TECHNICAL WORKING PAPER SERIES

THE PREDICTIVE ABILITY OF SEVERAL MODELS OF EXCHANGE RATE VOLATILITY

Kenneth D. West
Dongchul Cho

Technical Working Paper No. 152

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1994

This research is an outgrowth of work conducted jointly with Hali J. Edison, to whom we are greatly indebted. We also thank Neil Ericcson, Blake LeBaron, Wen-Ling Lin, three anonymous referees and an Associate Editor of this journal, and participants in a seminar at the Federal Reserve Board of Governors for helpful comments and discussions. West thanks the National Science Foundation, the Sloan Foundation, and the University of Wisconsin Graduate School for financial support. This paper is part of NBER's research programs in Asset Pricing and International Finance and Macroeconomics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #152
January 1994

THE PREDICTIVE ABILITY OF SEVERAL MODELS OF EXCHANGE RATE VOLATILITY

ABSTRACT

We compare the out-of-sample forecasting performance of univariate homoskedastic, GARCH, autoregressive and nonparametric models for conditional variances, using five bilateral weekly exchange rates for the dollar, 1973-1989. For a one week horizon, GARCH models tend to make slightly more accurate forecasts. For longer horizons, it is difficult to find grounds for choosing between the various models. None of the models perform well in a conventional test of forecast efficiency.

Kenneth D. West
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706
and NBER

Dongchul Cho
Department of Economics
Texas A and M University
College Station, TX 77843-4428

1. Introduction

This paper compares the forecasting performance of six models for a univariate conditional variance, using bilateral weekly data for the dollar versus the currencies of Canada, France, Germany, Japan and the United Kingdom, 1973-1989. The six models include a homoskedastic one, two GARCH models (Bollerslev (1986), Engle and Bollerslev (1986)), two autoregressions and a nonparametric one. We compare the out-of-sample realization of the square of the weekly change in an exchange rate with the value predicted by a model of the conditional variance, for horizons of one, twelve and twenty four weeks. The measure of performance that we focus on is mean squared prediction error (MSPE).

For twelve and twenty four week ahead forecasts of the squared weekly change, it is difficult to find grounds to choose one model over another. But at a one week horizon, we find that GARCH models have a slight edge over the other models. The GARCH mean squared prediction errors tend to be slightly smaller, and regressions of realized exchange rate squares on their estimated conditional variances tend to find somewhat more evidence of predictive power. But statistical tests typically cannot reject at conventional significance levels the null that the MSPE from the GARCH models are equal to those of other models, and standard regression tests for bias and efficiency strongly reject the null that the GARCH conditional variance differs from the realized exchange rate square by a white noise error. It appears that GARCH models leave something to be desired, even at the one week horizon.

Other papers have compared univariate volatility models in related frameworks. Using monthly stock return data and a one month ahead MSPE criterion, Akigray (1989) found GARCH models preferable to naive and ARMA ones, and Pagan and Schwert (1990a) found GARCH and ARMA models preferable to nonparametric and Markov switching ones. While we, too, find that GARCH models perform well, our results complement and extend these earlier ones in three ways.

First, and least important, we use exchange rate instead of stock price data. One would obviously like to know if what works with one type of data works with another as well. Second, we formally test for equality of MSPEs across models, using a straightforward asymptotic technique that may be of general interest. Third, we consider not only one period but multiperiod horizons as well. Since, in the end, we could not reject the null of equality of MSPEs across models, and since we found no grounds for preferring one estimator over another at horizons of more than one period, our endorsement of GARCH models is more moderate than it would have been had we not performed these tests and examined these horizons.

Before turning to the analysis, a final remark seems advisable. The literature on conditional volatility has grown enormously in recent years (see Bollerslev et al. (1992) for an excellent survey), and it is simply not practical to simultaneously study every model that has been proposed. While we feel that we have chosen a representative set of models, we recognize that some readers might prefer a different set. We hope that such readers will nonetheless find it useful that our analysis leads us to speculate that successful models will allow for what standard tests suggest is movement in <u>un</u>conditional variances.[1]

Section 2 describes our data and models, sections 3 and 4 our empirical results. Section 5 concludes. An Appendix available on request from the authors contains some results omitted from the paper to save space.

2. Data, Models, and Estimation Techniques

A. Data

Our exchange rates are measured as dollars per unit of foreign currency, between the U.S. and Canada, France, Germany, Japan and the United Kingdom.[2] The data are Wednesday, New York noon bid rates, as published in The Federal Reserve Bulletin. When Wednesday was a holiday we used Thursday data; when Thursday was a holiday as well we used Tuesday data. After an initial observation was lost due to differencing (see below), the sample for each country included the 863 observations from March 14, 1973, to September 20, 1989. Figures 1A1 to 1A5 plot the levels rather than differences of the series, with the vertical axis measured in cents per unit of foreign currency. Figures 1B1 to 1B5 will be discussed below.

Prior to our formal analysis, we took logarithmic differences of the series, and then multiplied by 100. That is, our exchange rate series is

$$e_t = 100*\ln(\text{exchange rate in week } t/\text{exchange rate in week } t\text{-}1),$$

and thus has the interpretation of percentage change in the level of the exchange rate. With a slight abuse of terminology, we will sometimes refer to our data as "exchange rates" rather than "percentage changes in exchange rates."

Table 1 contains some summary statistics on these data. Most standard errors and p-values in the remainder of the table also are robust to the possible presence of serial correlation and conditional heteroskedasticity, and are computed as described below. Table 1 is consistent with the results of many earlier studies (e.g., Baillie and Bollerslev (1989), Diebold and

Nerlove (1989), Engle and Bollerslev (1986)). Exchange rate changes appear to have zero unconditional means (line (1)), and, with the possible exception of Japan, appear to be serially uncorrelated (lines (5) to (7)).

Exchange rate changes are also very fat tailed. This leptokurtosis may be seen in Figures 1B1 to 1B5, each of which plots both a normal density whose mean and variance match sample estimates and a histogram of the data. More formal evidence is in panel A of Table 1. The standard deviation of exchange rate changes is about one per cent per week (line (2)); the maximum and minimum changes in this sample of size 863 are generally five or more standard deviations away from the mean (lines (8) and (12)), and the interquartile range is much less than two standard deviations (lines (9) and (11)). Excess kurtosis is greater than two and is significantly different from zero at any conventional significance level for all countries except Canada (line (4)). With the exception of Germany, there is no evidence of skewness (line (3)).

Panel B of Table 1 contains some summary statistics on squared exchange rates. The means and standard deviations (lines (13) and (14)) are presented for convenience of interpretation of our empirical results; they are redundant in the sense that the point estimates can be deduced from the appropriate entries in panel A. Rows (15) to (17) in panel B suggest that, in stark contrast to the levels, the squares of exchange rates are highly serially correlated. This, too, is a result consistent with many earlier studies.

B. Models and Estimation Techniques

The in-sample evidence in Table 1 that e_t is linearly unpredictable is supported by the stronger results from other studies, some of which use out-of- as well as in-sample evidence, that there is not even any nonlinear dependence in the conditional mean of e_t. The most salient reference is

Diebold and Nason (1990), who drew their data from exactly the same source as did we, but over the slightly shorter sample period 1973-1987. Despite much in-sample evidence of nonlinear dependence in the mean of e_t, they found little out-of-sample evidence of such dependence. Papers that come to similar conclusions using other data, sometimes with multivariate information sets, include Meese and Rogoff (1983) and Meese and Rose (1991). We therefore will limit ourselves to models in which the conditional mean of e_t is zero.

To define our models, some notation is needed. Let

(2-1a) $h_{t,j} = \text{var}_t(e_{t+j}) = E_t e_{t+j}^2 =$ (population) variance of e_{t+j}, conditional on information generated by past e_s, $s \le t$;

(2-1b) $\hat{h}_{mt,j} =$ fitted conditional variance of e_{t+j}, according to model m (e.g., model m is GARCH(1,1), or homoskedastic), estimated using data on past e_s, $s \le t$;

(2-1c) $h_t = h_{t,1}$, $\hat{h}_{mt} = \hat{h}_{mt,1}$;

(2-1d) R = endpoint of first sample used in estimation of regression parameters;

(2-1e) T = endpoint of last sample used in estimation.

Note our dating convention: what we denote h_t corresponds to what is often called h_{t+1} or σ_{t+1} (e.g., Engle (1982)). For concreteness in interpreting (2-1b) and (2-1c), it may help to note that in the tables below we report results for j=1, 12 and 24, corresponding to approximately to weekly, quarterly and semiannual horizons. To do so for a given horizon, we obtain for each model T-R+1 fitted values $\hat{h}_{mt,j}$, t=R,...,T, for models m=1,...,M, where the number of models M in the tables below is 6. We then compute the root mean squared prediction error (RMSPE) for model m at horizon j as

$$(2\text{-}2) \qquad [\ (T-R+1)^{-1}\Sigma_{t=R}^{T}(e_{t+j}^{2}-\hat{h}_{mt,j})^{2}\ ]^{1/2}$$

We focus on RMSPE because mathematical expectations have minimum RMSPE, so a good statistical model for the expected value of exchange rate squares will tend to have forecast errors whose average squared value is small.[3]

Column (1) of Table 2 lists the models we estimated, column (3) the acronyms used in some subsequent tables.[4] Column (2) gives the formula for the one period ahead conditional variance, except for the nonparametric estimator for which the formula for the arbitrary j period ahead forecast is given. Since all the other models are linear, multiperiod forecasts can be obtained by the usual recursive prediction formulas. Consistent with the assumption that exchange rate changes have zero conditional mean, in such forecasts the changes were assumed to be conditionally uncorrelated at all nonzero lags (i.e., $E_{t-1}e_t e_{t+j} = 0$ for all $j>0$).

The homoskedastic model (line (1)) simply set the conditional variance at all horizons equal to the sample mean of lagged e_t^2's.

