WP/03/131

IMF Working Paper

Predictive Ability of Asymmetric Volatility Models at Medium-Term Horizons

Turgut Kışınbay

INTERNATIONAL MONETARY FUND

 WP/03/131

IMF Working Paper

Monetary and Financial Systems Department

Predictive Ability of Asymmetric Volatility Models at Medium-Term Horizons

Prepared by Turgut Kışınbay[1]

Authorized for distribution by David Marston

June 2003

Abstract

Using realized volatility to estimate conditional variance of financial returns, we compare forecasts of volatility from linear GARCH models with asymmetric ones. We consider horizons extending to 30 days. Forecasts are compared using three different evaluation tests. With data from an equity index and two foreign exchange returns, we show that asymmetric models provide statistically significant forecast improvements upon the GARCH model for two of the datasets and improve forecasts for all datasets by means of forecasts combinations. These results extend to about 10 days in the future, beyond which the forecasts are statistically inseparable from each other.

JEL Classification Numbers: C22, C53, G15

Keywords: GARCH, high-frequency data, realized volatility, integrated volatility, and asymmetric volatility.

Author's E-Mail Address: tkisinbay@imf.org

[1] IMF and McGill University. I thank John W. Galbraith, Paul Kupiec, Matthew Thomas Jones, and seminar participants at Monetary and Financial Systems Department Seminar for comments; and Natalie Baumer and Adriana Rota for editorial assistance. The paper is based on the fourth chapter of my forthcoming McGill dissertation. Data were provided by the Toronto Stock Exchange, Market Data Services, and Olsen Associates, Switzerland. Address for correspondence: International Monetary Fund, 700 19th St. NW, Washington, D.C. 20431, U.S.A.; Tel.: 202-623-4158; Fax: 202-623-5611; e-mail: tkisinbay@imf.org

Contents | Page

I. INTRODUCTION

Volatility is central to finance. One can find many examples of the use of volatility in the theory and practice of asset pricing, optimal portfolio selection, and risk management. Naturally, the key role of volatility in finance has motivated an extensive literature on volatility modeling. Developments in modeling of conditional variance of economic time series, following Engle (1982) and many subsequent contributions, has permitted modeling and forecasting of volatility in financial markets.

Engle's original ARCH model and some of its extensions, including the popular GARCH model of Bollerslev (1986), have the property that the size of the innovations to the volatility process determines current volatility; the sign of the shock, however, does not have any significance. Thus these models are symmetric in the sense that positive and negative shocks have the same effect on conditional volatility. However, one of the established stylized facts of financial markets is that, generally, volatility after negative shocks is higher than volatility after positive shocks. Linear volatility models, such as the GARCH model, cannot capture such asymmetric effects. In this paper, we focus on the forecasting power of asymmetric models at various horizons. Given the common finding in many studies on financial market volatility that volatility responds asymmetrically to negative and positive shocks, the question remains whether we can improve our forecasts by employing asymmetric volatility models. And if we can, how far into the future can these gains be maintained?

In this paper, we will compare the forecasting performance of several asymmetric models with the standard GARCH(1,1) model. Galbraith and Kışınbay (2002), using the definitions of forecast content in Galbraith (2003), showed that the standard GARCH(1,1) model has forecast information content to a horizon of approximately thirty trading days, depending on the dataset in use. Here, we will use the GARCH(1,1) model as a benchmark model and compare the forecasting ability of other models with the benchmark model.

We aim to further previous studies in volatility forecasting literature in three ways. First, to evaluate forecasts of daily models, we use high-frequency data to obtain good estimates of volatility. Volatility is inherently unobservable, and one needs to obtain an estimate of it to evaluate model-based forecasts. In the pre-high-frequency data literature, squared daily returns were often used to obtain estimates of volatility, but as argued by Taylor and Xu (1997) and Andersen and Bollerslev (1998), among others, this strategy is not ideal for two reasons. The squared daily return is a noisy estimate of true volatility; and far better estimates of volatility can be obtained by exploiting information contained in high-frequency intraday data. The use of high-frequency data is relatively new in finance literature, and most of the previous work on forecasting with asymmetric models does not make use of such data. In this paper, we use *intraday* asset returns to *evaluate daily* asymmetric volatility model forecasts.

Second, we undertake an extensive analysis of forecasting power at various horizons and not merely focus on the one-day-ahead forecasts. For many problems in finance, volatility forecasts are relevant and necessary at various horizons. For example, investors and portfolio managers may want to have an idea about future volatility in the coming days. Some regulatory institutions enforce calculation of value-at-risk measures extending to 10 days. For option pricing, a forecast of volatility from the current date to the expiry date is required. Other examples can easily be found in the finance literature on the use of volatility forecasts at short-to-medium time horizons. Yet, much of the literature of forecasting with asymmetric models focus on one-day-ahead forecasts; only a few of them examine longer-term horizons, such as five- or ten-day-ahead forecasts, but a detailed analysis at longer horizons is generally not taken. Thus, little is known about the forecastibility of volatility with asymmetric models beyond one day or about the speed and decay of predictive power of such models as we move from short-to-medium horizons.

Finally, we carry out a detailed forecast evaluation exercise, using three different evaluation methods. First is the well-known Diebold-Mariano (1995) method, which tests for equal predictive ability of two competing forecasts. Using the Diebold-Mariano method, one can test whether one of the two competing models provide more (or less) accurate forecasts compared with the other. Thus, one can choose the best-performing model. However, choosing the best model is not necessarily the ultimate aim of the researcher. One model may perform better than the other, but the "worse" model can still provide useful information that is not contained in the best model; the so-called forecast-encompassing principle. If that is the case, predictive ability can be enhanced by combining alternative model forecasts. As noted by Poon and Granger (2003), the possibility of forecast encompassing has not been thoroughly examined in the volatility forecasting literature. We perform that analysis here, however. Finally, model forecasts are compared using regression-based forecast efficiency tests.

The paper is organized as follows. Section II provides a brief review of the hypotheses that are advanced to explain the observed asymmetry in equity return volatility. Section III presents the asymmetric models that are used to forecast volatility. Section IV provides the in-sample results for the models. Forecast evaluation techniques utilized in this study are presented in Section V, followed by the out-of-sample results in Section VI. Finally, Section VII summarizes the results and concludes the paper.

II. Theoretical Explanations of Asymmetry

There are two major theories to explain the asymmetry of volatility. The first explanation originated from Black (1976), and is called the *leverage effect*. Black reasons that when the price of a company's stock falls, its value of equity also falls. As a result, the company's leverage, or its debt-to-equity ratio, increases. Leverage is generally interpreted as an indicator of a company's riskiness: when the leverage ratio increases, the company is considered more risky, and a higher degree of risk is associated with higher volatility.

Christie (1982) tests Black's hypothesis by analyzing a cross-section of firms. He examines the relationship between debt-to-equity ratio of companies and the asymmetry of the volatility of their stock prices. Although he finds that there actually is a strong correlation between the asymmetry and leverage, the leverage itself is not sufficient to explain the asymmetric effects.

An alternative explanation for the asymmetry in stock price volatility is called the *volatility feedback hypothesis*; see Campbell and Hentschel (1992). According to this hypothesis, the causality runs from volatility to price: positive shocks to volatility increase future risk premia, and if the future dividends remain the same, then the stock price should fall.

Campbell and Hentschel find evidence in favor of their hypothesis, but they also find that the leverage effect also contributes to the asymmetric behavior of stock market volatility. Of course, the two hypotheses are not exclusive, and both effects may be present in the data. Recent developments on the asymmetric volatility literature are covered in Bekaert and Wu (2000), and Wu (2001). The former reviews the empirical literature, and provides some new evidence, while the latter summarizes the theoretical literature and provides some new theoretical results.

III. Data and Models

A. Data

Evaluation of volatility models is not straightforward. The volatility process is latent, and cannot be observed directly. Fortunately, increasing availability of high-frequency datasets allow us to obtain better measures of volatility, which can then be used to evaluate model based estimates of volatility. Use of high-frequency data to compute measures of volatility is motivated by results from the theory of continuous time finance, dating back to Merton (1980) and including many other subsequent contributions that are surveyed in Andersen, Bollerslev and Diebold (2002).

To set the stage for empirical analysis, and to present the key result of Andersen and Bollerslev (1998) that justifies the use of high-frequency data to obtain measures of volatility, we adopt a notation similar to theirs. Let p_t be the logarithm of the price of an asset at time *t*, and assume that the corresponding instantaneous returns, denoted dp_t, are generated by a continuous time martingale $dp_t = \sigma_t \cdot dW_{p,t}$, where σ_t (unobservable) denotes the instantaneous standard deviation, and $dW_{p,t}$ a standard Wiener process. Define the discretely observed series of continuously compounded returns with the sampling frequency of *m* observations per unit time as:

$$r_{m,t} = p_t - p_{t-\frac{1}{m}}.$$

With these definitions, the daily return ($m = 1$) corresponds to $r_{1,t} = p_t - p_{t-1}$, and the instantaneous return ($m \to \infty$) to $dp_t = r_{\infty,t}$. Assuming that the discrete returns are serially uncorrelated, and that the sample path for σ_t is continuous, the following result is presented in Andersen and Bollerslev (1998):

$$p\lim_{m\to\infty}\left(\int_0^1 \sigma_{t+\tau}^2 d\tau - \sum_{i=1}^{m} r_{m,t+\frac{i}{m}}^2\right) = 0 .$$

The first term in the parenthesis is called the integrated volatility and the second the realized volatility. The equation tells us that when the sampling frequency is high enough, the summation of squared returns gives us a good estimate of the latent volatility. Motivated by this result, we will use realized volatility to evaluate and compare forecasts of GARCH-type models based on daily data.

Notice that the realized volatility is an approximation to integrated quality. The theory outlined here suggests that when the sampling frequency increases, we would arrive at a better estimate of the true conditional variance. However, some of the underlying assumptions of the model are violated at very high frequencies, so the highest frequency is not necessarily the best one. Research and debates on the optimum sampling frequency continue, but generally it is recognized that sampling frequencies of less than five minutes suffer from serial correlation and various microstructure distortions. In this paper, we prefer to work with five-minute intervals to measure volatility.

