point $\tilde{x}$ can be expressed as $\hat{m}(\tilde{x}) = \sum_{i=1}^{n} K(\frac{x_i - \tilde{x}}{h}) y_i \Big/ \sum_{i=1}^{n} K(\frac{x_i - \tilde{x}}{h})$, where h is the optimal bandwidth. We choose $K(.)$, to be a Gausian multivariate kernel, $K(\psi_i) = (2\pi)^{-q/2} \exp(1/2\psi^{'}\psi)$. We consider h-optimal to be fixed for a given data set. The optimal bandwidth is calculated as $h = \sigma_y (4/(n(2q+1))^{1/(4+q)}$, where q is the number of conditioning variables. To predict the second order conditional moment at a point $\tilde{x}$, we use the formula, $\hat{\sigma}_u^2(\tilde{x}) = \sum K(\Psi_i)\hat{u}_i^2 \Big/ \sum K(\Psi_i)$, where $\hat{u}_i = y_i - m(x_i)$ [See Pagan and Schwert (1990) for more detail].

When forecasting out of sample, we replace the time varying conditional mean and conditional standard deviations in the Christoffersen and Diebold (1997) optimal prediction formula by their nonparametric estimators. In the pseudo-optimal prediction formula, we replace the constant conditional standard deviation with the sample standard deviation of the in sample nonparametric residuals. For both predictors, we estimate the inverse conditional c.d.f. nonparametrically. In the nonparametric estimation, the conditioning variables are lagged returns.

Data

To determine whether using the optimal predictor in the presence of asymmetric loss is empirically useful, we use data representative of that used in financial forecasting. We estimate models and predict returns for three exchange rates between the U.S. dollar and the Canadian Dollar, the Japanese Yen and the British Pound. We also estimate models and predict returns for five major market indexes: the Dow Jones Industrial Average, the NASDAQ Composite and the S&P 500, the NIKKEI, and FTSE. The frequency of the data is weekly. The data run from October 19, 1984 to July 27, 2001. Data from October 19, 1984 to July 28, 2000 is used for estimation and the last year is retained for evaluating out of sample losses.

Results

We first estimate n-GARCH (1,1) models for each return series. Although the efficient market hypothesis suggests that returns should be serially uncorrelated, we check for possible serial correlation in the returns. We find that all the return series on exchange rates are serially uncorrelated. Table 3 shows the estimated GARCH (1,1) models for returns on three exchange rates and diagnostic statistics for the standardized residuals. The estimates for the conditional variance parameters are all significant. The skewness coefficients of the standardized residuals show that they are slightly negatively skewed. The kurtosis coefficients are between five and seven, suggesting that the data do not have normal distributions. The Jarque-Bera statistics are all highly significant. The Ljung-Box Portmanteau tests for the serial correlation in the standardized and squared standardized residuals up to 10 lags indicate that the

residuals are white noise. This suggests that we can characterize the weekly volatility dependencies by a GARCH(1,1) model. When we look at the volatility persistence as indicated by $\alpha + \beta$, the parameter estimates sum close to .8 for Canada and .96 for Japan and .98 for U.K. This high persistence suggests that market volatility is predictable. Table 4 shows the estimated GARCH models for the returns on indices and diagnostic statistics for the standardized residuals. We find that the return series for S&P 500 and Dow Jones are AR(1) in the mean and the return series for Nasdaq, FTSE and NIKKEI are AR(2) in the mean. The estimates for the conditional variance parameters are all significant. The skewness coefficients of the standardized residuals show that they are slightly negatively skewed. The kurtosis coefficients are between four and seven, suggesting that the data do not have normal distributions. The Jarque-Bera statistics are all highly significant. The Ljung-Box Portmanteau tests for the serial correlation in the standardized and squared standardized residuals up to 10 lags indicate that the residuals are white noise. This suggests that we can characterize the weekly volatility dependencies by a GARCH(1,1) model. When we look at the volatility persistence as indicated by $\alpha + \beta$, the parameter estimates sum close to .97 for Dow Jones, FTSE and NIKKEI and sum to .99 for NASDAQ and S&P 500. This indicates that the return volatility for NASDAQ and S&P 500 may be the IGARCH (1,1) model of Engle and Bollerslev (1986).