Two GARCH models were used (lines (2) and (3)). Both were estimated by maximum likelihood assuming conditional normality, using analytical derivatives, with presample values of h and e^2 set to sample means. Lee and Hansen (1991) and Lumsdaine (1989) show that the conditional normality assumption is not necessary for the consistency and asymptotic normality of the estimators.[5] We chose GARCH(1,1) and IGARCH from a larger set of possible GARCH models after (1)analysis of some in-sample diagnostics seemed to suggest

GARCH(1,1) for Canada, Germany and the U.K., IGARCH for France and Japan, and (2)a little experimentation with ARCH(1), ARCH(2), GARCH(1,2) and GARCH(2,1) models suggested that MSPEs from these models are comparable or worse than the two we chose to study.

We also studied two autoregressive models, both of which were estimated by OLS. One autoregression used e_t^2 (line (4)). It is included because GARCH models imply ARMA processes for e_t^2 (see Bollerslev (1986)); OLS estimation of such autoregressions therefore might perform comparably to more complicated GARCH estimation (although under the GARCH null, such OLS estimation is asymptotically inefficient). As in Schwert (1989a, 1989b), whose work is based on that of Davidian and Carroll (1987), the other autoregression used $|e_t|$ (line (4)). Schwert suggests the factor of $(\pi/2)$ because the variance of a zero mean normally distributed random variable is $(\pi/2)$ times the square of the expected value of its absolute value. For both autoregressions, the lag length of 12 was chosen because for all countries in-sample results indicated that such a lag length was more than sufficient to produce a Q-statistic that implied white noise residuals.

Finally, we also tried a nonparametric estimator (line (6)). It can be interpreted as working off the basic definition $E(e_{t+j}^2|e_t) = \int_0^\infty e_{t+j}^2 f(e_{t+j}^2|e_t)de_{t+j}^2$, where $f(e_{t+j}^2|e_t)$ is the density of e_{t+j}^2 conditional on e_t. See Pagan and Ullah (1990a,1990b) for an excellent exposition. As in Pagan and Schwert (1990a) we used a Gaussian kernel, defined in column (2), with the bandwidth $b = \hat{\sigma}(R-j)^{-1/5}$, $\hat{\sigma}$ the sample standard deviation of e_t, $t=1,\ldots,R-j$, $j=1$, 12 or 24. We did not try any other kernel. We did a little experimentation with some alternative fixed bandwidths and information sets, comparing MSPEs, but found that these yielded similar results.

There remain two questions before we can begin our model evaluation. The first is where to begin the out of sample exercise. We arbitrarily began our forecasts at the midpoint of the sample, and the first sample for which we fit any models included the 432 observations from March 14, 1973 to June 17, 1981. Because the final 24 weeks of the sample (April 12, 1989 to September 20, 1989) were used only for forecast evaluation, the last observation of our final estimation sample was April 5, 1989. (In the notation of (2-1d), R=432 and T=839.) For our one week horizon, the predictions and realizations of e_t^2 spanned the 408 weeks from June 24, 1981, to April 12, 1989; the comparable 408 week period for the 12 and 24 week horizons may be obtained by shifting the one week dates forward by 11 and 23 weeks respectively.

The other question concerned what sample should be used for estimation as additional observations were added beyond the June 17, 1981, date at which our first sample ended. In our initial work, we estimated each of our models on both (1)rolling samples, in which the sample size used for estimation was fixed at 432, and what had been the initial observation as each additional observation was added, and (2)expanding samples, in which the sample size grew as additional observations were added. RMSPEs were quite similar for rolling and expanding samples, with those for rolling samples perhaps showing a slight tendency to be smaller (rolling RMSPEs were smaller in 63 of the 90 experiments [90 = 5 countries times 6 models times 3 horizons]). To keep the project manageable, we therefore decided to subject only the rolling estimators to detailed analysis.

C. Procedures for Asymptotic Inference

Most of our inference is based on asymptotic approximations described below. In addition to the usual reasons to be concerned about the finite sample accuracy of such approximations, there are grounds to be concerned about the applicability of regularity conditions typically underlying such approximations: exchange rate data may lack suitable higher order moments (e.g., Loretan and Phillips (1992)); one of our models uses a nonparametric estimator; more generally, the previous paragraph's observation that forecast quality did not deteriorate when we used rolling rather than expanding samples suggests that the usual conditions may not hold. Nevertheless, we conduct most our inference using such theory, for two reasons. First, a small Monte Carlo experiment to double check one piece of our asymptotic analysis suggested that the asymptotic approximation is unlikely to be very misleading if some minimal conditions do hold, and, second, the computational cost of using bootstrap methods throughout is enormous, given the nonlinear search required to estimate GARCH and IGARCH models.

To explain the asymptotic procedures that we used: Let P be the sample size. Under suitable regularity conditions, it is well known that if g_t is a zero mean, covariance stationary random vector, $P^{-1/2}\Sigma_{t=1}^{P} g_t \overset{A}{\sim} N(0,S)$, where $S \equiv \Sigma_{j=-\infty}^{\infty}\Gamma_j$, $\Gamma_j \equiv E g_t g_{t-j}'$ (e.g., Hannan (1973)); White (1984) summarizes some parallel results that apply when data satisfy some mixing conditions but possibly are not stationary. Suppose that g_t is a function of an underlying vector of parameters of interest, say, θ, and that $\hat{\theta}$ is estimated by setting $P^{-1}\Sigma_{t=1}^{P} g_t(\hat{\theta}) = 0$. A straightforward Taylor series argument yields $P^{1/2}(\hat{\theta}-\theta) \overset{A}{\sim} N(0,V)$, $V \equiv (E\partial g_t/\partial\theta)^{-1} S (E\partial g_t/\partial\theta)^{-1\prime}$; see Hansen (1982) for a formal argument in the stationary case, Gallant and White (1988) for the parallel argument, and more complicated formulas, under conditions that allow for the possibility that g_t is not stationary.

In our applications of this result, $\partial g_t/\partial\theta$ does not depend on θ and so

$E\partial g_t/\partial\theta$ is a matrix of known constants. The estimator of S that we used was that suggested by Newey and West (1987):

$$\hat{S} = \hat{\Gamma}_0 + \Sigma_{j=1}^{k}[1-j/(k+1)](\hat{\Gamma}_j + \hat{\Gamma}_j'), \tag{2-3}$$

where $\hat{\Gamma}_j$ is the j'th sample autocovariance of $\hat{g}_t$, $\hat{\Gamma}_j = P^{-1}\Sigma_{t=j+1}^{P}\hat{g}_t\hat{g}_{t+j}'$. The value of k in (2-3) was determined by a data dependent automatic rule that has certain asymptotic optimality properties (Newey and West (1993)): Let n be the integer part of $4(P/100)^{2/9}$, so that n=6 for estimates based on the 863 observations in the whole sample (e.g., Table 1), n=5 for estimates based on 408 observations in the forecasting sample (e.g., Table 4). Also, let w be vector of ones of the same dimension as g_t, $\hat{\sigma}_j = w'\hat{\Gamma}_j w$, $\hat{s}^{(0)} = \hat{\sigma}_0 + 2\Sigma_{j=1}^{n}\hat{\sigma}_j$, $\hat{s}^{(1)} = 2\Sigma_{j=1}^{n}j\hat{\sigma}_j$. Then k was set to the integer part of $1.1447\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^{2/3}\{\text{sample size}\}^{1/3}$. The resulting values for k in Table 1, for example, were Canada--4, France--1, Germany--6, Japan--7, and the U.K.--11. The values for the remaining tables are available on request.

Some details may be helpful in understanding how we used this framework. In Table 1, lines (1)-(4), (13) and (14), begin by defining the (4x1) vector $X_t = (e_t,\ e_t^2,\ e_t^3,\ e_t^4)'$. Let $\theta = (Ee_t,\ Ee_t^2,\ Ee_t^3,\ Ee_t^4)'$, $\hat{\theta} = (P^{-1}\Sigma_{t=1}^{P}e_t,\ P^{-1}\Sigma_{t=1}^{P}e_t^2,\ P^{-1}\Sigma_{t=1}^{P}e_t^3,\ P^{-1}\Sigma_{t=1}^{P}e_t^4)'$, $g_t = X_t - \theta$, $\hat{g}_t = X_t - \hat{\theta}$. Then $P^{1/2}(\hat{\theta}-\theta) \overset{A}{\sim} N(0,S)$, $S = \Sigma_{j=-\infty}^{\infty}\Gamma_j$, $\Gamma_j = Eg_t g_{t-j}'$. Given the estimate of S, standard errors on the relevant entries in Table 1 can be computed using the delta method. A similar method was used in panel A of Table 6 below.

In the modified Ljung-Box statistic, Table 1, lines (5)-(7), $\theta = (Ee_te_{t-1}, \ldots, Ee_te_{t-r})'$, r=10, 50 or 90, $\hat{\theta}$ the corresponding sample moments, $X_t = (e_te_{t-1}, \ldots, e_te_{t-r})'$, $g_t = X_t - \theta$, $\hat{g}_t = X_t - \hat{\theta}$. We assume that the conditional first moment of e_t is zero, which implies that g_t is serially uncorrelated, and that $Ee_t^2e_{t-i}e_{t-j}=0$ for $i\neq j$, which implies that EX_tX_t' is diagonal. With a little algebra, this validates the following: For j=0,...,r, let $\hat{\sigma}_j$ be the j'th element of $\hat{\theta}$, $\hat{\sigma}_j = P^{-1}\Sigma_{t=j+1}^{P}e_te_{t-j}$, and let $\hat{\kappa}_j = P^{-1}\Sigma_{t=j+1}^{P}e_t^2e_{t-j}^2$, $\hat{\rho}_j = \hat{\sigma}_j/\hat{\sigma}_0$. Then for any fixed r,

$$(2\text{-}4) \qquad P(P+2)\hat{\sigma}_0^2\Sigma_{j=1}^{r}(P-j)^{-1}(\hat{\rho}_j^2/\hat{\kappa}_j) \overset{A}{\sim} \chi^2(r).$$

If the data are conditionally homoskedastic, so that $Ee_t^2e_{t-j}^2 = Ee_t^2Ee_{t-j}^2 = \sigma_0^2$, $\hat{\kappa}_j \overset{p}{\to} \sigma_0^2$ and this statistic is asymptotically equivalent to the standard Ljung-Box statistic.