We compare forecasting power of various volatility models using two types of asset: an equity price index and two currencies, priced relative to the U.S. dollar. We consider daily logarithmic returns for all the models. High-frequency, intra-day data (bid, ask and index values of last trades) to evaluate forecasts are available on the equity index at 15-second intervals, and on the foreign exchange prices (bid and ask) at five-minute intervals.

Notice that the high-frequency data are only used to evaluate daily model forecasts. One could also use high-frequency data directly to produce forecasts, as in Andersen et al. (2003) and Galbraith and Kışınbay (2002), but our aim here is to assess forecasting ability of GARCH models only. Although there is evidence that forecasts obtained from realized volatility tend to perform better than GARCH-type forecasts, we believe that the latter is still interesting in many applications. High-frequency data are not easily accessible in many cases. For example, we do not have historical high-frequency data for many financial series of interest. Moreover, high-frequency data are difficult to assemble, filter, and currently the cost of using such data is substantially higher compared to daily returns. It is likely that in the future we will experience more widespread use of high-frequency data in finance, but daily volatility models will still remain interesting and useful, and both types of models will have their place.

The equity index that is used in this study is the Toronto Stock Exchange index of 35 large capitalization stocks (TSE 35). The daily dataset used for estimating the models covers the period 1987–1998 inclusive. For evaluating forecasts, we use high frequency intra-day data for the calendar year 1998. TSE 35 is recorded every 15 seconds in each trading day, which opens at 9:30 a.m. and closes at 4:00 p.m. For each trading day, we have approximately 1560 observations.

Calculation of the realized volatility involves simply summing up the squared intra-day returns at the chosen sampling frequency, but some adjustments have to be made to correct for deficiencies in the dataset. The first adjustment we make is related to the observation that the index value is usually not within the bid and ask range in the first few minutes of the trading day. On a typical trading day, after about two minutes of trading, the index value is within the bid and ask range, and from then on the midpoint of the range and the index value are invariably close to each other. To correct this, we use the midpoint of the range for the first two minutes of a trading day to calculate the index returns.

The second adjustment we make is related to the fact that the trading does not take place around the clock, but the closing value of the index at day t, and the opening value at day $t+1$ are different from each other. This difference constitutes a contribution to the volatility that cannot be captured by summing up the intra-day returns from 9:30 a.m. to 4:00 p.m. To avoid an underestimation of daily volatility , we add the squared return from close to subsequent opening to the sum of squared intra-day returns. Galbraith and Kışınbay (2002) contains a more detailed analysis of TSE 35 data and the filtering procedures applied to the dataset.

The foreign exchange data we use are from the HFDF-2000 dataset prepared by the Olsen & Associates. We have two foreign exchange series from this dataset, namely the deutschmark-U.S. dollar (USD/DEM) and yen-U.S. dollar (USD/JPY) series. The original source for the foreign exchange raw data that are used to form the HFDF-2000 dataset are the USD/DEM and USD/JPY bid-ask quotes displayed on the Reuters FXFX screen. As in most high-frequency datasets, the data have to be filtered before empirical analysis because of various microstructure frictions, outliers and other anomalies. Müller et al. (1990) and Dacorogna et al. (1993) contain a detailed analysis of the filtering procedures, and the construction of foreign exchange returns. Our description of the HFDF-2000 dataset will be brief here, as these datasets are widely used in the literature, and their properties are analyzed in detail in Andersen et al. (2001a).

For each day , 288 five-minute returns are obtained from the HFDF-2000 dataset. Each five-minute interval in the dataset is identified by a time-stamp, accompanied by a mid-price, and a bid-ask spread. The returns are calculated as the mid-quote price difference, and then multiplied by 10,000 so as to be presented in basis points. The five-minute mid-price, in turn, is an estimate obtained from a linear interpolation between the previous and following mid-price of the irregularly spaced tick-by-tick data. The bid-ask spread is obtained simply by taking the average of the values in the last five-minute interval. When there is no quote during this interval, the mean bid-ask spread is zero.

The sample covers the period from January 2, 1987 to December 31, 1998 inclusive. There are 1,262,016 five-minute returns in the dataset, but not all of them can be used in our analysis. Further adjustments are required before empirical analysis. First, note that our analysis involves obtaining forecasts from GARCH-type models using daily data, and then evaluating and comparing the forecasts using the estimate of daily conditional variance, that is, realized volatility. To do that, we first have to define a "day;" the definition of a day we adopt is the period from 21.05 GMT of the previous calendar day to 21:00 GMT on a given day. This definition of a day is motivated from the work of Bollerslev and Domowitz (1993), and is quite common in the literature, but other definitions also exist. See, for example, Barndorff-Nielsen et al.(2002) for an alternative convention.

After defining a day, low activity days, such as weekends and holidays, are removed from the dataset, as is done in Andersen et al. (2001a), among others. The excluded days include weekends, defined as the period from Friday 21:05 GMT to Sunday 21.00 GMT; the following fixed holidays: Christmas (December 24-26), New Year's Day (December 31-January 2), and the Fourth of July; and the following moving holidays: Good Friday, Easter Monday, Memorial Day, the Fourth of July (when it falls on a weekend), Labor Day, Thanksgiving (U.S.) and the day after Thanksgiving. A final adjustment involves eliminating the technical "holes" in the recorded data, which are defined as the periods for which the indicator variable (the bid-ask spread) has 144 or more zeros.

After the entire filtering process, we are left with 2,968 trading days for USD/DEM series corresponding to 2,968x288=854,784 five-minute returns, and 2970 days for USD/JPY series corresponding to 2,970x288=855,360 five-minute returns. One last adjustment is made to the USD/JPY dataset because of the exceptionally large volatility observations occurring during the second half of 1998, dubbed as the "Once-in-a-Generation Yen Volatility" by Cai et al. (2001). We terminate this dataset at the end of May 1998, leaving 2830 days of high-frequency data.

Given the computational difficulties associated with asymmetric volatility models, and the need for an initial estimation period, we cannot use the full dataset for estimation. Instead, we reserve the first nine years of the USD/DEM and USD/JPY datasets for the initial estimation of volatility models, and the remaining three years for out-of-sample forecast evaluation. For the TSE dataset, we have a long series of daily data and only one year of high-frequency data, and its implementation is straightforward.

B. Forecasting Models

The three most commonly used asymmetric volatility models are the EGARCH model of Nelson (1991), GJR-GARCH model of Glosten, Jagannathan and Runkle (1993), and APARCH model of Ding, Granger and Engle (1993). In addition to these "core" models, we present results for another asymmetric model, the TARCH model of Zakoian (1994) and two other models that are introduced to capture thick tails of the returns processes, namely, Taylor (1986)/Schwert (1989) GARCH model (TS-GARCH) and the generalized version of the Higgins and Bera's (1992) NARCH model.

Development of GARCH models has become an industry in itself, with new models being proposed frequently. Thus, it is hard to present all the proposed models in a single essay, let alone present empirical results given the computational difficulties. In this study, we restricted ourselves to a selection of commonly used and relatively successful models, hoping that the models that are analyzed in this study would be representative of the wide variety of models that have been proposed in the literature.

We assume that daily returns are defined as $r_t = \mu + \varepsilon_t$, where $\varepsilon_t = \sigma_t z_t$ and $z_t \sim i.i.d.(0,1)$.

EGARCH Model

The earliest extension of the GARCH model that allows for asymmetric effects is the Exponential GARCH (EGARCH) model of Nelson (1991). The EGARCH(p,q) model is given by:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} \beta_i \log(\sigma_{t-i}^2) + \sum_{j=1}^{q} \left[\alpha_j \varepsilon_{t-j} + \gamma_j \left(\left|\varepsilon_{t-j}\right| - E\left|\varepsilon_{t-j}\right|\right)\right]$$

In the convenient EGARCH(1,1) specification, the weighted innovation $g(\varepsilon_t) \equiv \alpha_1 \varepsilon_t + \gamma_1 \left(\left|\varepsilon_t\right| - E\left|\varepsilon_t\right|\right)$ is introduced to capture the asymmetric relation between returns and volatility changes. From the properties of ε_t, we know that $g(\varepsilon_t)$ has mean zero and is uncorrelated. Moreover, $g(\varepsilon_t)$ captures both *size* effect (the second term of the function) and the *sign* effect (the first term of the function). Therefore α_1 is expected to be negative and γ_1 to be positive (the usual ARCH effect). The asymmetry of the innovation can also be demonstrated by rewriting it in the following form:

$$g(\varepsilon_t) = \begin{cases} (\alpha_1 + \gamma_1)\varepsilon_t - \gamma_1 E\left(\left|\varepsilon_t\right|\right) & if \;\; \varepsilon_t \geq 0 \\ (\alpha_1 - \gamma_1)\varepsilon_t - \gamma_1 E\left(\left|\varepsilon_t\right|\right) & if \;\; \varepsilon_t < 0 \end{cases}$$

The expected value of the absolute value of the error term, $E\left|\varepsilon_t\right|$, depends on the assumed distribution of the errors. In his original contribution, Nelson (1991) assumed a generalized error distribution for the errors. We assume normal distribution for our estimations, and for the normal distribution we have $E\left|\varepsilon_t\right| = \sqrt{\frac{2}{\pi}}$

Notice that the logarithm of the conditional variance is a function of past shocks and consequently the nonnegativity of it is ensured; no restrictions on the parameters α_i, β_i and γ_i have to be imposed.

GJR-GARCH Model

Glosten, Jagannathan and Runkle (1993) introduced a popular volatility model (GJR-GARCH) that allows for asymmetric effects. The model is an extension of the GARCH model where it is assumed that the parameters of squared residuals depend on the sign of the shock. The main difference from the standard model is an additional variable in the conditional variance equation equal to the product of a dummy variable S_t^- and ε_{t-i}. The general model is of the form:

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \left(\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2 \right)$$

where S_t^- is a dummy variable and is equal to 1 when $\varepsilon_{t-i} < 0$ and is equal to 0 otherwise.