To determine if the results were sensitive to the assumption of conditional normality, we re-estimated the GARCH(1,1) models using a conditional t-distribution. The estimated models are shown in the second parts of Table 3 and Table 4. The GARCH coefficients are similar to those obtained under conditional normality. The

degrees of freedom parameters are all significantly different from three, rejecting conditional normality.

We then use the estimated n-GARCH(1,1) models to compute the optimal, pseudo-optimal and conditional mean predictors for each of the series over the forecast period. Figure 13 shows the ratio of the mean losses between the optimal predictor and the conditional mean, and also the ratio of the mean losses between the pseudo-optimal predictor and the conditional mean for each of the exchange rate return series for the n-GARCH(1,1) models. For all the series, it is striking that there is considerable gain by using the optimal or the pseudo optimal predictor versus the conditional mean predictor. The conditional mean predictor performs the worst, even for very low degrees of asymmetry. For the exchange rate series, use of the optimal predictor versus the pseudo-optimal predictor provides little or no reduction in average loss.

Figure 14 is the corresponding graph of the ratios of average losses for the t-GARCH(1,1) model for exchange rate returns. The results are close to the n-GARCH(1,1) model except for the return series on Japanese yen. For the return series on Japanese yen there is some improvement in the performance of the optimal predictor relative to the pseudo- optimal predictor for degrees of asymmetry equal to four and more.

Figure 15 shows the ratio of the mean losses between the optimal predictor and the conditional mean, and also the ratio of the mean losses between the pseudo-optimal predictor and the conditional mean for each exchange rate return series using the nonparametric estimators. We experimented with the number of conditioning variables to use in the estimation. For all series, we considered up to 5 lagged returns.

For all of the series, the average losses over the forecast period did not seem to improve when more than one lag was used. Therefore, we report results where one-lagged return is used for the estimation. The results for the nonparametric estimation, with the exception of Canada, are basically the same as the results obtained from the forecasts of the n-GARCH(1,1) model. For the return series on the Canadian exchange rate, the pseudo-optimal predictor seems to out perform the optimal predictor for degrees of asymmetry equal to two or more. The Christoffersen and Diebold (1997) optimal predictor ignores parameter uncertainty. The nonparametric estimator of the conditional variance for this series may be sufficiently noisy that it actually increases the losses relative to the pseudo-optimal predictor which requires an estimator of the single constant conditional variance parameter.

Figure 16 shows the ratio of the mean losses between the optimal predictor and the conditional mean and also the ratio of the mean losses between the pseudo-optimal predictor and the conditional mean for the returns on the five market indices using the n-GARCH(1,1) model. For the FTSE, NIKKEI, and S&P 500 there does not seem to be any difference using the pseudo-optimal versus the optimal predictor. For Dow-Jones and NASDAQ there is evidence that the optimal predictor out performs the pseudo- optimal predictor. This is particularly true, for the NASDAQ where the gain reaches about 30% for high degrees of asymmetry.

Figure 17 presents the ratio of the losses for the five returns on the market indices from t-GARCH(1,1) model. The results suggest that there is little gain using the optimal predictor versus the pseudo-optimal predictor for FTSE and NASDAQ. For the return series on NIKKEI, the pseudo-optimal predictor slightly outperforms the optimal predictor.

Figure 18, presents the results from the nonparametric model. The results are almost identical to the results from n-GARCH(1,1) model except for the S&P 500. For the S&P 500, unlike in the n-GARCH(1,1) case, the optimal predictor outperforms the pseudo-optimal predictor.

Conclusion

We consider the mean losses associated with using the optimal predictor, pseudo-optimal predictor and the conditional mean predictor when agents have asymmetric loss function. Our results provide strong empirical evidence to the Granger (1969). Conditional mean predictor performs very poorly compared to the optimal and the pseudo optimal predictors. For all series, loss associated with using the conditional mean predictor versus using the pseudo or the optimal predictor is considerably high even for moderate degrees of asymmetry, regardless of the variance model being used. This result suggests that if agents have any kind of asymmetry, the conditional mean predictor should not be used at all.

The comparison between the pseudo-optimal predictor and the optimal predictor is not very straightforward, as the results depend on the series and which variance model is used. It is evident that the results are very sensitive to the variance parameters being used. In general, the optimal predictor does tend to out perform the pseudo-optimal predictor. The difference can be very small 3-5% or as large as a 27% reduction in loss as in the case of the NASDAQ index.