In Tables 4 and 5 below, which report inference about forecasts or forecast errors, the conceptual experiment that underlies our asymptotic approximation is one in which both the number of observations used in estimation (R, in the notation of (2-2)) and the number used for forecasting (T-R+1≡P) go to infinity, with (T-R+1)/R approaching a finite constant (possibly zero).

Consider, for example, the one period ahead MSPE. For notational simplicity, assume stationarity (rather than, say, just mixing). Let h_{mt} be model m's population prediction of e_{t+1}^2 at time t (i.e., the prediction it would make if an infinite sized sample had been used in estimation). Let δ be the vector of the entire set of regression parameters, across all models (the constant for the homoskedastic model, the constant and coefficients on e_t^2 and h_{t-1} for the GARCH(1,1) model,...). Let $u_{mt+1} = e_{t+1}^2-h_{mt}$, $\sigma_m^2 = Eu_{mt+1}^2 = E(e_{t+1}^2-h_{mt})^2$, $\hat{u}_{mt+1} = e_{t+1}^2-\hat{h}_{mt}$, $\hat{\sigma}_m^2 = (T-R+1)^{-1}\Sigma_{t=R}^{T}(e_{t+1}^2-\hat{h}_{mt})^2$, $\hat{\theta} \equiv (\hat{\sigma}_1^2, \ldots , \hat{\sigma}_M^2)'$, $\theta \equiv (\sigma_1^2, \ldots , \sigma_M^2)'$, $g_{t+1}(\theta,\delta) = (u_{1t+1}^2-\sigma_1^2,\ldots,u_{Mt+1}^2-\sigma_M^2)'$, $g_{t+1}(\hat{\theta},\hat{\delta}) = (\hat{u}_{1t+1}^2-\hat{\sigma}_1^2, \ldots,$

$\hat{u}^2_{Mt+1}-\hat{\sigma}^2_M)'$. It may be shown that under suitable conditions, sampling error in $\hat{\delta}$ is irrelevant for asymptotic inference on θ (West (1993)), and we apply the logic above to $(T-R+1)^{-1}\Sigma_{t=R}^T g_{t+1}(\hat{\theta},\hat{\delta}) = P^{-1}\Sigma_{t=R}^T \hat{g}_{t+1}$. The implication is that $P^{1/2}(\hat{\theta}-\theta) \overset{A}{\sim} N(0,S)$, where the (i,q) element of the MxM matrix S is $\Sigma_{j=-\infty}^{\infty}E(u^2_{it}-\sigma^2_i)(u^2_{qt-j}-\sigma^2_q)$. A test statistic for the equality of the MSPEs across all M models is constructed as follows. Let B be the (M-1)xM matrix whose first column is (-1,-1,...,-1)' and whose (M-1) other columns contain the identity matrix; the null is that $B\theta=0$. Then for $\hat{S}$ constructed as in (2-3),

$$(2\text{-}5)\qquad (T-R+1)[\hat{\theta}'B'(B\hat{S}B')^{-1}B\hat{\theta}] = P[\hat{\theta}'B'(B\hat{S}B')^{-1}B\hat{\theta}] \overset{A}{\sim} \chi^2(M-1),$$
$$\hat{S} = \hat{\Gamma}_0+\Sigma_{j=1}^k[1-j/(k+1)](\hat{\Gamma}_j+\hat{\Gamma}_j'),\ \hat{\Gamma}_j = P^{-1}\Sigma_{t=R+j}^T\hat{g}_t\hat{g}_{t+j}'.$$

Note that since we select k as described, and do not constrain k to be 0, we allow the forecast errors to be serially correlated. Similar formulas apply for the 12 and 24 period ahead predictions, with, e.g., $u_{m,t+12} = e^2_{t+12}-h_{m,t+12}$ and $\sigma^2_{m,12} = E(e^2_{t+12}-h_{m,t+12})^2$.

3. Basic Empirical Results

To frame our discussion, Table 3A presents estimates of the GARCH(1,1) model for the first of our rolling samples. The Appendix available on request has parallel estimates for the other models; we present GARCH(1,1) here because of its simplicity and because, as we shall see, it worked relatively well in forecasting. For the benefit of those familiar with GARCH, we briefly note that the estimates suggest, as usual, considerable persistence, since $\alpha+\beta$ is estimated to be above 0.80 in all five countries, above 0.90 in France, Germany and Japan; the null that $\alpha+\beta=1$ could not be rejected at the five percent level for France and Japan (not reported in the table).

What is of particular interest to us is how such parameters translate into for RMSPEs at various horizons. Suppose e_t^2 is stationary, ($\alpha+\beta<1$, under a GARCH(1,1) parameterization). With the exception of IGARCH, all our estimators will then yield the essentially the same predictions in population for a sufficiently long horizon, since all will predict that the e_t^2 will be near its unconditional mean. Accordingly, the RMSPEs will also be essentially the same. We use the GARCH(1,1) estimates in Table 3A to get an idea of how long a horizon is needed for this to occur.

Table 3B reports the ratio of the population RMSPEs of a homoskedastic model to that of a GARCH(1,1) model, for each of our three horizons, and for each of the five sets of estimates of α and β reported in Table 3A.[7] According to columns (1) and (5) in Table 3B, the Table 3A estimates for Canada and the U.K. suggest sufficiently rapid mean reversion that our proposed comparisons of 12 and 24 week horizons are probably not of interest. On the other hand, columns (2) to (4) indicate that other Table 3A estimates imply as sharp a difference in RMSPEs at one or both of these longer horizons as occurs at a one period horizon for the U.K. parameters in column (5).

We will <u>not</u> attempt to squeeze an interpretation of the results of our out-of-sample comparison into the Table 3 figures. Even under a GARCH(1,1) null the Table 3 figures will be misleading insofar as sampling error has affected the point estimates of α and β. Rather, we interpret Table 3 as presenting in-sample evidence that it may be possible to distinguish different estimators at horizons of as long as 24 weeks.

Table 4 presents our attempts to do so, for forecasts of 1 and 12 weeks as well as 24 weeks ahead, in panels A, B and C respectively. In each panel, under each country are two columns. The second, labelled "RMSPE" gives the root mean squared prediction error, computed according to equation (2-2). The other column, labelled "Rank," indicates the relative size of that model's RMSPE, 1 indicating the best (smallest) RMSPE, 6 the worst (largest). The rows labelled "H_A", "H_B" and "H_C" (at the bottom of each panel) will be discussed below.

We begin with two general comments, before beginning a comparison of the models. First, as one would expect, given the noisiness of exchange rate data, these out-of-sample RMSEs generally are larger than the in-sample RMSEs reported in line (14) of Table 1. That is, the out-of-sample predictions using the estimated conditional variances are usually less accurate than an in-sample prediction using the in-sample unconditional variance. Second, and somewhat surprisingly, there does not appear to be a tendency for RMSPEs to increase at longer horizons; the median Table 4 values for the 1, 12 and 24 week horizons are 4.746, 4.791 and 4.503, for example. In the context of GARCH(1,1) models, the implication is that mean reversion occurs as rapidly as in, say, column (5) of Table 3B. The figures in columns (2) to (4) of that Table suggest otherwise, so there is a clear conflict between the out-of-sample and in-sample evidence.

Turn now to comparing the models. At the 1 week horizon, panel A indicates that one of the two GARCH models had the smallest RMSPE for all five countries. The IGARCH model was probably the most consistent performer overall, being best in three countries (France, Germany and U.K.), second and third best in the other two (Japan and Canada). At the 12 week horizon (panel B), the best model was either the homoskedastic (Canada, France and Germany) or autoregression in absolute values (Japan, U.K.). At the 24 week horizon (panel C), depending on the country, one of four different models had the lowest RMSPE, GARCH(1,1) being the only model that was best in two countries (Germany, U.K.). But the most consistent performer at 24 weeks was probably the autoregression in exchange rate squares, which was second in four countries and first in one (Japan).

Which model performs best, then, varies from country to country and horizon to horizon; if there is an underlying pattern, it is difficult for us to discern, and, at least superficially, Table 1 suggests that it might be largely a matter of chance which model produces the smallest RMSPE.[8] That performance is quite similar across models is also suggested by casual inspection of the point estimates of the RMSPEs; even at a one period horizon, in only one case is the worst model's RMSPE more than five percent larger than the best model's (U.K.); once again, such point estimates are surprising in light of Table 3B.

For an additional measure of similarity of RMSPEs, we turn to formal statistical testing of the hypothesis that these are the same across various models, for a given horizon. In Table 4, the "H_A", "H_B" and "H_C" rows at the bottom of each panel give statistics and, in brackets, p-values assuming an asymptotic chi-squared distribution, for the following three hypotheses:

(3-1) H_A: MSPEs for all six models are equal ($\chi^2(5)$).

H_B: MSPEs for the best model and the homoskedastic model are equal ($\chi^2(1)$).

H_C: MSPEs for the homoskedastic, GARCH(1,1), and two autoregressive models are the same ($\chi^2(3)$).

Hypothesis A is an obvious one.[9] Tests of hypothesis B were performed because the homoskedastic model is the simplest one, and therefore probably the model of most appeal if, in fact, performance is similar across models. Tests of hypothesis C were performed because the formal asymptotic theory that underlies the test makes assumptions that rule out our nonparametric estimator and possibly the IGARCH estimator as well.

Table 4 indicates that the H_A test of the null of equal RMSPEs across all models is rejected at the .05 level in four of our fifteen experiments (Canada and France, 12 and 24 week horizons) and once at the .10 but not .05 level (Canada, 1 week horizon). This suggests that the seeming similarity of point estimates of RMSPEs might be misleading, at least for Canada and France. In no case, however, can one reject at conventional significance levels the null that the homoskedastic model's RMSPE is the same as that of the best model: the lowest of p-value for H_B is 0.217 (U.K., one week horizon). The H_C test of equal RMSPEs for the homoskedastic, GARCH(1,1), and two AR models rejects at the .05 level for France for all three horizons, again suggesting that the seeming similarity of point estimates of RMSPEs might be misleading for France.