For the GJR-GARCH(1,1) case, we would expect $\gamma_1 > 0$, such that a negative shock increases the conditional variance more than a positive shock of the same size. The effect of a squared shock ε_{t-1}^2 on the conditional variance depends on the sign of the shock: it is proportional to α_1 when ε_{t-1} is positive, and to $\alpha_1 + \gamma_1$ when ε_{t-1} is negative. Obviously, increasing the value of γ_1 in relation to α_1 increases the asymmetry in the model.

APARCH Model

Another model that allows asymmetric effects of positive and negative shocks on conditional volatility is proposed in Ding, Granger, and Engle (1993). The Asymmetric Power ARCH (APARCH) model is:

$$\sigma_t^\delta = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i} \right)^\delta$$

where ε_t is independently and identically distributed with unit variance, $\omega > 0$, δ, α_i and $\beta_j \geq 0$, and $-1 < \gamma_i < 1$ $(i = 1,\ldots,q)$

The main difference between the APARCH model and many other GARCH-type volatility models is the introduction of the power term, δ, as a free parameter to be estimated. In other models, δ is generally assumed to be 1 or 2, but mostly 2. Arguably, this is because of the common assumption of normally distributed errors. A well known property of normal distribution is that the whole distribution can be defined by its first two moments. Consequently, the variance is modeled often times. However, normality is rarely a good assumption in finance, and to describe the data adequately, one has to deal with skewness and kurtosis most of the time. Thus making an assumption about the power term may not be optimal; hence the potential usefulness of the APARCH model.

Asymmetry in the APARCH model is captured by the term γ. In the APARCH(1,1) case, when $\gamma > 0$, negative shocks lead to higher volatility than positive shocks, and when $\gamma < 0$, positive shock lead to higher volatility than negative shocks. Thus the usual leverage effect is captured when $\gamma > 0$.

It is important to note that, the APARCH model is a quite general model, which nests several other models we have presented; see Laurent and Peters(2002).

In addition to the three "core" asymmetric models presented above, we will also present results for the following three models.

TARCH Model

Zakoian's (1994) Threshold GARCH model has the following form:

$$\sigma_t = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j} + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i} \right)$$

Notice that the TARCH model is very similar to the GJR-GARCH model; the only difference is that in the TARCH model, standard deviation is modeled whereas in the GJR-GARCH model the variance is modeled.

TS-GARCH Model

Taylor (1986) and Schwert (1990) presented a model that is based on standard deviations, rather than variance, which is quite common on volatility modeling literature. The model is not an asymmetric model; it is introduced to capture fat thick tails that is commonly observed in data. The model is:

$$\sigma_t = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j} + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| \right)$$

N(G)ARCH Model

Higgins and Bera (1992) introduced the Nonlinear ARCH model. Here, we present its generalized version, the NGARCH model:

$$\sigma_t^{\delta} = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\delta} + \sum_{i=1}^{q} \alpha_i \left(\left| \varepsilon_{t-i} \right| \right)^{\delta}$$

Higgins and Bera's original NARCH model contained only ARCH lags, where $\beta_j = 0 \ (j = 1,2,\cdots,p)$ but the model can be easily extended to include GARCH lags, as presented above.

IV. In-Sample Results

Tables 1–3 present point estimates, robust t-statistics, log likelihoods and R^2 statistics for TSE, DEM and JPY datasets respectively. All models are estimated using the G@RCH 2.3, an OX application for estimating ARCH models; see Laurent and Peters (2002). To obtain robust inference about the estimated models, we compute the robust standard errors as suggested by Bollerslev and Wooldridge (1992). For the TSE dataset, all the coefficients are statistically significant at conventional levels. Asymmetry is clearly not rejected in any of the models. Moreover, all the asymmetry coefficients have the expected signs. We will now examine models one by one, starting with the coefficient estimates. For the EGARCH model, the parameter γ_1 has a negative sign, as expected, and it is highly significant. The asymmetry parameter of the APARCH model, γ, has the expected positive sign and is significant. In the GJR-GARCH model, asymmetry is captured by the coefficient of the dummy variable γ, and it is positive as expected. Similarly, γ is positive in the TARCH model.

Results for DEM dataset are presented in Table 2. As expected from exchange rate data, the asymmetry coefficients are not significant in these models. In EGARCH model, the asymmetry parameter is γ_1 and it is not significant. In APARCH, GJR-GARCH and TARCH models, the asymmetry coefficient is denoted by γ and is not significant in any of the models. The results for the JPY dataset, which are presented in Table 3, are similar: the asymmetry coefficients are not significant in any of the models.

Our results are consistent with the common finding in the literature that the asymmetric volatility effects are observed for the equity data, but not for the foreign exchange data. However, we will still examine their out-of-sample performance, as a good in-sample fit doesn't always transfer to good predictive ability, or vice versa.

V. Forecast Evaluation Techniques

In this section, we will present the forecast evaluation techniques that are utilized to compare the accuracy of out-of-sample forecasts of the models. To evaluate the models, we will employ three different techniques. First, we will report results for the Diebold-Mariano (DM) (DM) (1995) test for equal predictive ability. Second, we will use the forecast encompassing principle to test whether competing forecasts embody useful information that is not contained in the "best" forecast. For this, we will employ the forecast encompassing test recently suggested by Harvey, Leybourne and Newbold (HLN) (1998). The null hypothesis of the HLN test is that the model *i* forecast encompasses the model *j* forecast. That is, all the relevant information of the model *j* is contained in model *i*. A rejection of the null hypothesis suggests that the alternative model *j* forecasts contain incremental information to model *i* forecasts.

Finally, we will run the following three forecast evaluation regressions and report the regression R^2s.

i) $$(v_{t+s})^{1/2} = a_0 + a_1 \cdot (v_{t+s|GARCH})^{1/2} + u_{1,t+s}$$

ii) $$(v_{t+s})^{1/2} = b_0 + b_1 \cdot (v_{t+s|ALTERNATIVE})^{1/2} + u_{2,t+s}$$

iii) $$(v_{t+s})^{1/2} = c_0 + c_1 \cdot (v_{t+s|GARCH})^{1/2} + c_2 \cdot (v_{t+s|ALTERNATIVE})^{1/2} + u_{3,t+s},$$

where v_{t+s} is the *s*-day-ahead realized volatility at time *t*, $v_{t+s|GARCH}$ is the *s*-day-ahead GARCH(1,1) forecast at time *t*, and $v_{t+s|ALTERNATIVE}$ is the *s*-day-ahead forecast of the alternative model. The first two regressions are employed for finding out the R^2s when only one model is in use, and the third regression includes both model forecasts to find out whether a combination of the two forecasts can improve our predictive ability. Notice that we focus on the forecasts for the standard deviation rather than the variance, since the former is less sensitive to extreme outliers, which are given less weight; for discussions, see Andersen et al. (2003), and Poon and Granger (2003). For the same reason, we report results based on the absolute value of errors (mean absolute deviations), instead of using squared errors.

VI. Out-of-Sample Results

A. Results for the TSE Dataset

Figures 1–3 show the estimated relative mean absolute deviations of errors (RMAD) up to 30 trading days. Our benchmark model is the GARCH(1,1) model, and a RMAD less than one indicates that, on average, the alternative model has smaller absolute forecast errors than the benchmark model. Figures 1(a-e) contains the results for the TSE dataset for each of the models we consider in this study. A first look at the results suggests that some of the asymmetric models provide smaller mean absolute deviations (MADs) compared to the benchmark GARCH (1,1) model. For example, the APARCH model results in Figure 1e show that this model has smaller mean absolute errors than the benchmark model at all the horizons. Similarly, the EGARCH model has an RMAD less than one for up to 21 days. The TARCH model is another good performer, providing smaller forecasts errors compared to the benchmark model up to 23 days. The GJR-GARCH model provides smaller forecasts errors at almost all the horizons, but the gains in percentage terms are small, rarely exceeding 2 percent.

Finally, the TS-GARCH and NARCH model forecasts are very close to the benchmark model. Notice that these two models are not asymmetric models, but are intended to capture thick tails in the data, rather than the nonlinear effects. The result that they do not reduce forecast errors as in other models support the usefulness of asymmetric models for stock market data.

Numerical values for reductions in MADs can be observed on the vertical axis. The most promising model in terms of reductions in forecast errors is the EGARCH model, providing close to 10 percent reduction in errors in short horizons, and with good relative performance up to about 20 days. The APARCH and the TARCH models provide reductions in MADs starting from approximately 6 percent in short horizons, but then these percentage gains get smaller in longer horizons. The gains form the GJR-GARCH model are generally not more than 2 percent. The next step is to test whether these differences in MADs are significant.

Table 4 contains the results of the Diebold-Mariano tests. The test results suggest that the reductions in the MADs for the EGARCH and APARCH models are statistically significant at short horizons. For the APARCH model, the superior predictive ability is significant even at the 30-day horizon, yet the gains are moderate at longer horizons, and not significant at intermediate horizons. The TARCH model also provides statistically significant improvements in the very short term, but not at longer horizons. Other models' forecasts are very close to the GARCH model in that statistically, it is not possible to reject the hypothesis of equal predictive ability.

Table 5 reports the results of the Harvey-Leybourne-Newbold (HLN) forecast encompassing test for the TSE dataset. The first column records the p-values for the null hypothesis that the benchmark The GARCH(1,1) model encompasses the alternative model. The second column records the results of the tests for the opposite hypothesis, whether the alternative model encompasses the benchmark model or not.