Is this reduction in the mean loss financially important? A parallel argument to the difference between economic significance and statistical significance can be

carried out here. Cleary, if you are a hedge fund manager, even 1% loss reduction would be of critical importance. However for different utility functions, or moderate degrees of asymmetry, agents might be indifferent between using the optimal versus the pseudo optimal predictor.

References

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307-327.

Christoffersen, P. F. and Diebold F. X. (1996), "Further results on Forecasting and Model Selection Under Asymmetric Loss," *Journal of Applied Econometrics*, Vol. 11, No. 5, Special Issue: Econometric Forecasting (Sep.-Oct., 1996), 561-571.

Christoffersen, P. F. and Diebold F. X. (1997), "Optimal Prediction Under Asymmetric Loss," Econometric Theory, 13, 808-817.

Diebold, F. X. and Nason J. A (1990), "Nonparametric Exchange Rate Prediction" *Journal of International Economics*, 28, 315-332.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

Engle, R.F., T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews* ,5, 1-50.

Granger, C. W. J. (1969), "Prediction with a generalized cost of error function," *Operational Research Quarterly*, 20, 199-207.

Granger, C. W. J. and P. Newbold (1986), *Forecasting Economic Time Series,* 2nd ed. Orlando: Acedemic Press.

Nadaraya, E. A. (1964),"On Estimating Regression," *Theory of Probability and its Applications*, 9,141-142.

Pagan, A. R. and Schewert G. W. (1990) "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267-290.

Pagan A. and Ullah, A. (1999), "Nonparametric Econometrics" Cambridge.

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

Stockman, A.C. (1987), "Economic Theory and Exchange Rate Forecasts," *International Journal of Forecasting*, 3, 3-15.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhya* , Series A, 26,359-372.

Varian, H. (1974), "A Bayesian Approach to Real Estate Assessment. In S.E. Feinberg and A. Zellner(eds.), *Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage*, 195-208.

Zellner, A. (1986), "Bayesian Estimation and Prediction Using Asymmetric Loss Functions," *Journal of American Statistical.*

Table 3. Estimated n-GARCH(1,1) and t-GARCH(1,1) models for returns on exchange rates.

Model	parameters	FX Canada	FX Japan	FX U.K.
	ψ		0.116	0.028
n-GARCH(1,1)		0.093	(0.086)	(0.018)
		(0.041)		
	α		0.064	0.059
		0.135	(0.022)	(0.030)
		(0.041)		
	β		0.898	0.926
		0.656	(0.036)	(0.030)
		(0.101)		
	Skewness	0.296	-0.973	-0.226
	Kurtosis	5.87	7.349	5.467
	Jaque-Bera	295.5	779.358	216.003
	$Q(10)$	9.831	11.713	15.290
	$Q^2(10)$	4.960	6.751	9.635
t-GARCH(1,1)	ψ	0.053	0.191	0.026
		(.016)	(.410)	(.008)
	α	0.150	0.080	0.049
		(.062)	(.032)	(0.015)
	β	0.745	0.859	0.937
		(.323)	(.261)	(.326)
	υ	5.255	4.410	8.299
		(0.543))	(.294)	(1.378)
	Skewness	0.143	-0.854	-.287
	Kurtosis	4.25	5.65	3.84
	Jaque-Bera	18.15	27.95	48.12
	$Q(10)$	8.712	9.056	12.994
	$Q^2(10)$	5.138	4.322	6.734

Table 4. Estimated n-GARCH(1,1) and t-GARCH(1,1) models
for returns on market indexes.