These asymptotic tests may well be deceptive in finite samples, even if the asymptotic theory eventually yields a good approximation. One indication that this may be the case is that of the four rejections at the .05 level of equality of all six models, three occur in experiments in which the homoskedastic model is the best (Canada, 12 and 24 week; France, 24 week). If, indeed, a homoskedastic model were generating the data, at least four of the other five models would produce exactly the homoskedastic forecast in an infinitely large sample (the possible exception is IGARCH, whose asymptotic behavior under these conditions is unclear to us). But this suggests a tendency to reject too much, not too little, a result that we have found in related Monte Carlo studies using data generated by GARCH processes (Newey and West (1993)).

But to double check the possibility that our asymptotic tests are instead rejecting too infrequently, we undertook two exercises. First, we examined a seventh model, which set $\hat{h}_t = e_t^2$--the conditional variance in week t is equal to the realized square of the exchange rate. (Reminder to readers familiar with the GARCH literature: what we call h_t here is usually called h_{t+1}.) To our knowledge, this has <u>not</u> been seriously proposed as a model for exchange rate volatility, for the good reason that it is not an appealing one: the RMSPEs for the one week horizon, for example, are: Canada 0.933; France 7.119; Germany 6.574; Japan 5.593; U.K. 7.466. These are a good 25 percent above the Table 4A figures for the other models. We use it here to see if our asymptotic tests have enough power to recognize the substantial difference between this model and the others. And they do, as is indicated by the following summary of test results. Of 15 $\chi^2(6)$ tests of the equality of RMSPEs across all seven models, 11 reject at the .05 level, 13 at the .10 level. Of 15 $\chi^2(1)$ tests of the equality of the RMSPE from this additional model and that of the <u>worst</u> of the six models reported in Table 4, 13 reject at the .05 level, 14 at the .10 level (the exception was U.K., one week horizon, which rejects at the .15 level). It seems, then, that whatever the problems with our asymptotic tests, these tests <u>do</u> have enough power to reject an egregiously poor model at conventional significance levels.

The second exercise we undertook to check the validity of our asymptotic tests was a small Monte Carlo experiment. Because of space constraints, we limit ourselves to the succinct statement that the experiment suggested that if our asymptotic procedures have a small sample bias, that bias is towards rejecting too much, not too little. A detailed discussion of the experiment is available in the Appendix that is available on request.

4. Additional Empirical Results

It seems that to a first approximation all our models are equally good as predictors of exchange rates squares. To compare them from a slightly different perspective, we conducted a standard efficiency test (e.g., Pagan and Schwert (1990a)), estimating by OLS the regression

$$(4\text{-}1)\qquad e_{t+1}^2 = b_0 + b_1\hat{h}_{mt} + \epsilon_{t+1}.$$

If, indeed, $E_t e_{t+1}^2 = \hat{h}_{mt}$, one should get $b_0=0$, $b_1=1$. One should also find that ϵ_{t+1} is serially uncorrelated. But a quick look at the autocorrelations of the residuals suggested that this was rarely if ever the case. So we do not formally test for the absence of serial correlation, and instead correct the variance-covariance matrix of the estimated parameter vector for conditional heteroskedasticity as well as serial correlation, using the techniques described above.

Results are in Table 5. Asymptotic standard errors for b_0 and b_1 are given in parentheses beneath the point estimates. For all five countries, the $\chi^2(2)$ column gives the point estimate and asymptotic p-value for H_0: $b_0=0$, $b_1=1$. The "**" and "*" after the estimates of $\hat{b}_1$ indicate significant differences from <u>zero</u>, not one.

We note first that the rankings by R^2 are quite similar to those by

RMSPE. This indicates that models with relatively low RMSPEs also have RMSPEs whose variance component is relatively low, since R^2 reflects the variance but not bias-squared component of MSPE. Some new information is yielded by the other estimates. Of the 30 $\chi^2(2)$ tests of H_0: $b_0=0$, $b_1=1$, 27 reject at the .10 level (the exceptions are GARCH(1,1) for Japan, IGARCH for Japan and U.K.), 25 at the .05 level (the additional exceptions are the two autoregressions for Canada). The standard errors on $\hat{b}_0$ and $\hat{b}_1$ yield compatible conclusions.

Perhaps unsurprisingly, then, none of the models pass this efficiency test: the Monte Carlo simulation indicates that this test has good power, being very likely to reject the null (not reported in the Table). More encouraging is that 7 of the estimates of b_1 are significantly different from zero at the .05 level, five of these being for GARCH models (see the "**" entries). This shows that there is some predictive power in the estimated conditional variances. For future reference, note the marked tendency of the models to have predictive power for Canadian data.

We also performed the efficiency test in Table 5 for the 12 and 24 week horizons. For these horizons, the results did not help discriminate between models, and we therefore limit ourselves to a summary of the results. Of the 60 $\chi^2(2)$ tests, 58 reject at the .10 level (the exceptions are GARCH(1,1) for Canada 12 week and U.K. 24 week), 56 at the .05 level (the additional exceptions are homoskedastic for Canada 12 and 24 week). More troubling is that while $\hat{b}_1$ was different from zero at the .05 level 7 times, only two of those estimates were positive (GARCH(1,1) and IGARCH for U.K., 12 week).

Overall, then, it seems that at the one period horizon there is some evidence favoring GARCH models: while Table 4 cannot reject the null that the RMSPEs are the same for all models, GARCH models do tend to produce lower

RMSPEs, and Table 5 suggests that they have markedly more predictive power for next period's e_{t+1}^2. On the other hand, at longer horizons, we find little grounds for preferring one model over another.

This is a disappointing, and surprising, result. It seems that mean reversion in the conditional variance occurs rapidly enough that no model dominates the others at 12 week or longer horizons. This suggests that the in-sample fits overstate the conditional predictability of exchange rate squares. Lamoureux and Lastrapes (1990) have shown that occasional discrete shifts in the mean level of volatility cause substantial upward bias in estimates of the persistence of volatility. We close this section with some evidence that such shifts may have occurred here, and thus may help account for our inability to sharply distinguish one model from another.

Panel A of Table 6 reports split sample estimates of the standard deviation of e_t. As one can see, the point estimate is markedly higher in the second half of the sample for all countries except perhaps Canada.[10] In addition, line (3) of the table indicates that the null of equality is rejected at the .05 level for France, Japan and the U.K., at the .10 level for Germany.

Given that we began forecasting at the sample midpoint, the choice of the midpoint as a date to test for a shift is natural, but nonetheless still arbitrary. In panel B we report a Pagan and Schwert (1990b) test for the constancy of the unconditional variance of e_t that does not require a priori specification of a date. The details of the test are described in the notes to the table. As indicated in the table, the null of constancy is rejected at the .05 level for all countries but Canada, for which it is not rejected at even the .20 level. See row (1) of panel B.

Rows 2 to 4 of Table 6 report the results of applying this test on three subsamples for each country: the first half of our total sample (March 14, 1973 to June 17, 1981), the middle two-fourths (April 24, 1977 to July 31, 1985), and the last half (June 24, 1981 to September 20, 1989). Of the 15 tests for constant variances (15 = 3 subsamples times 5 countries), only two tests rejected at the .05 level (Japan, beginning and middle subsamples).

Now, if the data were driven by a stationary model that allows time varying conditional variances, it would not be surprising if tests such as those in Table 6 found evidence of shifts in variance at short but not long horizons. We, however, find the converse. And, as briefly noted in section 2, forecast quality was no better for expanding than for rolling samples, which also seems to suggest a failure of the stationarity assumption.

In this study, we followed many others (e.g., Engle et al. (1990)) and implicitly allowed for a failure of stationarity by using rolling samples. We conjecture that it will be productive to explore models that explicitly allow for seeming or actual movement in the unconditional variance of e_t. The sort of movement that one wants to capture appears to be slow enough that it might not be detectable in samples that are eight years long, but rapid enough that it is marked in samples sixteen years long.

Canadian data were unusual in that Table 4's tests of equality of RMSPEs tended to find differences across models, and Table 5's efficiency tests were unusually likely to be able to find predictive power in the estimated conditional variances. Perhaps the distinctive results for Canada are no accident, but instead are linked to the stationary behavior of its exchange rates.

5. Conclusions

The in-sample evidence summarized in Tables 1 and 3 strongly suggests that a homoskedastic model should be dominated by the other models that we studied. This did not turn out to be the case. We speculate that models that allow for seeming or actual drift in unconditional moments may result in superior performance. Possibilities include processes that allow occasional discrete jumps (Jorion (1988)) and models with time varying parameters (Chou et al. (1990)).

Footnotes

1. We find this to also be a message, perhaps implicit, in the studies using stock price data by Pagan and Schwert (1990a,1990b) and Chou et al. (1991), as well in Loretan and Phillips (1992).

2. We also obtained Italian data. But in-sample statistics such as those reported in Table 1 suggest a nonzero unconditional mean. Fitted GARCH models tended to be explosive, with $\hat{\alpha}+\hat{\beta}>1$ in the notation of Table 2; apparently this resulted in part from the nonzero sample mean since removing this mean lessened the tendency to get explosive estimates. We dropped Italy rather than fit means as well as variances.

3. In related work (West, Edison and Cho (1993)) we consider an alternative measure of model quality, which also tends to favor GARCH.

4. We also used these models in another paper (West, Edison and Cho (1993)), and some of the prose in the remainder of this subsection also appears in that paper.

5. For efficiency reasons, one might nonetheless prefer to assume, say, a conditional t distribution, if the conditional density is in fact t. Our reading of the in-sample evidence is that this is not essential (e.g., Baillie and Bollerslev (1989) found little support for the use of a t in weekly exchange rate data).

6. Diebold and Mariano (1991) have independently suggested conducting inference on forecast errors using similar techniques, and Diebold (1988) suggested our modification of the Ljung-Box statistic in the specific case of a GARCH data generating process.

7. These population figures ignore the effects of sampling error in the estimation of model parameters. Reinsel (1980) and Ericcson and Marquez (1989), among others, have suggested a refinement to the computation of the RMSPE that accounts for sampling error in such estimation. But inspection of

their formulae and simulation results indicates that the refinement has a noticeable effect only when the following ratio is much larger than in our application: (number of regressors) / (sample size). While neither of these papers considers data that are conditionally heteroskedastic, we take the message to be that such a refinement is unlikely to much affect the Table 3B figures.