Forecast encompassing tests results provide further evidence on the predictive power of asymmetric models for the TSE dataset. The null hypothesis that the benchmark model encompasses the EGARCH model is strongly rejected at short horizons, as reported in the first column of Table 5a . The second column reports the results for the null hypothesis that the EGARCH model encompasses the benchmark model, and this hypothesis cannot be rejected at any horizon. Thus in short and medium horizons the EGARCH model seems to perform significantly better than the GARCH model, according to the results of the HLN test. More specifically, this result is supported up to horizon seven ($s = 8$). Additional evidence on this is provided by the R^2s of forecast evaluation regressions. In all horizons, the R^2 of the EGARCH model is higher than that of the GARCH model. Moreover, the R^2 of the third regression, where both models are included, is rarely higher than the R^2 of the EGARCH model; therefore it seems that combining the two models' forecasts would not improve our forecasts significantly.

Similar results can be observed for the APARCH model by examining Table 5b. The significance of the better performance of the APARCH model extends to horizon 30 at the 10 percent level. APARCH model has higher R^2s than the GARCH model at all horizons, and there is little improvement in the predictive ability when forecasts are combined, as can be observed from the fifth column. Results for the GJR-GARCH model are also similar: The GJR-GARCH model encompasses the benchmark model in short horizons ($s \leq 5$), the model has higher or equal R^2s than the benchmark model, and forecast combinations provide little improvements in forecasting ability, if any.

The TS-GARCH model (Table 5d) does not perform better than the benchmark model. Generally, the benchmark model encompasses the TS-GARCH model, and has higher R^2s. At all horizons, TS-GARCH model has lower R^2s than other models we have considered so far.

The TARCH model is another successful model for this dataset. Figure 5e shows that this model has lower MADs than the benchmark model at almost all horizons. These reductions in errors are significant up to the tenth horizon. As in other models, the asymmetric model has higher R^2s than the benchmark model, with little possibility of forecast improvements from forecast combinations. Finally, the performances of the GARCH and NARCH models (Table 5f) are quite similar. Both models encompass each other at all horizons, suggesting that they have similar predictive power.

B. Results for the USD/DEM Dataset

Next, we examine the results for the exchange rate data, starting with the USD/DEM. From Figure 2, it can be readily observed that none of the models that are considered in this study improve upon the benchmark model. The alternative models have higher MADs than the benchmark model at all the horizons, and generally their performances are even worse at longer horizons. Moreover, the poor performances of the alternative models are significant, according to the DM test. The second panel of Table 4 shows that the higher MADs of the alternative models are significant at the 5 percent level[2], with the exception of the EGARCH model. For the latter model, the null hypothesis of equal predictive ability cannot be rejected.

Forecast encompassing tests for the USD/DEM data suggests that even though the alternative models have poor forecasting performance relative to the benchmark GARCH model, they can still be useful when they are used along with the benchmark model. Thus, forecast combinations can improve volatility forecasts for USD/DEM data. Evidence on this issue is

---

[2] In fact, unreported p-values show that they are generally significant at the one percent level.

provided in Table 6. It can be observed from panels 6b to 6f that generally the GARCH model does not encompass the alternative model at horizons beyond the very short term, meaning that alternative models have useful information that is not contained in the GARCH model at medium term horizons. For example, the first two columns of panel 6c, where we compare the GJR-GARCH model with the GARCH model, show that neither model encompasses the other one at any of the horizons. Thus forecasts of these two models can be combined to improve the volatility forecast in a statistically significant way. Consistent with the results of the encompassing tests, the third, fourth, and fifth columns show that the combined forecasts are generally 1 or 2 percent higher than individual forecasts. Given that the highest R^2 for the USD/DEM data is 26 percent, 1 to 2 percent gains could be economically significant.

Similar results are obtained for the APARCH, TS-GARCH, TARCH, and NARCH models: the benchmark GARCH model provides the lowest MADs, and encompass alternative models at the very short term ($s = 1$ or 2, depending on the model), but neither model encompasses each other beyond the very short-term. Forecast gains arising from combined forecasts, although generally small, can be economically significant in some cases. Lastly, the benchmark model generally encompasses the EGARCH model.

C. Results for the USD/JPY Dataset

We finally examine the results for the USD/JPY dataset. Interesting results emerge from Figures 3a-f. First, all the model forecasts are very close at short horizons; relative MADs are within the range of 1 percent for short horizons. APARCH, TARCH and NARCH models provide some statistically significant improvements in short horizons compared to the benchmark model, but the gains are small. Nevertheless, we find this result interesting. Previously, we discussed the in-sample properties of the models, and showed that the asymmetry coefficients of the models are insignificant. Yet, three of our models perform better than the more parsimonious benchmark model, a result supported by the DM test.

At longer horizons none of the models do better than the benchmark model, and in some cases the worst performance is significant.

Forecast encompassing test results for the USD/JPY datasets are reported in the third panel of Table 7(a-e). For forecast horizons less than ten-days, generally all the alternative models except for the GJR-GARCH model encompass the GARCH model, but not vice versa. At longer horizons, no model encompasses each other. For the EGARCH model, when $s \leq 7$, at the 10 percent significance level, the EGARCH model encompasses the benchmark model, but not vice versa. However, regression based comparisons suggest very similar predictive ability. There are some differences at the very short term, but beyond horizon four, they are almost always the same. Similar results hold for APARCH, and TS-GARCH models when $s \leq 10$; and for the TARCH and NARCH models when $s \leq 7$.

A comparison of GJR-GARCH model with the benchmark model reveals that neither model encompasses each other. Given the striking closeness of the relative MADs, this result is not surprising.

VII. Summary and Conclusions

The main contribution of this paper is to demonstrate the usefulness of asymmetric volatility models for forecasting at short-to-medium-term horizons. To do so, we use three different datasets: an equity index dataset, and two foreign exchange datasets.

A major difficulty in evaluation of volatility models arises from the inability to observe the volatility process. Fortunately, the increasing availability of high-frequency datasets allow us to obtain better measures of volatility, which can be employed to evaluate model-based estimates of volatility. Our contributions to the volatility forecasting literature are based on the use of such high-frequency-based volatility estimates to evaluate the forecasting power of various volatility models.

Our focus in this paper is on asymmetric GARCH-type models. We attempt to contribute to this line of research in three ways. First, as mentioned in the previous paragraph, we use a better measure of volatility to evaluate the forecasts of the asymmetric models. The use of high-frequency datasets is relatively new in the literature, and most, if not all, of the previous studies that examine the forecasting power of asymmetric models do not make use of such datasets, but rather use squared daily returns, which is a less preferred strategy.

The second contribution of the paper is to analyze the forecasting ability of asymmetric models at various horizons. For many practical purposes, forecasts beyond the one-step-ahead horizon are interesting, but there is little research on the predictive ability of volatility models (symmetric or asymmetric) beyond the very short term. Finally, we evaluate models using several alternative methods, including the use of forecast encompassing tests. Our results are as follows:

First, the results for the TSE equity index data suggest that asymmetric models are better predictors of index volatility than the standard linear GARCH model. The EGARCH, APARCH and TARCH models perform especially well. Given the good in-sample fit, and the results of previous studies in this line of research, this result is not surprising. Here, we not only confirmed the previous results by using a better measure of volatility but also showed that the superior forecasting ability of asymmetric models may extend to five-to-ten days in the future, depending on the model and forecast evaluation technique. Beyond about ten days, the forecasts are generally statistically similar, though we were able to detect significant gains using the APARCH model even at the 30-day horizon. Thus, asymmetric models improve our forecasts in the short term compared with the benchmark model, and can extend the maximum horizon where we still have some useful information that is not contained in the historical average volatility, though the evidence on the latter point is weak.

Second, we show that asymmetric volatility models can provide forecast improvements even for the foreign exchange data. For the USD/DEM dataset, we show that all the alternative models but the EGARCH model provide statistically higher MADs compared with the GARCH model, according to the Diebold-Mariano test. Yet we show that some of the alternative models still do contain some information that is not contained in the GARCH model, and they can be used along with the benchmark model and improve forecasts further. However, the gains are limited to 1 to 2 percent increases in the predictive R^2 and may or may not be economically significant.

When we turn to the USD/JPY dataset, we show that three of our models forecast statistically better than the linear model in the very short term. This result is surprising, in that asymmetry is generally not expected from foreign exchange data and in agreement with this finding, the in-sample asymmetry results are not significant. But a good in-sample fit does not always guarantee a good out-of-sample performance, and using three different forecast evaluation techniques we show that some of the alternative models provide forecasts that are significantly better than the benchmark model. As in USD/DEM dataset, however, the gains are small.

To conclude, for all the three datasets we consider, asymmetric models were useful for forecasting volatility; the gains for the exchange rate data, however, may not be economically significant.

Table 1. Parameter Estimates for the TSE Data

We report the estimates of the parameters of the models estimated using Quasi-Maximum Likelihood Estimation. The values in the parenthesis represent the robust *t*-statistics.

	GARCH	EGARCH	APARCH	GJR	TS-GARCH	TARCH	NARCH
Cst(M)	0.05	0.04	0.04	0.04	0.06	0.04	0.05
	(3.67)	(2.93)	(3.09)	(2.74)	(4.48)	(3.63)	(3.88)
Cst(V)	0.03	-0.37	0.03	0.03	0.03	0.03	0.03
	(4.84)	(-3.68)	(4.99)	(5.24)	(4.65)	(4.76)	(4.79)
GARCH(Beta1)	0.83	-0.02	0.87	0.84	0.86	0.73	0.84
	(38.64)	(-1.34)	(51.84)	(41.26)	(54.81)	(5.27)	(42.57)
GARCH(Beta2)		0.95				0.14	
		(58.67)				(1.09)	
ARCH(Alpha1)	0.12	0.91	0.11	0.06	0.13	0.12	0.13
	(7.950	(12.52)	(8.22)	(4.02)	(10.75)	(7.77)	(8.830
EGARCH(Theta1)		-0.06					
		(-6.01)					
EGARCH(Theta2)		0.19					
		(8.50)					
APARCH(Gamma1)			0.30			0.35	
			(4.68)			(5.52)	
APARCH(Delta)			1.25				1.58
			(7.15)				(7.10)
GJR(Gamma1)				0.09			
				(4.64)			
No. Observations	2772	2772	2772	2772	2772	2772	2772
No. Parameters	4	7	6	5	4	6	5
Log Likelihood	-3047.94	-3030.74	-3030.67	-3036.19	-3051.91	-3031.91	-3046.6

Table 2. Parameter Estimates for the DEM Dataset

We report the estimates of the parameters of the models estimated using Quasi-Maximum Likelihood Estimation. The values in the parenthesis represent the robust *t*-statistics.