Model	parameters	Dow-Jones	FTSE	NASDAQ	NIKKEI	S&P 500
n-GARCH(1,1)	ψ	0.142 (0.062)	0.194 (0.077)	0.284 (0.113)	0.266 (0.091)	0.066 (0.037)
	α	0.117 (0.039)	0.099 (0.058)	0.242 (0.079)	0.120 (0.035)	0.114 (0.035)
	β	0.856 (0.042)	0.865 (0.058)	0.748 (0.057)	0.853 (0.036)	0.877 (0.033
	Skewness	-0.407	-0.334	-0.576	-0.761	-0.411
	Kurtosis	3.774	4.893	5.379	6.611	3.623
	Jaque-Bera	43.277	138.787	239.297	527.280	36.531
	$Q(10)$	10.609	8.182	7.031	10.772	14.482
	$Q^2(10)$	9.680	14.385	5.049	3.250	8.736
t-GARCH(1,1)	ψ	0.122 (.0493)	0.227 (.0759)	0.274 (0.114)	0.246 (0.085)	0.063 (.0148)
	α	0.102 (.332)	0.072 (.026)	0.154 (0.634)	0.160 (0.059)	0.094 (.041)
	β	0.875 (.256)	0.878 (.383)	0.812 (.340)	0.824 (.213)	0.895 (.472)
	υ	9.326 (1.269)	7.961 (2.551)	5.622 (1.250)	6.393 (1.935)	11.348 (2.599)
	Skewness	-0.363	-0.225	-0.671	-.643	-0.398
	Kurtosis	3.244	3.981	4.253	5.123	3.450
	Jaque-Bera	10.684	17.022	20.567	48.434	7.540
	$Q(10)$	8.043	6.003	7.561	8.089	12.554
	$Q^2(10)$	7.569	12.945	5.119	4.589	5.943

Figure 13. Ratio of Average Losses, n-GARCH(1,1).

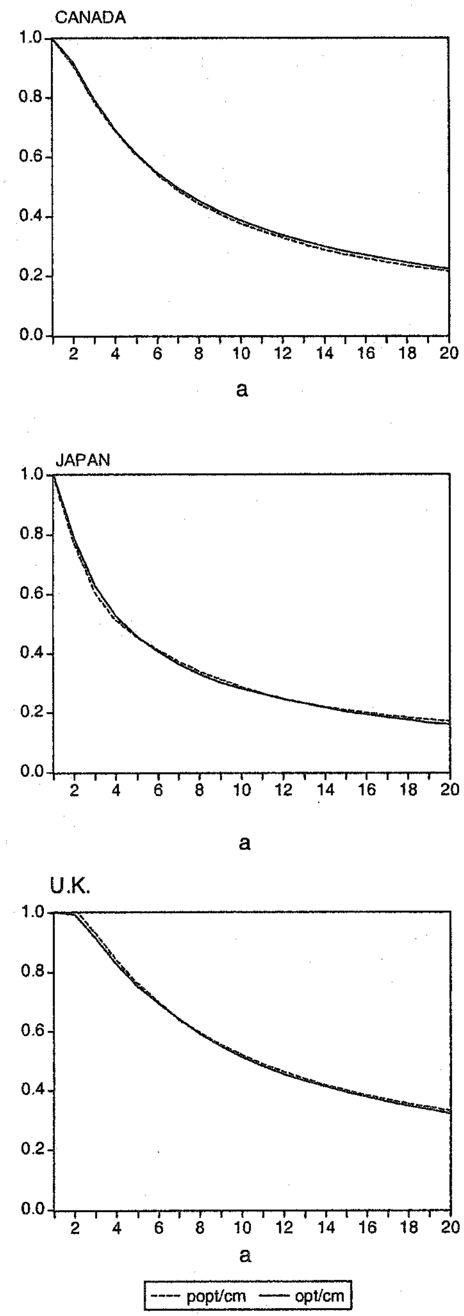


Figure 14. Ratio of Average Losses, t-GARCH(1,1).