8. Consistent with this statement, and with the literature surveyed in Clemen (1989), a prediction formed by averaging the six forecasts typically performs better than any of the individual forecasts, at least at the longer horizons. Of the 15 comparisons, the ranking of the average forecast was: 1--7 times (2 of the 7 occur for a one period horizon; IGARCH performs roughly comparably here); 2--6 times, 3--once; 4--once. Details are in the Appendix that is available on request.

9. For computational convenience, we computed tests for equality of the MSPE's rather than the asymptotically equivalent tests for the RMSPE's; for expositional convenience, in all discussion apart from the statement of the tests in the preceding paragraph in the text, we refer to these as tests on the RMSPE's.

10. This raises the question of whether our exercise would produce different results if applied to split samples, a question also raised by a referee who noted that in the mid-1980s central banks attempted to drive down the dollar. We computed one week ahead RMSPEs for samples running from (1)6/17/81 to 9/18/85 (number of predictions = 223), and (2)9/18/85-4/5/89 (number of predictions = 185); the split date was chosen because the Plaza Accord was announced on 9/22/85. The RMSPEs were generally higher in the later sample. But GARCH or IGARCH still fared relatively well: one or the other was best in all five comparisons in the early sample, in three of the five comparisons in the later sample. Details are in the Appendix that is available on request.

References

Akgiray, Vedat, 1989, "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts," Journal of Business 62, 55-80.

Baillie, Richard T., and Bollerslev, Tim, 1989, "The Message in Daily Exchange Rates: A Conditional Variance Tale," Journal of Business and Economic Statistics 7, 297-305.

Bollerslev, Tim, 1986, "Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics 31, 307-327.

Bollerslev, Tim, Chou, Ray Y., and Kenneth F. Kroner, 1992, "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics 52, 5-59.

Chou, Ray, Engle, Robert F., and Alex Kane, 1991, "Measuring Risk Aversion from Excess Returns on a Stock Index," NBER Working Paper No. 3643.

Clemen, Robert T., 1989, "Combining Forecasts: A Review and Annotated Bibliography," International Journal of Forecasting 5, 559-583.

Davidian, M. and R.J. Carroll, 1987, "Variance Function Estimation," Journal of the American Statistical Association 82, 1079-1091.

Diebold, Francis X., 1988, Empirical Modeling of Exchange Rate Dynamics, Berlin: Springer-Verlag.

Diebold, Francis X., and Roberto S. Mariano, 1991, "Comparing Predictive Accuracy I: An Asymptotic Test," manuscript, University of Pennsylvania.

Diebold, Francis X., and James Nason, 1990, "Nonparametric Exchange Rate Prediction?", Journal of International Economics 28, 315-332.

Diebold, Francis X., and Marc Nerlove, 1989, "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," Journal of Applied Econometrics 4, 1-21.

Engle, Robert F., 1982, "Autoregressive Conditional Heteroskedasticity, with Estimates of the Variance of United Kingdom Inflation," Econometrica 50, 987-1007.

Engle, Robert F. and Tim Bollerslev, 1986, "Modelling the Persistence of Conditional Variances, " Econometric Reviews 5, 1-50.

Engle, Robert F., Che-Hsiung Hong, and Alex Kane, 1990, "Valuation of Variance Forecasts with Simulated Options Markets," manuscript, UCSD.

Ericcson, Neil R. and Jamie R. Marquez, 1989, "Exact and Approximate Multi-Period Mean-Square Forecast Errors for Dynamic Econometric Models," Federal Reserve Board of Governors International Finance Discussion Paper No. 348.

Gallant, A. Ronald and Halbert White, 1988, A Unified Theory of Estimation and

Inference for Nonlinear Dynamic Models, New York: Basil Blackwell.

Hannan, Edward J., 1973, "Central Limit Theorems for Time Series Regressions," Z. Wahrscheinlichkeitstheorie verw. Geb. 26, 157-170.

Hansen, Lars Peter, 1982, "Asymptotic Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-1054.

Haubrich, Joseph G. and Andrew W. Lo., 1989, "The Sources and Nature of Long-Term Memory in the Business Cycle," NBER Working Paper No. 2951.

Jorion, Philippe, 1989, "On Jump Processes in the Foreign Exchange and Stock Markets," Review of Financial Studies 1, 427-445.

Lamoureux, Christopher G. and William D. Lastrapes, 1990, "Persistence in Variance, Structural Change and the GARCH Model," Journal of Business and Economic Statistics, 225-234.

Lee, Sang-Won and Bruce E. Hansen, 1991, "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," manuscript.

Loretan, Mico S. and Peter C.B. Phillips, 1992, "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets," forthcoming, Journal of Empirical Finance.

Lumsdaine, Robin L., 1989, "Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models," manuscript, Harvard University.

Meese, Richard A. and Kenneth S. Rogoff, 1983, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?", Journal of International Economics 14, 3-24.

Meese, Richard A. and Andrew K. Rose, 1991, "An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination", Review of Economic Studies 58, 603-618.

Newey, Whitney K., and Kenneth D. West, 1987, "A Simple, Positive Semidefinite, Hetroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica 55, 703-708.

Newey, Whitney K., and Kenneth D. West, 1993, "Automatic Lag Selection in Covariance Matrix Estimation," NBER Technical Working Paper No. 144.

Pagan, Adrian R., and G. William Schwert, 1990a, "Alternative Models for Conditional Stock Volatility," Journal of Econometrics 45, 267-290.

Pagan, Adrian R., and G. William Schwert, 1990b, "Testing for Covariance Stationarity in Stock Market Data," Economics Letters 33, 165-170.

Pagan, Adrian R., and Aman Ullah, 1990a, "Chapter 3: Methods of Density Estimation," manuscript, University of Rochester.

Pagan, Adrian R., and Aman Ullah, 1990b, "Chapter 4: Non-Parametric Estimation of Conditional Moments," manuscript, University of Rochester.

Reinsel, Greg, 1980, "Asymptotic Properties of Prediction Errors for the Multivariate Autoregressive Model Using Estimated Parameters," Journal of the Royal Statistical Society B, 328-333.

Schwert, G. William, 1989a, "Business Cycles, Financial Crises and Stock Volatility," 83-126 in K. Brunner and A. H. Meltzer (eds.), IMF Policy Advice, Market Volatility, Commodity Price Rules and Other Essays, Carnegie Rochester Series on Public Policy No. 31.

Schwert, G. William, 1989b, "Why Does Stock Market Volatility Change over Time?", Journal of Finance 44, 1115-1154.

West, Kenneth D., 1993, "Asymptotic Inference About Predictive Ability," manuscript, University of Wisconsin.

West, Kenneth D., Edison, Hali J., and Dongchul Cho, 1993, "A Utility Based Comparison of Some Models of Exchange Rate Volatility," Journal of International Economics 35, 23-45.

White, Halbert, 1984, Asymptotic Theory for Econometricians, New York: Academic Press.

Appendix

West and Cho, "The Predictive Ability of Alternative Models of Exchange Rate Volatility"
January 1994

This not-for-publication appendix contains results omitted from the body of the paper to save space. Following are:

I. Parameter estimates or summary statistics, T=432

- A. GARCH(1,1)
- B. IGARCH
- C. AR(12) in |e(t)|
- D. AR(12) in e(t)**2

II. Diagnostics on GARCH models

- A. Ljung-Box Statistics for e(t)/sqrt(h(t))
- B. Ljung-Box Statistics for e(t)**2/h(t)
- C. GARCH(1,2) estimates
- D. GARCH(2,1) estimates
- E. LM test for one and two additional coefficients, and t-test for IGARCH

III. Mean squared errors for additional GARCH models, using German data

IV. Estimates with Italian data

V. Mean squared errors for additional nonparametric models, using German data

VI. Out of sample RMSPEs for models estimated on expanding samples

VII. Bandwidths used in estimation of variance-covariance matrices

VIII. Chi-squared statistics for tests of model in which h(t) = e(t)**2

IX. Regression Tests of Efficiency, 12 and 24 Week Horizons

X. Monte Carlo Experiment

XI. RMSPE for prediction made using average forecast

XII. RMSPE for split samples

I. Parameter estimates, T=432

A. GARCH(1,1)

1. Canada

	CONSTANT	ALPHA	BETA
EST	0.00000552	0.26072752	0.54451673
S.E.	0.00000131	0.02411562	0.06148268

2. France

	CONSTANT	ALPHA	BETA
EST	0.00001396	0.34547912	0.61122294
S.E.	0.00000266	0.05482124	0.04430997

3. Germany

	CONSTANT	ALPHA	BETA
EST	0.00002022	0.29526189	0.61345651
S.E.	0.00000447	0.04876142	0.05331147

4. Japan

	CONSTANT	ALPHA	BETA
EST	0.00000074	0.05327541	0.94455421
S.E.	0.00000035	0.00769470	0.00690244

5. U.K.

	CONSTANT	ALPHA	BETA
EST	0.00001888	0.11460277	0.73289811
S.E.	0.00000733	0.04635227	0.09907802

B. IGARCH

1. Canada

	ALPHA	
EST	0.12549316	0.87450684
S.E.	0.00630164	

2. France

	ALPHA	
EST	0.07941096	0.92058904
S.E.	0.00547796	

3. Germany

	ALPHA	
EST	0.08882797	0.91117203
S.E.	0.00033945	

4. Japan

	ALPHA	
EST	0.04363261	0.95636739
S.E.	0.00301078	

5. U.K.

	ALPHA	
EST	0.00007805	0.99992195
S.E.	0.00102633	

C. AR(12) in |e(t)|

```
DEPENDENT VARIABLE      1      EABSCAN
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420      SKIPPED/MISSING          0
USABLE OBSERVATIONS    420      DEGREES OF FREEDOM     407
R**2              .06733482     RBAR**2          .03983609
SSR           .50911213E-02     SEE          .35367920E-02
DURBIN-WATSON  2.02197733
Q(  60)=    42.1373        SIGNIFICANCE LEVEL  .961263

DEPENDENT VARIABLE      2      EABSFRA
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420      SKIPPED/MISSING          0
USABLE OBSERVATIONS    420      DEGREES OF FREEDOM     407
R**2              .15970723     RBAR**2          .13493201
SSR           .25915306E-01     SEE          .79795971E-02
DURBIN-WATSON  2.00114184
Q(  60)=    44.7489        SIGNIFICANCE LEVEL  .929113

DEPENDENT VARIABLE      3      EABSGER
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420      SKIPPED/MISSING          0
USABLE OBSERVATIONS    420      DEGREES OF FREEDOM     407
R**2              .16477312     RBAR**2          .14014727
SSR           .30872286E-01     SEE          .87093790E-02
DURBIN-WATSON  1.98183755
Q(  60)=    40.9947        SIGNIFICANCE LEVEL  .971228

DEPENDENT VARIABLE      4      EABSJAP
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420      SKIPPED/MISSING          0
USABLE OBSERVATIONS    420      DEGREES OF FREEDOM     407
```

```
R**2              .14225244      RBAR**2            .11696259
SSR           .27958344E-01      SEE            .82881674E-02
DURBIN-WATSON  2.00295282
Q(  60)=   36.6313       SIGNIFICANCE LEVEL  .992541
```

```
DEPENDENT VARIABLE     5     EABSUKG
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .12777411    RBAR**2            .10205737
SSR           .22111741E-01    SEE            .73707939E-02
DURBIN-WATSON  1.98334136
Q(  60)=   40.5622       SIGNIFICANCE LEVEL  .974435
```