	GARCH	EGARCH	APARCH	GJR	TS-GARCH	TARCH	NARCH
Cst(M)	-0.02	-0.01	-0.02	-0.02	-0.02	-0.02	-0.02
	(-1.26)	(-0.96)	(-1.18)	(-1.16)	(-1.08)	(-1.11)	(-1.20)
Cst(V)	0.02	-0.55	0.02	0.02	0.03	0.03	0.02
	(3.21)	(-6.59)	(3.00)	(3.11)	(3.43)	(3.22)	(3.06)
GARCH(Beta1)	0.92	0.97	0.91	0.91	0.91	0.91	0.91
	(54.29)	(93.96)	(53.31)	(50.36)	(55.69)	(53.60)	(56.14)
ARCH(Alpha1)	0.05	3.87	0.06	0.06	0.07	0.07	0.06
	(5.19)	(0.42)	(5.45)	(4.03)	(6.26)	(6.08)	(5.49)
EGARCH(Theta1)		0.00					
		(-0.04)					
EGARCH(Theta2)		0.02					
		(0.52)					
APARCH(Gamma1)			-0.01			0.03	
			(-0.09)			(0.31)	
APARCH(Delta)			1.37				1.37
			(4.87)				(5.05)
GJR(Gamma1)				-0.01			
				(-0.58)			
No. Observations	2224	2224	2224	2224	2224	2224	2224
No. Parameters	4	6	6	5	4	5	5
Log Likelihood	-2389.34	-2385	-2387.27	-2389.16	-2388.36	-2388.31	-2387.27

Table 3. Parameter Estimates for the JPY Dataset

We report the estimates of the parameters of the models estimated using Quasi-Maximum Likelihood Estimation. The values in the parentheses represent the robust t-statistics.

	GARCH	EGARCH	APARCH	GJR	TS-GARCH	TARCH	NARCH
Cst(M)	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(-1.16)	(-0.89)	(-0.99)	(-1.02)	(-0.84)	(-1.19)	(-1.04)
Cst(V)	0.02	-0.55	0.02	0.02	0.03	0.03	0.02
	(3.58)	(-6.50)	(3.13)	(3.54)	(3.60)	(3.49)	(3.10)
GARCH(Beta1)	0.90	0.19	0.90	0.90	0.89	0.89	0.90
	(49.90)	(0.88)	(46.95)	(47.92)	(42.43)	(42.56)	(47.71)
GARCH(Beta2)		0.72					
		(3.53)					
ARCH(Alpha1)	0.06	3.30	0.07	0.07	0.09	0.08	0.07
	(5.44)	(0.81)	(4.87)	(4.59)	(6.13)	(6.09)	(4.98)
EGARCH(Theta1)		0.00					
		(-0.58)					
EGARCH(Theta2)		0.07					
		(0.89)					
APARCH(Gamma1)			-0.02			0.04	
			(-0.35)			(0.51)	
APARCH(Delta)			1.62				1.58
			(5.48)				(5.73)
GJR(Gamma1)				-0.01			
				(-0.81)			
No. Observations	2227	2227	2227	2227	2227	2227	2227
No. Parameters	4	7	6	5	4	5	5
Log Likelihood	-2315.74	-2314.99	-2314.63	-2315.4	-2316.99	-2316.86	-2314.69

Table 4. Relative Mean Absolute Value of Forecast Errors

Table 4a. TSE 35

Horizon (s)	GARCH MAD	EGARCH RMAD	APARCH RMAD	GJR RMAD	TS-GARCH RMAD	TARCH RMAD	NARCH RMAD
1	1.12	0.91*	0.94*	0.98*	1.01**	0.94*	1.00
2	1.19	0.92*	0.95**	1.00	1.00	0.95	0.99
3	1.24	0.91*	0.95**	0.96*	1.00	0.94**	1.00
4	1.28	0.93**	0.96**	0.99	1.00	0.96	1.00
5	1.27	0.93*	0.95**	0.99	0.99	0.95	0.99
6	1.31	0.93**	0.96	1.00	0.98	0.96	0.99
7	1.29	0.94**	0.97	1.00	0.99	0.97	0.99
8	1.26	0.93	0.97	0.99	0.99	0.97	0.99
9	1.25	0.95	0.97	0.99	1.00	0.98	1.00
10	1.30	0.93**	0.96**	0.98**	0.98	0.95	0.99
15	1.31	0.96	0.97**	0.98	1.00	0.97	1.00
20	1.28	1.00	0.99**	0.99	1.01	0.99	1.00
25	1.29	1.02	0.99**	0.99	1.02	1.00	1.01
30	1.34	1.00	0.98**	0.99	1.01	0.99	1.00

1/ TSE 35 represents Toronto Stock Exchange index of 35 large capitalization stocks.

Table 4b. USD/DEM

Horizon (s)	GARCH MAD	EGARCH RMAD	APARCH RMAD	GJR RMAD	TS-GARCH RMAD	TARCH RMAD	NARCH RMAD
1	0.15	1.01	1.01	1.01*	1.02*	1.02	1.01
2	0.15	0.99**	1.01	1.01*	1.03*	1.02**	1.01
3	0.16	0.99	1.01	1.01*	1.04*	1.03*	1.01
4	0.16	0.99	1.02**	1.01*	1.04*	1.03*	1.01
5	0.17	0.99	1.02*	1.01*	1.05*	1.04*	1.02*
6	0.17	0.98	1.03*	1.01*	1.05*	1.04*	1.02*
7	0.17	0.99	1.03*	1.01*	1.06*	1.05*	1.03*
8	0.18	0.99	1.03*	1.01*	1.07*	1.06*	1.03*
9	0.18	1.00	1.04*	1.01*	1.07*	1.06*	1.04*
10	0.18	1.00	1.04*	1.01*	1.07*	1.07*	1.04*
15	0.19	1.01	1.05*	1.01*	1.10*	1.09*	1.05*
20	0.20	1.01	1.06*	1.02*	1.11*	1.10*	1.06*
25	0.21	1.02	1.07*	1.01*	1.13*	1.12*	1.07*
30	0.22	1.03	1.08*	1.01*	1.15*	1.14*	1.08*

Table 4c. USD/JPY

Horizon (s)	GARCH MAD	EGARCH RMAD	APARCH RMAD	GJR RMAD	TS-GARCH RMAD	TARCH RMAD	NARCH RMAD
1	0.23	1.01	0.99*	1.00	0.99	0.98*	0.99*
2	0.23	1.00	0.99**	1.00	0.99	0.99	0.99**
3	0.24	1.00	0.99	1.00	0.99	1.00	1.00
4	0.24	1.00	1.00	1.00	1.00	1.00	1.00
5	0.24	1.01	1.00	1.00	1.01	1.01	1.00
6	0.24	1.01	1.00	1.00	1.01	1.01	1.00
7	0.24	1.01	1.00	1.00	1.02	1.02	1.00
8	0.24	1.01	1.00	1.00	1.02	1.02**	1.00
9	0.24	1.01	1.00	1.00	1.02	1.02	1.00
10	0.24	1.01	1.00	1.00	1.03**	1.02**	1.00
15	0.25	1.02	1.01	1.00	1.04*	1.04*	1.01
20	0.25	1.02	1.01	1.00	1.05**	1.05**	1.01
25	0.25	1.03	1.02	1.00	1.07**	1.07*	1.02
30	0.25	1.04	1.02	1.00	1.08**	1.08**	1.02

Notes: We report the relative mean absolute value (RMAD) of forecast errors for each model at horizons up to 30 days. First column reports mean standard deviation of errors for the benchmark GARCH(1,1) model, and remaining columns report relative mean standard deviation of errors for the alternative models. Diebold-Mariano test results are used to test for equal predictive ability. A * (**) indicates that the differences between the benchmark model and the alternative model are significant at the 5 (10 percent) level.

Table 5. Forecast Encompassing Tests TSE 35

Table 5a. EGARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.42	0.50	0.52	0.52
2	0.00	0.32	0.44	0.46	0.46
3	0.00	0.36	0.41	0.44	0.44
4	0.00	0.40	0.36	0.39	0.39
5	0.01	0.25	0.35	0.39	0.39
6	0.02	0.22	0.32	0.35	0.35
7	0.06	0.14	0.33	0.35	0.35
8	0.10	0.12	0.32	0.34	0.34
9	0.12	0.14	0.31	0.33	0.33
10	0.14	0.13	0.28	0.31	0.31
15	0.17	0.12	0.20	0.22	0.22
20	0.44	0.16	0.15	0.17	0.18
25	0.37	0.15	0.10	0.11	0.11
30	0.39	0.15	0.03	0.06	0.07

Table 5b. APARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.08	0.50	0.53	0.53
2	0.01	0.17	0.44	0.47	0.47
3	0.01	0.07	0.41	0.44	0.45
4	0.02	0.10	0.36	0.39	0.40
5	0.03	0.14	0.35	0.39	0.40
6	0.03	0.13	0.32	0.35	0.36
7	0.03	0.35	0.33	0.35	0.35
8	0.07	0.31	0.32	0.34	0.34
9	0.06	0.28	0.31	0.33	0.33
10	0.05	0.22	0.28	0.31	0.32
15	0.10	0.36	0.20	0.22	0.23
20	0.10	0.35	0.15	0.18	0.19
25	0.13	0.49	0.10	0.12	0.13
30	0.07	0.08	0.03	0.05	0.10