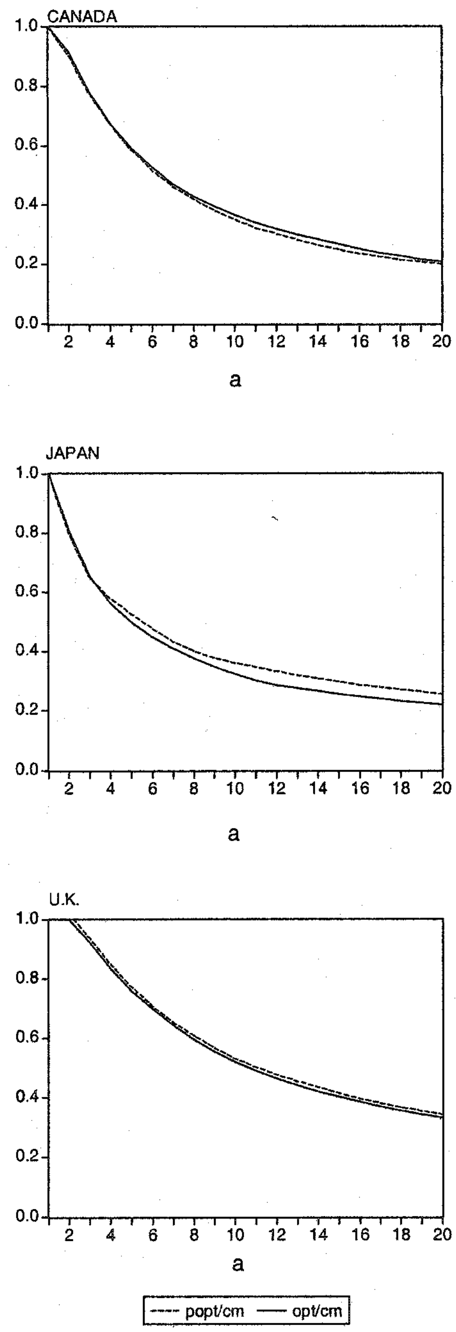


Figure 15. Ratio of Average Losses, nonparametric model.

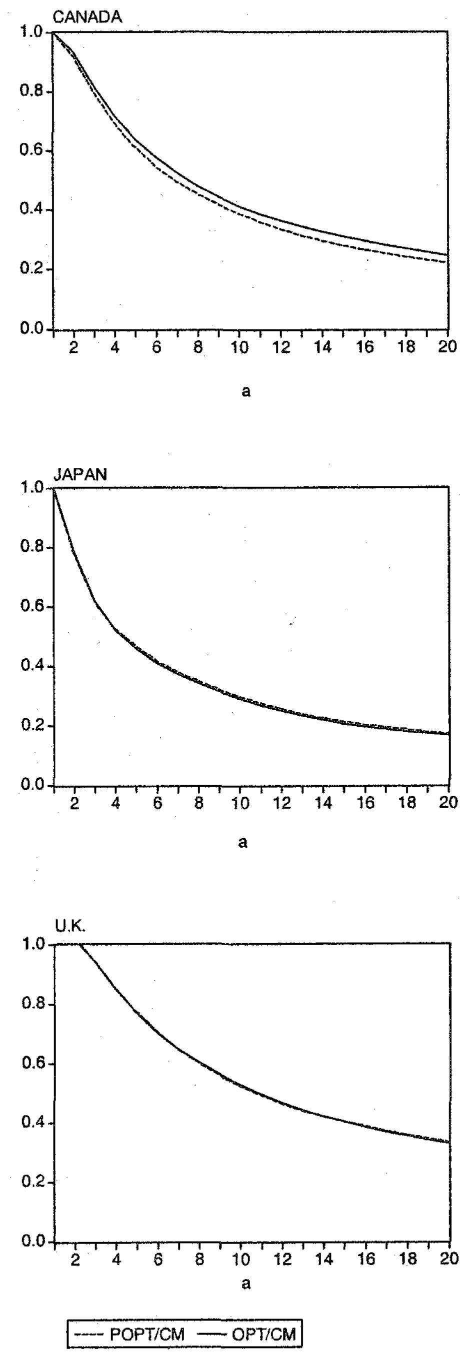


Figure 16. Ratio of Average Losses, n-GARCH(1,1).