D. AR(12) in e(t)**2

```
DEPENDENT VARIABLE     1     E2CAN
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .01788668    RBAR**2           -.01106998
SSR           .34738170E-05    SEE            .92386022E-04
DURBIN-WATSON  2.00462169
Q(  60)=    9.69061      SIGNIFICANCE LEVEL  1.00000

DEPENDENT VARIABLE     2     E2FRA
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .09352165    RBAR**2            .06679502
SSR           .34819148E-04    SEE            .29249057E-03
DURBIN-WATSON  1.99921420
Q(  60)=   52.2613       SIGNIFICANCE LEVEL  .751031

DEPENDENT VARIABLE     3     E2GER
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .11641248    RBAR**2            .09036076
SSR           .63289013E-04    SEE            .39433648E-03
DURBIN-WATSON  1.98712955
Q(  60)=   37.1952       SIGNIFICANCE LEVEL  .990932

DEPENDENT VARIABLE     4     E2JAP
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .05469587    RBAR**2            .02682449
SSR           .47371604E-04    SEE            .34116293E-03
DURBIN-WATSON  2.00251115
Q(  60)=   32.3521       SIGNIFICANCE LEVEL  .998669

DEPENDENT VARIABLE     5     E2UKG
FROM 1973: 6: 6 UNTIL 1981: 6:17
TOTAL OBSERVATIONS     420     SKIPPED/MISSING            0
USABLE OBSERVATIONS    420     DEGREES OF FREEDOM       407
R**2              .05186329    RBAR**2            .02390840
SSR           .26232018E-04    SEE            .25387424E-03
DURBIN-WATSON  1.99853815
Q(  60)=   34.3075       SIGNIFICANCE LEVEL  .996907
```

IIA. Ljung-Box Statistics for e(t)/sqrt(h(t))

			DEGREES OF FREEDOM		
COUNTRY	SAMPLE	MODEL	10	50	90

Canada	1-432	GARCH(1,1)	21.87	66.71	117.55
		IGARCH	25.69	72.33	154.49
France	1-432	GARCH(1,1)	44.06	71.26	113.83
		IGARCH	41.77	64.57	112.71
Germany	1-432	GARCH(1,1)	37.95	75.43	116.95
		IGARCH	45.53	84.20	123.71
Japan	1-432	GARCH(1,1)	36.08	70.26	92.30
		IGARCH	38.22	75.73	112.58
U.K.	1-432	GARCH(1,1)	24.53	70.46	117.80
		IGARCH	21.39	62.33	108.87

IIB. Ljung-Box Statistics for e(t)**2/h(t)

COUNTRY	SAMPLE	MODEL	10	50	90
Canada	1-432	GARCH(1,1)	6.81	56.35	82.84
		IGARCH	20.48	47.39	58.08
France	1-432	GARCH(1,1)	8.28	48.75	97.07
		IGARCH	14.05	39.95	79.14
Germany	1-432	GARCH(1,1)	6.89	53.61	84.87
		IGARCH	20.91	48.53	71.98
Japan	1-432	GARCH(1,1)	63.70	73.62	93.19
		IGARCH	16.97	28.33	43.63
U.K.	1-432	GARCH(1,1)	2.64	32.49	58.19
		IGARCH	20.04	64.58	100.09

IIC. GARCH(1,2) estimates

COUNTRY	SAMPLE	CONSTANT(10-5)	ALPHA	BETA1,	BETA2
Canada	1-432	.385	.270	.334	.272
		(.114	.025	.181	.174)
France	1-432	1.053	.208	.736	.00000152
		(.314	.075	.318	.250)
Germany	1-432	1.904	.279	.640	.00000100
		(.538	.071	.258	.189)
Japan	1-432	.189	.107	.00700	.880
		(.071	.015	.016	.018)
U.K.	1-432	1.757	.093	.808	.00000107
		(1.549	.084	.898	.766)

IID. GARCH(2,1) estimates

COUNTRY	SAMPLE	CONSTANT(10-5)	ALPHA	BETA1,	BETA2
Canada	1-432	.383	.328	.00000146	.660
		(.175	.038	.112	.103)
France	1-432	1.977	.239	.159	.514
		(.380	.076	.073	.057)
Germany	1-432	2.252	.284	.037	.578
		(.559	.070	.083	.073)

Japan	1-432	.163	.055	.00000139	.943
		(.074	.040	.042	.012)
U.K.	1-432	2.254	.069	.071	.678
		(.809	.052	.063	.112)

IIE. LM test for one and two additional coefficients, and t-test for IGARCH

There are two entries under each of the "one more" and "two more" columns. The first gives the standard LM test, the second the TR-squared version of the test.

The column labeled "IGARCH" gives the t-statistic for testing alpha+beta=1 in the GARCH(1,1) model.

COUNTRY	SAMPLE	MODEL	ONE MORE	(TR2)	TWO MORE	(TR2)	IGARCH
Canada	1-432	GARCH(1,1)	2.185	0.845	2.287	0.884	-3.112
France	1-432	GARCH(1,1)	3.543	1.588	3.552	1.592	-1.194
Germany	1-432	GARCH(1,1)	1.394	0.726	1.465	0.762	-2.631
Japan	1-432	GARCH(1,1)	6.477	1.381	6.530	1.392	-0.480
U.K.	1-432	GARCH(1,1)	1.142	0.324	1.213	0.344	-2.548

III. Mean squared errors for additional GARCH models, using German data

These are one week ahead out of sample mean squared errors. All estimates used expanding rather than rolling samples. The estimates for IGARCH duplicates that given below, and is included for comparison.

		Horizon	
	1	12	24
GARCH(2,0)	25.11	22.38	20.08
GARCH(2,1)	23.67	52.04	290.27
GARCH(1,2)	23.24	24.47	20.78
IGARCH	22.05	23.35	20.42

IV. Estimates with Italian data

Sample mean: -.0011
(.0005)

	T	m	α	β
GARCH(1,1)	432	.1141e-4	.4350	.7563
	863	.0634e-4	.7452	.2609

V. Additional estimates for nonparametric model, using German data

A. These are one week ahead out of sample mean squared errors. All estimates used expanding rather than rolling samples. Alternative bandwidths, b=k x sigma x (N-j)**(-.2):

k=.1	k=.5	k=1.5	Reference: k=1.0
27.12	22.78	22.39	22.47

B. Mean squared error, one week ahead forecast, information set = {e(t),e(t-1),e(t-2),e(t-3)}

33.54

VI. Out of sample RMSPEs, expanding samples

FOR HORIZON 1 MSPES ARE

	CANADA	FRANCE	GERMANY	JAPAN	U.K.
HOMO	0.716	5.156	4.701	4.399	5.765
(1,1)	0.706	5.415	4.801	4.370	5.581
IG	0.706	5.150	4.696	4.360	5.643
E2(12)	0.702	5.231	4.848	4.410	5.854
¦E¦(12)	0.707	5.170	4.789	4.407	5.683
NONP	0.749	5.207	4.740	4.453	6.633

FOR HORIZON 12 MSPES ARE

	CANADA	FRANCE	GERMANY	JAPAN	U.K.
HOMO	0.698	5.212	4.751	4.460	5.816
(1,1)	0.698	5.880	4.934	4.506	5.704
IG	0.740	5.274	4.832	4.481	5.765
E2(12)	0.697	5.220	4.755	4.447	5.736
¦E¦(12)	0.703	5.274	4.798	4.481	5.708
NONP	0.704	5.227	4.756	4.468	5.862

FOR HORIZON 24 MSPES ARE

	CANADA	FRANCE	GERMANY	JAPAN	U.K.
HOMO	0.696	5.086	4.494	4.454	5.791
(1,1)	0.701	5.908	4.534	4.560	5.741
IG	0.758	5.075	4.519	4.516	5.831
E2(12)	0.696	5.081	4.492	4.445	5.748
¦E¦(12)	0.702	5.130	4.529	4.493	5.776
NONP	0.703	5.128	4.545	4.461	5.969

VII. Bandwidths used in estimation of variance-covariance matrices

Here are the values of k (in the notation of section II of the paper):

TABLE 4

	Canada	France	Germany	Japan	U.K.
			Horizon = 1		
Ha	10	1	4	4	8
Hb	10	4	6	6	8
Hc	10	0	4	4	8
			Horizon = 12		
Ha	9	3	1	3	8
Hb	n.a.	n.a.	n.a.	2	9
Hc	9	3	1	3	7
			Horizon = 24		
Ha	9	1	1	4	8
Hb	n.a.	1	2	3	9
Hc	9	0	0	3	8

TABLE 5

	Canada	France	Germany	Japan	U.K.
homo	12	4	6	6	9
(1,1)	2	8	17	8	20
ig	7	0	7	6	10
e2AR	6	3	2	2	9
¦e¦AR	8	4	7	3	6
nonp	4	4	5	6	10

TABLE 6

Canada	France	Germany	Japan	U.K.
11	15	14	15	15

VIII. Chi-squared statistics for tests of model in which h(t) = e(t)**2

Hd: RMSPEs are the same for all 7 models (degrees of freedom = 6)
He: RMSPE from this model = that of next worst model

Horizon=1

HD	16.88[0.010]	9.55[0.145]	14.39[0.026]	22.75 [0.001]	11.86[0.065]
HE	5.31[0.021]	5.73[0.017]	3.98[0.046]	16.40 [0.000]	2.08[0.150]

Horizon=12

HD	20.50[0.002]	15.06[0.020]	9.02[0.172]	27.58 [0.000]	12.24[0.057]
HE	6.34[0.012]	5.28[0.022]	4.28[0.038]	5.32 [0.021]	4.51[0.034]

Horizon=24

HD	23.69[0.001]	21.34[0.002]	21.92[0.001]	14.14 [0.028]	16.64[0.011]
HE	8.61[0.003]	3.79[0.052]	4.02[0.045]	6.45 [0.011]	9.67[0.002]

IX. Regression Tests of Efficiency, 12 and 24 Week Horizons

As in Table 5, for each model, the first row gives: b0, b1, R2, chi-squared(2); the second row gives the asymptotic s.e.s on b0 and b1 and the asymptotic p-value for the test. For the horizons of 12 and 24, k=11 and 23.