Table 5c. GJR-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.12	0.50	0.53	0.54
2	0.01	0.26	0.44	0.46	0.47
3	0.02	0.24	0.41	0.43	0.44
4	0.06	0.44	0.36	0.37	0.37
5	0.08	0.49	0.35	0.36	0.36
6	0.17	0.35	0.32	0.32	0.32
7	0.25	0.24	0.33	0.33	0.33
8	0.20	0.31	0.32	0.32	0.32
9	0.15	0.42	0.31	0.31	0.31
10	0.11	0.49	0.28	0.30	0.30
15	0.22	0.35	0.20	0.22	0.22
20	0.32	0.30	0.15	0.17	0.19
25	0.49	0.21	0.10	0.12	0.15
30	0.39	0.22	0.03	0.05	0.09

Table 5d. TS-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.27	0.05	0.50	0.48	0.50
2	0.24	0.11	0.44	0.42	0.44
3	0.20	0.02	0.41	0.39	0.41
4	0.07	0.46	0.36	0.36	0.36
5	0.13	0.40	0.35	0.36	0.36
6	0.13	0.45	0.32	0.33	0.33
7	0.16	0.17	0.33	0.33	0.33
8	0.23	0.36	0.32	0.32	0.32
9	0.26	0.27	0.31	0.31	0.31
10	0.26	0.37	0.28	0.29	0.29
15	0.31	0.28	0.20	0.20	0.20
20	0.33	0.37	0.15	0.14	0.15
25	0.35	0.44	0.10	0.09	0.11
30	0.22	0.34	0.03	0.03	0.03

Table 5e. TARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.21	0.50	0.53	0.53
2	0.01	0.31	0.44	0.46	0.46
3	0.01	0.08	0.41	0.44	0.44
4	0.02	0.12	0.36	0.40	0.41
5	0.04	0.18	0.35	0.39	0.40
6	0.04	0.16	0.32	0.36	0.37
7	0.03	0.38	0.33	0.35	0.35
8	0.10	0.35	0.32	0.35	0.35
9	0.10	0.33	0.31	0.33	0.34
10	0.09	0.27	0.28	0.32	0.32
15	0.15	0.35	0.20	0.22	0.23
20	0.17	0.32	0.15	0.17	0.18
25	0.23	0.38	0.10	0.11	0.11
30	0.15	0.18	0.03	0.05	0.09

Table 5f. NARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.25	0.31	0.50	0.49	0.50
2	0.25	0.37	0.44	0.44	0.44
3	0.21	0.45	0.41	0.40	0.41
4	0.07	0.11	0.36	0.36	0.36
5	0.13	0.23	0.35	0.36	0.36
6	0.12	0.20	0.32	0.32	0.33
7	0.15	0.34	0.33	0.33	0.33
8	0.24	0.36	0.32	0.32	0.32
9	0.26	0.40	0.31	0.31	0.31
10	0.26	0.37	0.28	0.29	0.29
15	0.33	0.49	0.20	0.20	0.20
20	0.33	0.47	0.15	0.15	0.15
25	0.37	0.47	0.10	0.10	0.11
30	0.22	0.27	0.03	0.03	0.03

Notes: Results of the forecast encompassing test of Harvey, Leybourne, and Newbold (1998) are reported. The first column records the results for the null hypothesis that the GARCH(1,1) model encompasses the alternative model. The second column tests whether the alternative model encompasses the GARCH(1,1) model. The third column is the coefficient of determination for the GARCH model, the fourth column for the alternative model, and the fifth column for the two models together.

Table 6. Forecast Encompassing Tests USD/DEM

Table 6a. EGARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.22	0.00	0.26	0.18	0.28
2	0.06	0.01	0.19	0.16	0.20
3	0.05	0.03	0.17	0.14	0.17
4	0.11	0.03	0.16	0.13	0.16
5	0.26	0.03	0.14	0.11	0.15
6	0.18	0.10	0.12	0.10	0.12
7	0.23	0.09	0.11	0.09	0.12
8	0.25	0.11	0.09	0.08	0.10
9	0.36	0.10	0.09	0.06	0.10
10	0.35	0.12	0.08	0.06	0.08
15	0.38	0.23	0.08	0.06	0.09
20	0.49	0.23	0.08	0.07	0.08
25	0.30	0.13	0.08	0.07	0.08
30	0.09	0.04	0.05	0.05	0.05

Table 6b. APARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.27	0.01	0.26	0.23	0.28
2	0.09	0.00	0.19	0.17	0.21
3	0.05	0.00	0.17	0.14	0.18
4	0.02	0.00	0.16	0.13	0.18
5	0.01	0.00	0.14	0.11	0.17
6	0.00	0.00	0.12	0.10	0.15
7	0.00	0.00	0.11	0.09	0.15
8	0.00	0.00	0.09	0.08	0.12
9	0.00	0.00	0.09	0.07	0.11
10	0.00	0.00	0.08	0.06	0.10
15	0.00	0.00	0.08	0.07	0.10
20	0.00	0.00	0.08	0.07	0.10
25	0.00	0.00	0.08	0.06	0.09
30	0.00	0.00	0.05	0.04	0.05

Table 6c. GJR-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.00	0.26	0.25	0.28
2	0.00	0.00	0.19	0.19	0.21
3	0.00	0.00	0.17	0.16	0.18
4	0.00	0.00	0.16	0.15	0.17
5	0.00	0.00	0.14	0.13	0.15
6	0.00	0.00	0.12	0.11	0.13
7	0.00	0.00	0.11	0.11	0.13
8	0.00	0.00	0.09	0.09	0.11
9	0.00	0.00	0.09	0.08	0.10
10	0.00	0.00	0.08	0.07	0.08
15	0.00	0.00	0.08	0.08	0.09
20	0.00	0.00	0.08	0.08	0.09
25	0.00	0.00	0.08	0.07	0.08
30	0.00	0.00	0.05	0.04	0.05

Table 6d. TS-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.14	0.00	0.26	0.22	0.27
2	0.04	0.00	0.19	0.16	0.21
3	0.02	0.00	0.17	0.13	0.18
4	0.00	0.00	0.16	0.12	0.17
5	0.00	0.00	0.14	0.10	0.16
6	0.00	0.00	0.12	0.09	0.15
7	0.00	0.00	0.11	0.08	0.14
8	0.00	0.00	0.09	0.07	0.12
9	0.00	0.00	0.09	0.06	0.11
10	0.00	0.00	0.08	0.05	0.10
15	0.00	0.00	0.08	0.06	0.10
20	0.00	0.00	0.08	0.06	0.09
25	0.00	0.00	0.08	0.06	0.08
30	0.00	0.00	0.05	0.04	0.05

Table 6e. TARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.46	0.00	0.26	0.22	0.27
2	0.16	0.00	0.19	0.16	0.20
3	0.07	0.00	0.17	0.13	0.18
4	0.02	0.00	0.16	0.12	0.17
5	0.01	0.00	0.14	0.10	0.16
6	0.00	0.00	0.12	0.09	0.14
7	0.00	0.00	0.11	0.08	0.14
8	0.00	0.00	0.09	0.07	0.11
9	0.00	0.00	0.09	0.06	0.10
10	0.00	0.00	0.08	0.05	0.09
15	0.00	0.00	0.08	0.06	0.09
20	0.00	0.00	0.08	0.06	0.09
25	0.00	0.00	0.08	0.06	0.08
30	0.00	0.00	0.05	0.04	0.05

Table 6f. NARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.43	0.02	0.26	0.23	0.28
2	0.14	0.01	0.19	0.17	0.21
3	0.08	0.00	0.17	0.14	0.18
4	0.04	0.00	0.16	0.13	0.18
5	0.02	0.00	0.14	0.11	0.17
6	0.01	0.00	0.12	0.10	0.15
7	0.00	0.00	0.11	0.09	0.15
8	0.00	0.00	0.09	0.08	0.12
9	0.00	0.00	0.09	0.07	0.11
10	0.00	0.00	0.08	0.06	0.10
15	0.00	0.00	0.08	0.07	0.10
20	0.00	0.00	0.08	0.07	0.10
25	0.00	0.00	0.08	0.06	0.09
30	0.00	0.00	0.05	0.04	0.05

Notes: Results of the forecast encompassing test of Harvey, Leybourne, and Newbold (1998) are reported. The first column records the results for the null hypothesis that the GARCH(1,1) model encompasses the alternative model. The second column tests whether the alternative model encompasses the GARCH(1,1) model. The third column is the coefficient of determination for the GARCH model, the fourth column for the alternative model, and the fifth column for the two models together.

Table 7. Forecast Encompassing Tests USD/JPY

Table 7a. EGARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.06	0.00	0.21	0.19	0.21
2	0.01	0.20	0.16	0.17	0.17
3	0.00	0.31	0.15	0.16	0.16
4	0.04	0.18	0.15	0.15	0.15
5	0.16	0.11	0.13	0.12	0.13
6	0.06	0.28	0.13	0.13	0.13
7	0.08	0.29	0.13	0.13	0.14
8	0.15	0.20	0.12	0.11	0.12
9	0.12	0.19	0.10	0.10	0.10
10	0.08	0.31	0.09	0.09	0.10
15	0.24	0.22	0.09	0.09	0.09
20	0.26	0.22	0.10	0.11	0.11
25	0.31	0.28	0.07	0.09	0.09
30	0.40	0.21	0.06	0.08	0.08

Table 7b. APARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.11	0.21	0.22	0.22
2	0.01	0.08	0.16	0.17	0.17
3	0.02	0.14	0.15	0.16	0.16
4	0.08	0.25	0.15	0.15	0.15
5	0.11	0.32	0.13	0.13	0.13
6	0.08	0.24	0.13	0.13	0.13
7	0.10	0.28	0.13	0.14	0.14
8	0.10	0.30	0.12	0.12	0.12
9	0.09	0.27	0.10	0.10	0.10
10	0.10	0.29	0.09	0.10	0.10
15	0.21	0.45	0.09	0.09	0.10
20	0.28	0.45	0.10	0.10	0.10
25	0.29	0.49	0.07	0.08	0.10
30	0.35	0.41	0.06	0.07	0.10