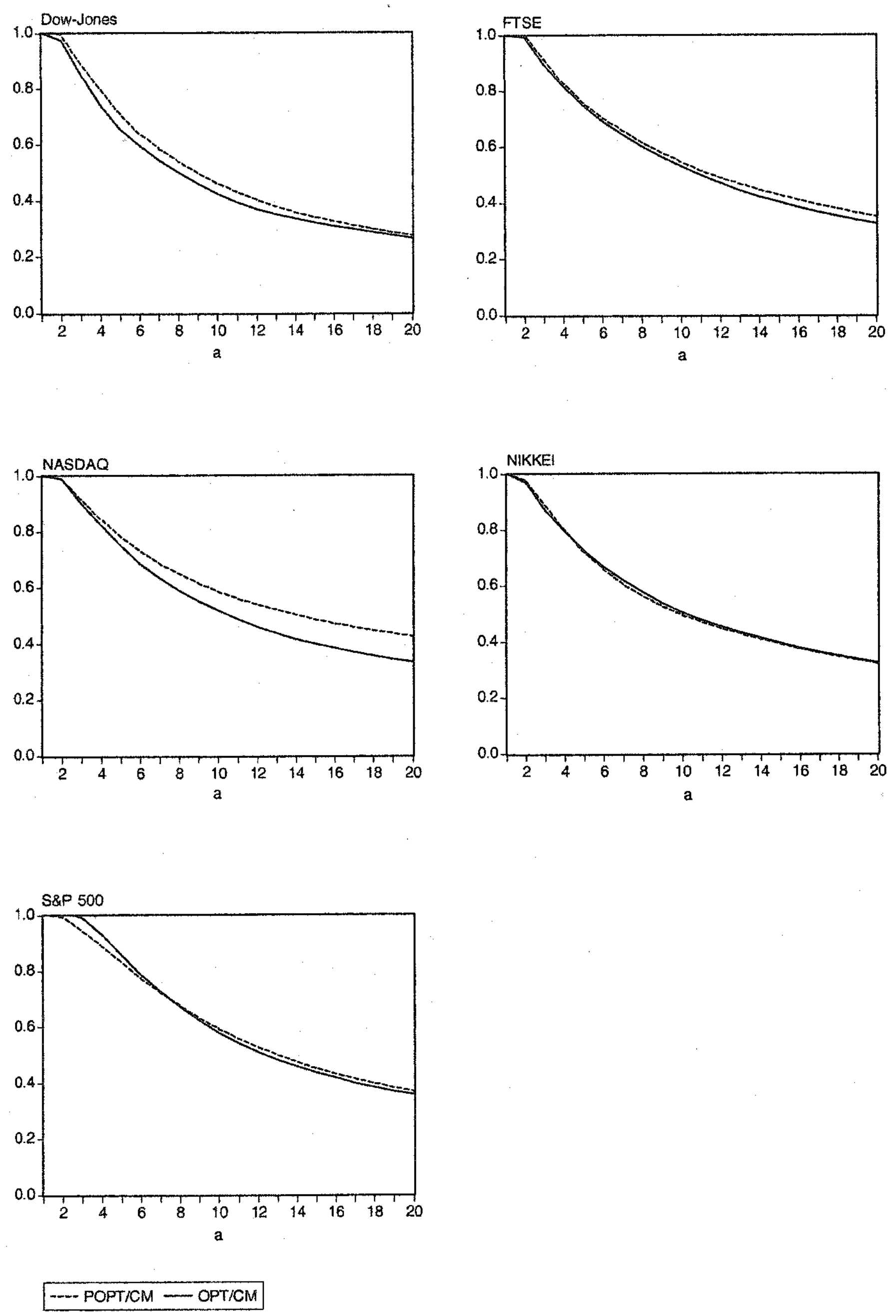


Figure 17. Ratio of Average Losses, t-GARCH(1,1).