HORIZON = 12

Canada

homo	1.031107	-2.086976	0.004045	2.720807
	0.635373	1.935881		0.256557
(1,1)	0.239761	0.306861	0.002823	3.130597
	0.138759	0.395703		0.209026
ig	0.325918	0.070601	0.000641	21.860954
	0.081591	0.200034		0.000018
e2AR	0.635848	-0.857535	0.004974	7.458198
	0.244076	0.684274		0.024014
¦e¦AR	0.941119	-2.087409	0.011682	10.628475
	0.297204	0.951702		0.004921
nonp	0.778693	-1.278986	0.014399	14.128027
	0.245351	0.656947		0.000855

France

homo	3.954564	-0.700857	0.004090	12.516717
	1.117870	0.491569		0.001914
(1,1)				
	3.013126	-0.131535	0.002664	110.636750
	0.563462	0.124768		0.000000
ig	3.252295	-0.259269	0.002222	25.576913
	0.787861	0.257790		0.000003
e2AR	4.471275	-0.889827	0.010840	27.636828
	0.946556	0.367142		0.000001
¦e¦AR	4.755856	-1.115419	0.014636	24.963501
	1.011162	0.426649		0.000004
nonp	3.529092	-0.469122	0.004009	18.214883
	0.907890	0.350131		0.000111

Germany

homo	2.856675	-0.140829	0.000165	8.807277
	0.983517	0.432525		0.012233
(1,1)	3.516629	-0.373886	0.004351	62.679105
	0.638720	0.189023		0.000000
ig	2.397748	0.063889	0.000227	33.606139
	0.440861	0.167530		0.000000
e2AR	2.964122	-0.178361	0.000814	54.762376
	0.472303	0.159319		0.000000
¦e¦AR	3.417485	-0.415819	0.002155	20.968764
	0.760616	0.312536		0.000028
nonp	2.563571	-0.000120	0.000000	12.649732
	0.733208	0.285712		0.001791

Japan

homo	3.036565	-0.380043	0.000591	5.479012
	1.418420	0.716212		0.064602
(1,1)	1.828322	0.201602	0.001376	5.673686
	0.824140	0.335469		0.058610
ig	1.877058	0.203426	0.001263	3.685731
	0.978864	0.424832		0.158363
e2AR	1.716029	0.279284	0.001609	5.287437
	0.753440	0.356621		0.071096
¦e¦AR	1.395504	0.482387	0.005937	5.368715
	0.747720	0.406798		0.068265
nonp	1.850532	0.218355	0.001224	8.552924
	0.645056	0.274249		0.013892

U.K.

homo	3.795168	-0.520904	0.002444	10.309707
	1.204796	0.474912		0.005771
(1,1)	1.518268	0.443546	0.036217	24.192541
	0.377649	0.119811		0.000006
ig	1.431124	0.519181	0.023400	6.345943
	0.574939	0.249774		0.041879
e2AR	1.643902	0.463512	0.016752	3.829037
	0.900187	0.383171		0.147413
¦e¦AR	1.153074	0.744758	0.022204	4.052592
	0.946938	0.455275		0.131823
nonp	3.098772	-0.174430	0.000815	44.379059
	0.631164	0.179535		0.000000

HORIZON = 24

Canada

homo	1.044784	-2.146363	0.004287	2.157743
	0.713870	2.168065		0.339979
(1,1)	0.498992	-0.424593	0.002829	39.868374

	0.121621	0.254102		0.000000
ig	0.378053	-0.089148	0.001024	176.278541
	0.076067	0.098691		0.000000
e2AR	1.047510	-2.133744	0.006223	4.553071
	0.505369	1.484385		0.102639
¦e¦AR	0.707019	-1.293817	0.001763	2.631784
	0.567396	1.994728		0.268235
nonp	0.511663	-0.520394	0.003564	19.689142
	0.140400	0.355605		0.000053

France

homo	3.645760	-0.556886	0.002706	10.137907
	1.145648	0.504704		0.006289
(1,1)	1.421710	0.260133	0.023519	62.522168
	0.527811	0.129706		0.000000
ig	1.762603	0.284765	0.002808	8.411464
	0.655154	0.246662		0.014910
e2AR	3.497705	-0.467882	0.002061	8.716246
	1.184789	0.509745		0.012802
¦e¦AR	3.193164	-0.359209	0.001271	9.800441
	1.080187	0.517417		0.007445
nonp	3.544731	-0.484216	0.004382	17.067878
	0.932468	0.363921		0.000197

Germany

homo	2.442757	0.022702	0.000005	6.816709
	0.965627	0.429929		0.033096
(1,1)	2.099971	0.154788	0.000431	8.559282
	0.731753	0.292273		0.013848
ig	1.632552	0.330894	0.006760	16.809337
	0.418656	0.168766		0.000224
e2AR	2.239759	0.116155	0.000184	10.249447
	0.699890	0.292640		0.005948
¦e¦AR	1.911209	0.297094	0.001007	9.835476
	0.695115	0.348887		0.007316
nonp	2.531624	-0.019580	0.000010	31.031218
	0.459384	0.201048		0.000000

Japan

homo	2.193381	0.041427	0.000007	2.800451
	1.557754	0.777924		0.246541
(1,1)	2.406710	-0.055748	0.000118	30.332684
	0.638799	0.208558		0.000000
ig	2.357051	-0.040506	0.000050	8.636156
	0.879299	0.357531		0.013325
e2AR	2.278134	-0.001711	0.000000	2.933709
	1.377205	0.651277		0.230650
¦e¦AR	1.874575	0.226185	0.000310	6.136690

	0.906833	0.491861		0.046498
nonp	1.681275	0.291734	0.002343	20.068651
	0.393200	0.173606		0.000044

U.K.

homo	3.842574	-0.561865	0.002865	8.133512
	1.382245	0.547753		0.017133
(1,1)	1.354690	0.476903	0.055629	26.092876
	0.375641	0.108220		0.000002
ig	2.250555	0.180471	0.002850	17.956412
	0.547917	0.207882		0.000126
e2AR	1.972567	0.327213	0.002997	4.862904
	1.031143	0.469124		0.087909
¦e¦AR	1.848324	0.439689	0.002452	6.546378
	1.065496	0.555244		0.037885
nonp	3.121600	-0.194439	0.002357	88.465924
	0.587645	0.141139		0.000000

X. Monte Carlo Experiment

We generated 100 samples of size 863, according to a process described in the next paragraph. We generated only 100 samples because of computational constraints; even this small experiment required nonlinear, iterative estimation of 40,800 GARCH models and 40,800 IGARCH models. For each of the 100 samples, we repeated the Table 4 calculations, fitting and then forecasting with all six models 408 times, fitting our models first to observations 1 to 432, then to observations 2 to 433,..., and finally to observations 408 to 839. We then computed RMSPEs and χ^2 statistics.

Our data generating process assumed e_t ~ iid $N(0,\sigma_1^2)$ for observations 1 to 432, e_t ~ iid $N(0,\sigma_2^2)$ for observations 433 to 863. We allowed for two distinct variances in part because of evidence presented below that the unconditional variance in these data appears to have drifted upward during the sample, in part because estimates of our GARCH models did not converge in an initial attempt to use a constant variance model. Recall that the conceptual experiment that underlies our asymptotic inference is one in which the number of observations used for estimation (=432, in the sample that we have) and the number used for forecasting (=408) both go to infinity. Here, we further assume that there is a one-time shift in the variance in the first observation that is forecast. Intuition suggests that our estimators might then yield equal out-of-sample RMSPEs in a large sample. (We have not, however, proved this formally.)

The standard deviations σ_1 and σ_2 were set to match estimates for U.K. exchange rates, σ_1=1.093 (3/14/73-6/17/81), σ_2=1.663 (6/14/81-9/20/89). The choice of the U.K. was arbitrary.

The attached has the results. The first part of the table repeats the ranks and RMSPEs reported in Table 4. Beneath the RMSPEs are 95 percent confidence intervals, for which the lower bound is the third smallest RMSPE, the upper bound the 98th smallest. The point estimates all fall within the 95 percent confidence intervals.

The bottom of the table repeats the χ^2 statistics and asymptotic p-values from Table 4, and adds the Monte Carlo p-values. The Monte Carlo p-value of "0.99" figure for H_A for the one week horizon, for example, indicates that 99 of the 100 samples yielded a χ^2 statistic larger than the 3.76 produced in our actual data. Seven of the remaining eight Monte-Carlo p-values also are higher than those derived from the asymptotic chi-square distribution, confirming a pattern of a small sample bias towards undersized tests that we have seen in related Monte Carlo work (Newey and West (1992)). There does not, then, appear to be a tendency for these tests to reject too infrequently.