Table 7c. GJR-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.35	0.22	0.21	0.21	0.22
2	0.34	0.48	0.16	0.16	0.17
3	0.26	0.41	0.15	0.15	0.15
4	0.29	0.42	0.15	0.14	0.15
5	0.24	0.36	0.13	0.13	0.13
6	0.18	0.29	0.13	0.13	0.13
7	0.19	0.29	0.13	0.13	0.13
8	0.12	0.20	0.12	0.12	0.12
9	0.10	0.17	0.10	0.10	0.10
10	0.15	0.26	0.09	0.09	0.09
15	0.22	0.33	0.09	0.09	0.09
20	0.43	0.29	0.10	0.10	0.10
25	0.43	0.49	0.07	0.07	0.07
30	0.45	0.38	0.06	0.07	0.07

Table 7d. TS-GARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.33	0.21	0.23	0.23
2	0.00	0.29	0.16	0.18	0.18
3	0.01	0.40	0.15	0.16	0.16
4	0.04	0.42	0.15	0.15	0.15
5	0.08	0.28	0.13	0.12	0.13
6	0.06	0.36	0.13	0.13	0.13
7	0.09	0.27	0.13	0.13	0.14
8	0.12	0.19	0.12	0.11	0.12
9	0.10	0.22	0.10	0.10	0.10
10	0.10	0.21	0.09	0.09	0.10
15	0.24	0.13	0.09	0.09	0.09
20	0.29	0.09	0.10	0.11	0.11
25	0.32	0.13	0.07	0.09	0.10
30	0.39	0.10	0.06	0.09	0.10

Table 7e. TARCH vs. GARCH

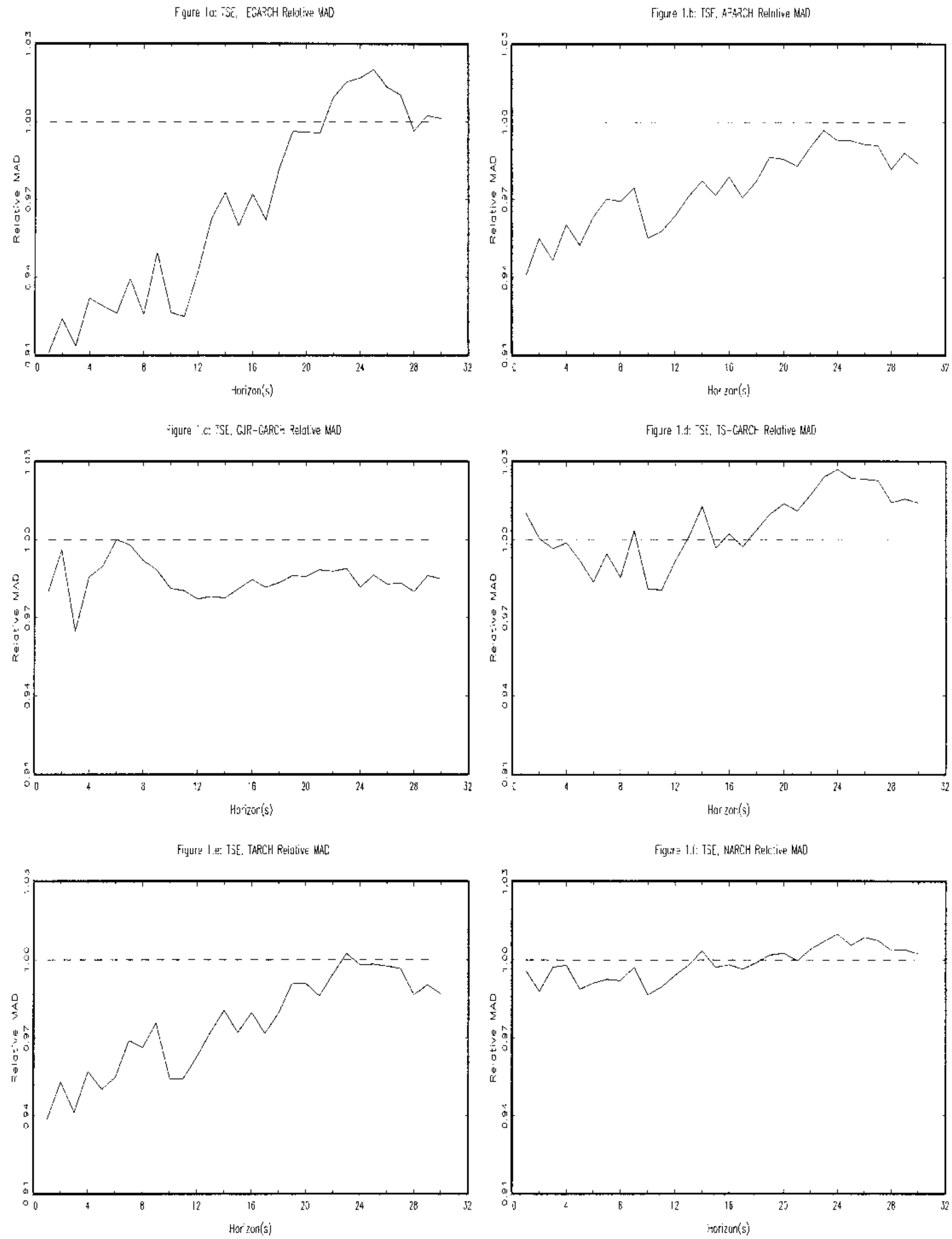
Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.20	0.21	0.23	0.23
2	0.00	0.21	0.16	0.18	0.18
3	0.01	0.36	0.15	0.16	0.16
4	0.04	0.45	0.15	0.15	0.15
5	0.09	0.29	0.13	0.13	0.13
6	0.08	0.36	0.13	0.13	0.13
7	0.11	0.27	0.13	0.13	0.13
8	0.14	0.18	0.12	0.11	0.12
9	0.12	0.21	0.10	0.10	0.10
10	0.12	0.22	0.09	0.09	0.10
15	0.25	0.14	0.09	0.09	0.09
20	0.29	0.09	0.10	0.11	0.11
25	0.32	0.13	0.07	0.09	0.10
30	0.38	0.10	0.06	0.09	0.10

Table 7f. NARCH vs. GARCH

Horizon (s)	$H_{G,A}$	$H_{A,G}$	Rg^2	Ra^2	Rc^2
1	0.00	0.03	0.21	0.22	0.22
2	0.00	0.04	0.16	0.17	0.17
3	0.01	0.10	0.15	0.16	0.16
4	0.06	0.22	0.15	0.15	0.15
5	0.11	0.33	0.13	0.13	0.13
6	0.09	0.26	0.13	0.13	0.13
7	0.11	0.31	0.13	0.13	0.13
8	0.13	0.35	0.12	0.12	0.12
9	0.11	0.32	0.10	0.10	0.10
10	0.11	0.32	0.09	0.10	0.10
15	0.23	0.47	0.09	0.09	0.09
20	0.27	0.46	0.10	0.10	0.11
25	0.28	0.48	0.07	0.08	0.10
30	0.34	0.40	0.06	0.07	0.10

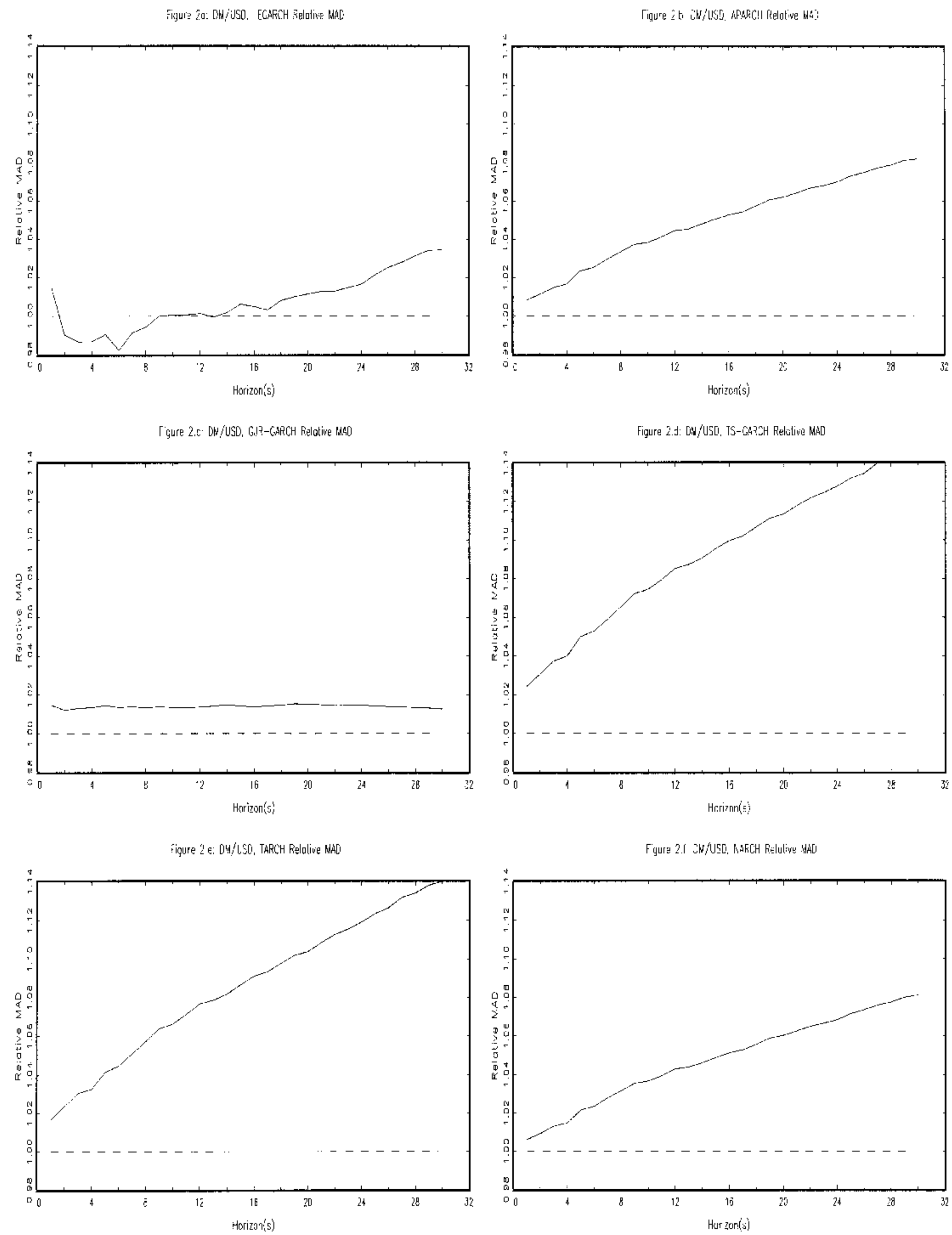
Notes: Results of the forecast encompassing test of Harvey, Leybourne, and Newbold (1998) are reported. The first column records the results for the null hypothesis that the GARCH(1,1) model encompasses the alternative model. The second column tests whether the alternative model encompasses the GARCH(1,1) model. The third column is the coefficient of determination for the GARCH model, the fourth column for the alternative model, and the fifth column for the two models together.