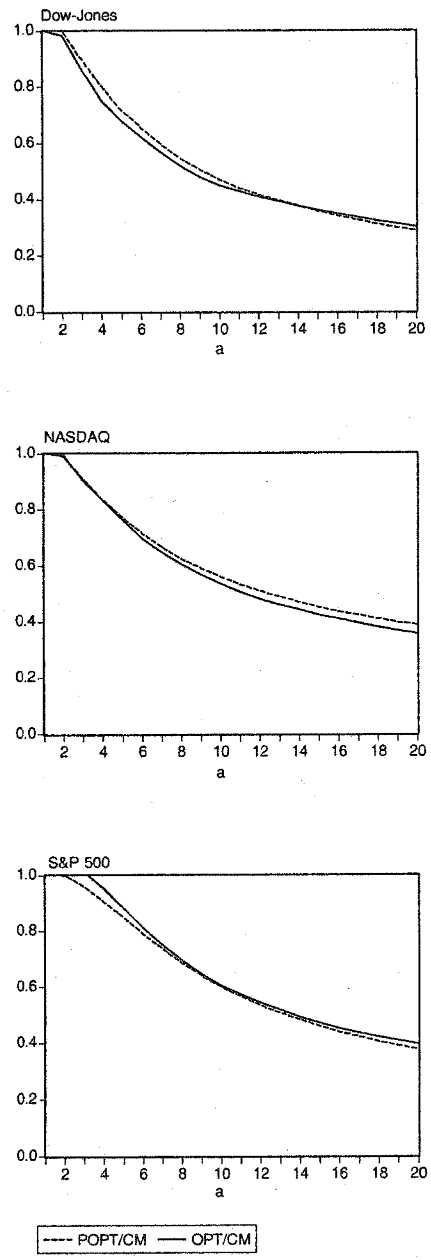
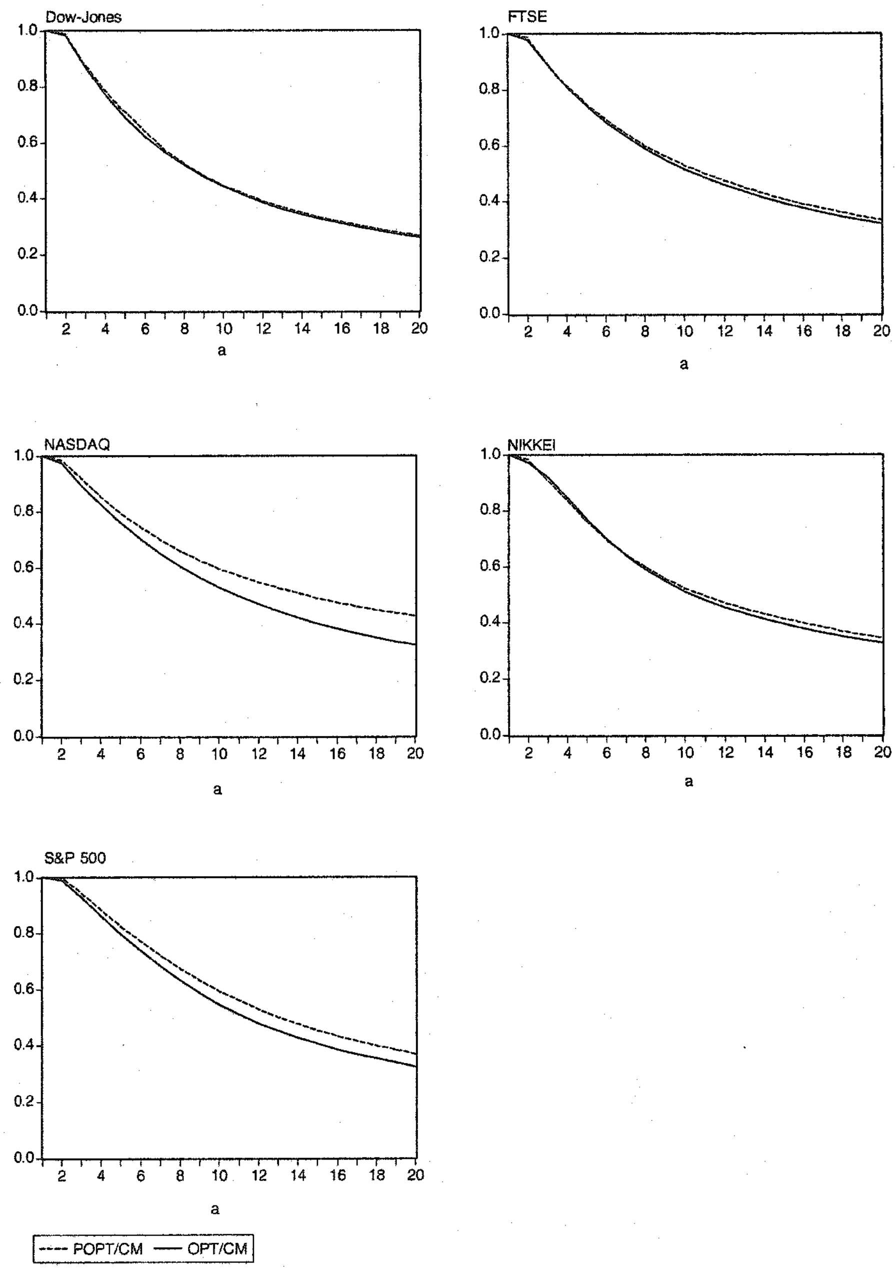


Figure 18. Ratio of Average Losses, nonparametric model.