Figure 1. Relative Mean Absolute Deviations - TSE



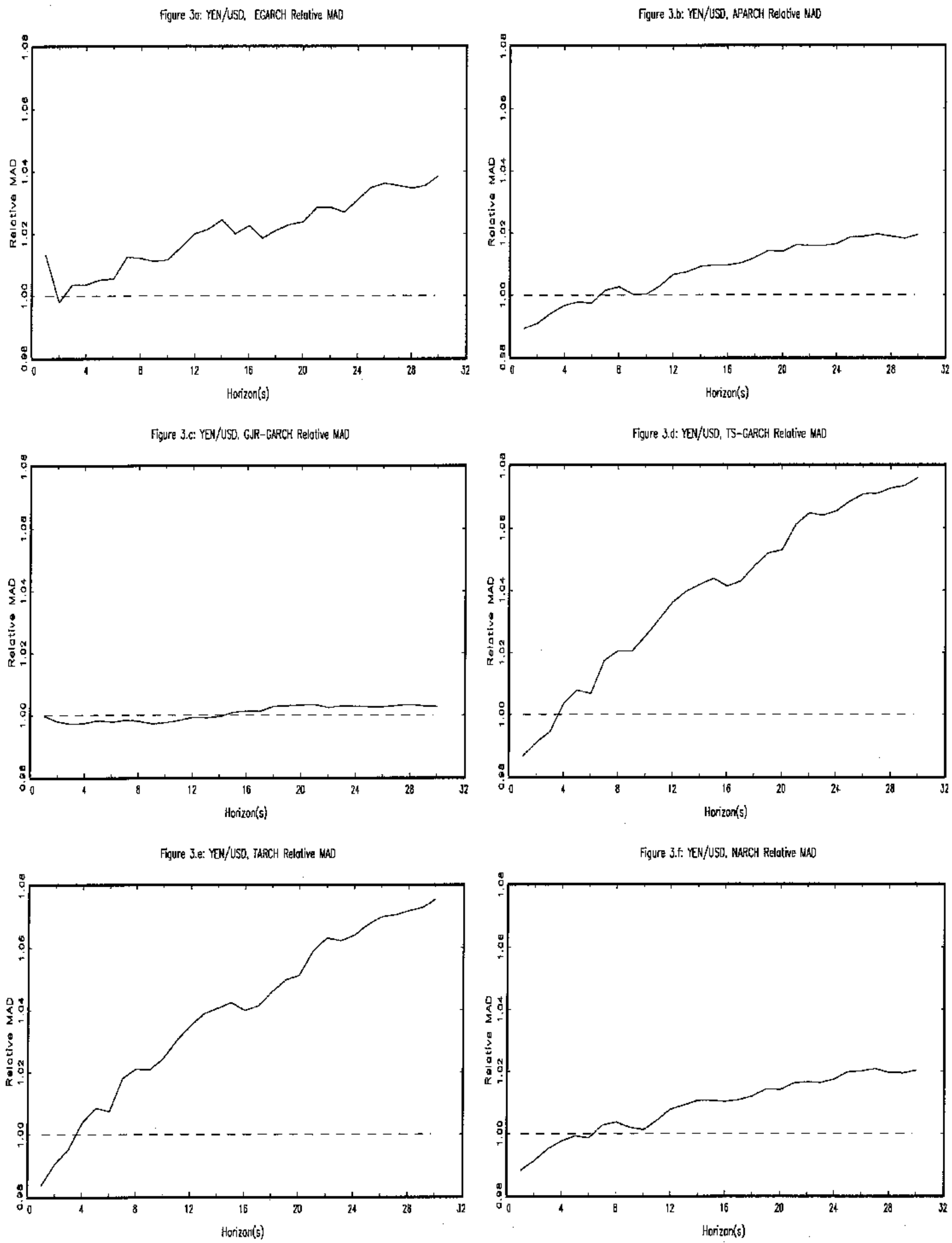
1/ TSE – Toronto Stock Exchange
2/ MAD – Mean Absolute Deviation

Figure 2. Relative Mean Absolute Deviations - USD/DEM



1/ MAD – Mean Absolute Deviation

Figure 3. Relative Mean Absolute Deviations - USD/JPY



1/ MAD – Mean Absolute Deviation

References

Andersen, T.G. and T. Bollerslev, 1998, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review* 39(4), pp. 885–905.

Andersen, T.G., T. Bollerslev, F.X. Diebold, 2002, "Parametric and Nonparametric Volatility Measurement," NBER Technical Working Paper, No. 279.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001a, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* 96, pp. 42–55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001b, "Exchange Rate Returns Standardized by Realized Volatility are Nearly Gaussian," *Multinational Finance Journal,* Vol. 4, pp.159–79.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica,* Vol. 71, pp. 579–625.

Barndorff-Nielsen, O.E., B. Nielsen, N. Shephard, and C. Ysusi, 2002, "Measuring and Forecasting Financial Variability Using Realized Variance With and Without a Model," Working paper, (Oxford, England: Oxford University).

Bekaert, G. and G. Wu 2000, "Asymmetric Volatility and Risk in Equity Markets," *The Review of Financial Studies,* Vol.13, pp.1–42.

Black, French 1976, "Studies of Stock Market Volatility Changes," *1976 Proceedings of the American Statistical Association, Business and Economics Section*, pp. 177–81.

Bollerslev, T. 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics,* Vol. 31, pp. 307–327.

Bollerslev, T. and I. Domowitz 1993, "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *Journal of Finance,* Vol. 48, pp. 1421–43.

Bollerslev, T. and J. Wooldridge 1992, "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews,* Vol.11, pp.143–172.

Cai, J.Y., Y.L. Cheung, R.S.K. Lee, and M. Melvin 2001, "'Once-in-a-Generation' Yen Volatility in 1998: Fundamentals, Intervention and Order Flow," *Journal of International Money and Finance,* Vol. 20, pp. 327–47.

Campbell, J., and L. Hentschel 1992, "No News is Good News: an Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics,* Vol.31, pp. 281-318.

Christie, A. 1982, "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics,* Vol.10, pp. 407-432.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen, and O.V. Pictet 1993, "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance,* Vol. 12, pp. 413–438.

Diebold, F. X. and R. Mariano 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* Vol. 13, pp. 253–263.

Ding, Z., C.W.J. Granger, R.F. Engle,1993, "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance,* Vol. 1, pp. 83–106.

Engle, R.F.,1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol. 50, pp. 987–1007.

Galbraith, J.W., 2003, "Content Horizons for Univariate Time Series Forecasts." *International Journal of Forecasting,* Vol. 19, pp. 43-55.

Galbraith, J.W. and T. Kışınbay 2002, "Information Content of Volatility Forecasts at Medium-term Horizons," CIRANO Working Paper, 2002–21.

Glosten, L.R., R. Jagannathan, and D. Runkle 1993, "On the Relationship Between GARCH and Symmetric Stable Processes: Finding the Source of Fat Tails in Data," *Journal of Finance,* Vol. 48, pp.1779–1802.

Harvey, I. D., S. J. Leybourne, and P. Newbold 1998, "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics,* Vol.16, pp.254-59.

Higgins, M. and A. Bera 1992, "A Class of Nonlinear ARCH Models," *International Economic Review,* Vol. 33, pp.137–158.

Laurent, S. and J-P. Peters 2002, A Tutorial for G@RCH 2.3, a Complete Ox Package for Estimating and Forecasting ARCH Models," Working Paper, (Liège, Belgium: Faculté d'Economie, de Gestion et de Sciences Sociales, Université de Liège).

Merton, R. 1980, "On Estimating the Expected Return on the Market," *Journal of Financial Economics,* Vol. 8, pp.323–361.

Műller, U.A., M.M. Dacorogna, R.B. Olsen, O.V. Pictet, M. Schwarz, and C. Morgenegg, 1990, "Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Scaling Law, and Intraday Analysis" *Journal of Banking and Finance,* Vol.14, pp.1189–1208.

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

Poon, S. and C. W. J. Granger 2003, "Forecasting Volatility in Financial Markets: A Review," Working Paper, (Strathclyde, Scotland: University of Strathclyde), Forthcoming in *Journal of Economic Literature*.

Schwert, W. 1990, "Stock Volatility and the Crash of '87," *Review of Financial Studies* 3, pp. 77-102.

Taylor, S.J, 1986, *Modeling Financial Time Series* (New York: Wiley & Sons).

Taylor, S.J. and X. Xu 1997, "The Incremental Volatility Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance,* Vol. 4, pp. 317-340.

Wu, G. 2001, "The Determinants of Asymmetric Volatility," *The Review of Financial Studies,* Vol.14, pp. 837–859.

Zakoian J.M. 1994, "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control,* Vol.18, pp. 931–995.