Appendix A

Derivation of population moments

Proof of Theorem 1

The population MSE is

$$
\begin{aligned}
MSE_p &= E[(r_t^2 - h_t)^2] \\
&= E(r_t^4 - 2r_t^2 + h_t^2) \\
&= E(r_t^4) - 2E(r_t^2 h_t) + E(h_t^2). \qquad \text{(A.1)}
\end{aligned}
$$

Substituting for the powers of r_t from (1) and using the law of iterated expectations, for the first term in (A.1) we have $E(r_t^4) = E[E(h_t^2 \cdot z_t^4 \mid \Phi_{t-1})] = \kappa\, \mathrm{E}(\mathrm{h}_t^2)$. Again, using (1) and iterated expectations, the second term in (A.1) is $E(r_t^2 h_t) = E[E(h_t^2 z_t^2 \mid \Phi_{t-1})] = E(h_t^2)$. Substituting these values back into (A.1), we get

$$
MSE_p = (\kappa - 1)E(h_t^2) \qquad \text{(A.2)}
$$

The second moment of the conditional variance , h_t, can be calculated as

$$
\begin{aligned}
E(h_t^2) &= \mathrm{E}[(\psi + \alpha\, \mathrm{r}_{t-1}^2 + \beta\, \mathrm{h}_{t-1})^2] \\
&= \psi^2 + \alpha^2 E(r_{t-1}^4) + \beta^2 \mathrm{E}(\mathrm{h}_{t-1}^2) + 2\alpha\beta\, \mathrm{E}(\mathrm{r}_{t-1}^2 h_{t-1}) \qquad \text{(A.3)}
\end{aligned}
$$

From Bollerslev (1986, Theorem 2), we have that $E(r_t^2) = E(h_t) = \psi/(1-\alpha-\beta)$. Substituting this and the expressions for $E(r_{t-1}^4)$ and $E(r_{t-1}^2 h_{t-1})$ as given above back into (A.3), we get

$$
E(h_t^2) = \frac{\psi(1+\alpha+\beta)}{(1-\alpha-\beta)} + (\alpha^2\kappa + \beta^2 + 2\alpha\beta)\, \mathrm{E}(\mathrm{h}_{t-1}^2)\,.
$$

This is a first order linear difference equation in h_t. Assuming $\alpha^2\kappa + \beta^2 + 2\alpha\beta < 1$, this equation can be solved as

$$E(\mathrm{h}_{t-1}^2) = \frac{\psi(1+\alpha+\beta)}{(1-\alpha-\beta)[1-(\alpha^2\kappa+\beta^2+2\alpha\beta)]}.$$

Substituting this expression for $E(h_t^2)$ into (A.2) we get

$$MSE_p = (\kappa - 1)\frac{\psi^2(1+\alpha+\beta)}{(1-\alpha-\beta)[1-(\alpha^2\kappa+\beta^2+2\alpha\beta)} \qquad \text{(A.4)}$$

q.e.d

Proof of Theorem 2

Expanding the square and using (1), the population HMSE is

$$\begin{aligned} HMSE_p &= E\left[\left(\frac{r_t^2}{h_t}-1\right)^2\right] \\ &= E\left(\frac{r_t^4}{h_t}\right) - 2E\left(\frac{r_t^2}{h_t}\right) + 1 \\ &= E(z_t^4) - 2E(z_t^2) + 1 \\ &= \kappa - 1. \end{aligned}$$

q.e.d

Proof of Theorem 3

Substituting from (1), the population value for MAE is

$$\begin{aligned} MAE_p &= E(|\, r_t^2 - h_t \,|) \\ &= E\left(|\, r_t^2 / \sqrt{h_t} - 1 \,| \cdot h_t\right) \\ &= E(|\, z_t^2 - 1 \,| \cdot h_t). \end{aligned}$$

Then $E(|\, z_t^2 - 1 \,| \cdot h_t) = E[E(|\, z_t^2 - 1 \,| \cdot h_t \mid \Phi_{t-1})] = E(|\, z_t^2 - 1 \,|) \cdot E(h_t)$. Using $E(h_t)$ as given above

$$MAE_p = E(|\, z_t^2 - 1 \,|) \cdot \frac{\psi}{1-\alpha-\beta}.$$

q.e.d