Monte Carlo Results, U.K. Data

95 percent confidence intervals around point estimates of RMSPEs, from Monte Carlo:

	One Week Horizon		Twelve Week Horizon		Twenty Four Week Horizon	
	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE
homo	4	5.745 (4.284,7.536)	5	5.794 (4.152,7.557)	4	5.770 (4.142,7.443)
(1,1)	2	5.632 (4.267,7.785)	4	5.756 (4.129,8.011)	1	5.708 (4.136,8.031)
ig	1	5.563 (4.278,7.514)	2	5.692 (4.136,7.561)	5	5.834 (4.131,7.410)
e2AR	5	6.033 (4.321,7.804)	3	5.726 (4.158,7.792)	2	5.721 (4.133,7.639)
¦e¦AR	6	5.726 (4.336,7.649)	1	5.674 (4.170,7.653)	3	5.729 (4.186,7.529)
nonp	3	6.537 (4.309,7.723)	6	5.841 (4.319,7.594)	6	5.943 (4.188,7.750)

Hypothesis tests:

	χ^2	p-value		χ^2	p-value		χ^2	p-value	
		Asymptotic	Monte Carlo		Asymptotic	Monte Carlo		Asymptotic	Monte Carlo
H_A	3.76	[0.584]	[0.99]	8.49	[0.131]	[0.89]	5.82	[0.324]	[1.00]
H_B	1.52	[0.217]	[0.13]	1.29	[0.256]	[0.26]	0.07	[0.789]	[0.91]
H_C	3.59	[0.310]	[0.84]	5.66	[0.129]	[0.84]	1.95	[0.583]	[1.00]

Notes:
1. The entries in the "Rank" column and the point estimates of the RMSPEs are repeated from the Table 4 entries for the U.K., as are the χ^2 statistics and asymptotic p-values for H_A, H_B and H_C.
2. Parameters for the Monte Carlo simulation were chosen to match U.K. data in certain respects. Details are described on the previous page. For a given model and horizon, the numbers in parentheses are interpreted as: 2.5% of the artificial data sets had RMSPEs smaller than the first number, 2.5% had RMSPEs higher than the second number.
3. The Monte-Carlo p-value gives the fraction of artificial data sets for which the χ^2 statistic was higher than the one computed from the actual data.

XI. RMSPE for prediction made using average forecast

(Comparable to Table 4)

horizon	Canada		France		Germany		Japan		U.K.	
1	1	0.696	1	5.155	2	4.695	2	4.335	2	5.687
12	3	0.698	2	5.237	2	4.760	1	4.416	1	5.652
14	4	0.699	1	4.997	1	4.474	2	4.423	1	5.662

XII. RMSPE for the split samples. One-week

(Comparable to Table 4, Panel A)

A. 6/17/81 - 9/18/85

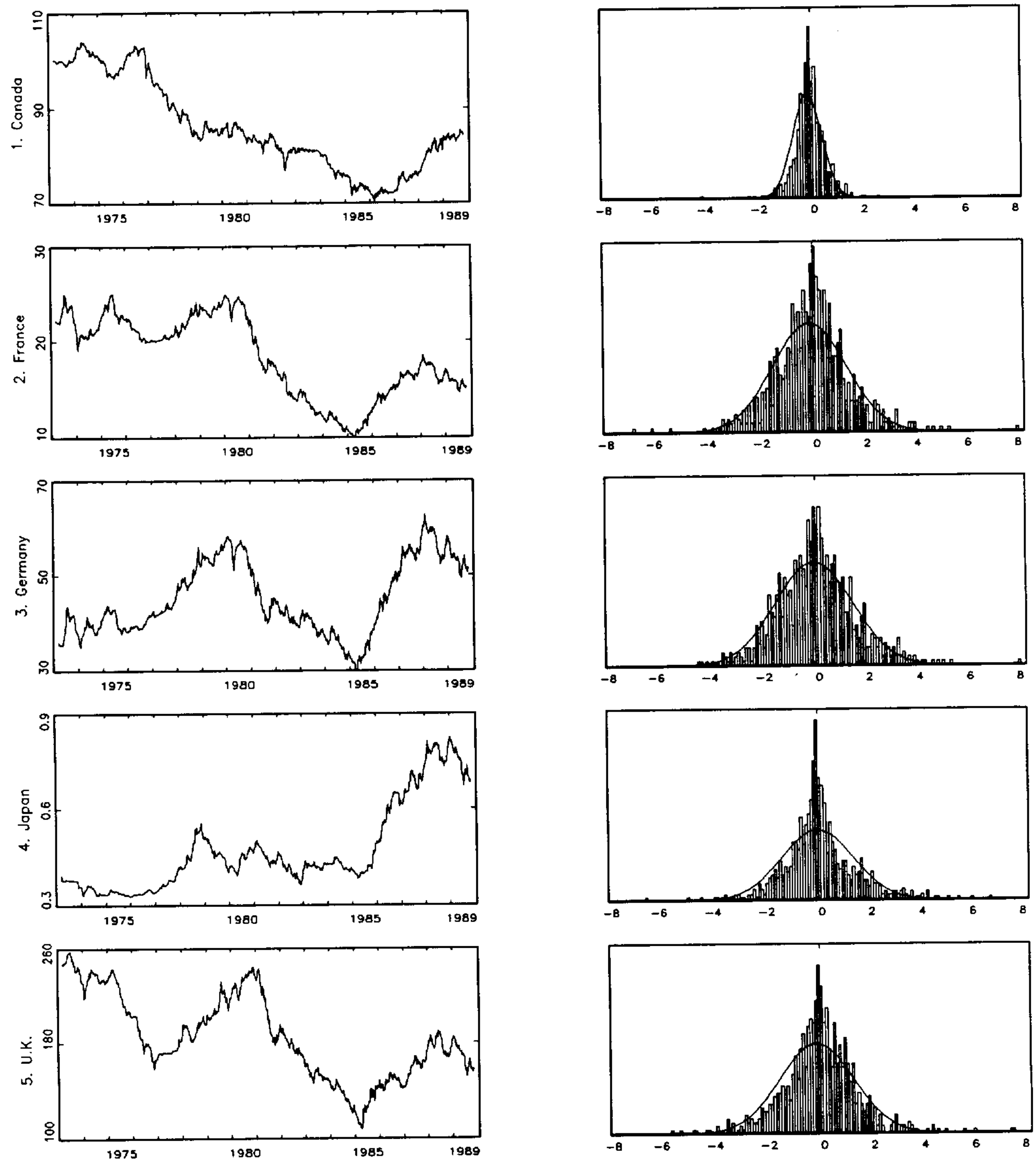
	Canada		France		Germany		Japan		U.K.	
homo	5	0.617	2	5.224	3	3.781	6	3.262	4	6.249
(1,1)	1	0.582	6	5.454	4	3.792	2	3.141	2	6.059
ig	2	0.599	1	5.219	1	3.731	1	3.119	1	5.955
e2AR	3	0.600	5	5.316	5	3.882	3	3.155	5	6.413
|e|AR	4	0.603	4	5.283	6	3.899	4	3.156	3	6.152
nonp	6	0.634	3	5.261	2	3.779	5	3.250	6	7.516

B. 9/18/85 - 4/5/89

	Canada		France		Germany		Japan		U.K.	
homo	2	0.814	2	5.097	1	5.602	2	5.412	3	5.085
(1,1)	4	0.821	6	5.225	5	5.737	1	5.396	2	5.082
ig	3	0.815	1	5.091	3	5.625	3	5.445	1	5.061
e2AR	5	0.825	5	5.221	6	5.925	6	5.539	6	5.552
|e|AR	1	0.807	3	5.100	2	5.622	4	5.498	5	5.177
nonp	6	0.842	4	5.129	4	5.641	5	5.530	4	5.144

Figure 1: Basic Data

A. Time Series of Levels B. Histograms of Log Differences



Notes:
1. In panel A, the vertical axis is measured in cents per unit of foreign currency.
2. In panel B, the vertical axis of the histograms is the relative frequency of the data falling in every .1 interval from -8 to 8. The corresponding densities were computed in the same way for each country using the normal density with the sample mean (row (1) in Table 1) and the sample variance (row (2)) for each country.

Table 1

Summary Statistics

A. e_t

	Canada	France	Germany	Japan	U.K.
(1)Mean	-0.020	-0.044	0.042	0.068	-0.052
	(0.020)	(0.050)	(0.056)	(0.056)	(0.055)
(2)Standard Deviation	0.552	1.408	1.466	1.361	1.406
	(0.029)	(0.053)	(0.064)	(0.063)	(0.080)
(3)Skewness	-0.423	0.103	0.480	0.385	0.261
	(0.455)	(0.245)	(0.203)	(0.233)	(0.270)
(4)Excess Kurtosis	4.981	2.490	2.133	2.587	3.092
	(2.636)	(0.846)	(0.900)	(0.715)	(0.857)
(5)Modified L-B(10)	7.47	20.89	13.43	20.52	13.57
	[.681]	[.022]	[.200]	[.025]	[.194]
(6)Modified L-B(50)	61.84	63.47	53.74	71.92	54.10
	[.122]	[.096]	[.333]	[.023]	[.321]
(7)Modified L-B(90)	111.99	98.46	87.99	122.26	88.47
	[.058]	[.254]	[.540]	[.005]	[.526]
(8)Min	-4.164	-6.825	-4.488	-6.587	-5.691
(9)Q1	-0.313	-0.851	-0.850	-0.641	-0.773
(10)Median	-0.030	0.000	0.023	-0.027	-0.034
(11)Q3	0.272	0.675	0.855	0.606	0.708
(12)Max	2.550	7.741	8.113	6.546	7.397

B. e_t^2

	Canada	France	Germany	Japan	U.K.
(13)Mean	0.305	1.983	2.147	1.854	1.978
	(0.030)	(0.148)	(0.190)	(0.166)	(0.223)
(14)Standard Deviation	0.809	4.196	4.395	4.001	4.446
	(0.213)	(0.598)	(0.763)	(0.513)	(0.768)
(15)L-B(10)	34.27	37.82	56.72	51.92	98.12
	[.000]	[.000]	[.000]	[.000]	[.000]
(16)L-B(50)	52.50	129.59	134.75	101.16	322.19
	[.377]	[.000]	[.000]	[.000]	[.000]
(17)L-B(90)	65.41	178.42	166.25	138.44	337.07
	[.976]	[.000]	[.000]	[.000]	[.000]

Notes:

1. The variable e_t is the percentage change in the weekly exchange rate. The sample includes 863 weekly observations from March 14, 1973 to September 20, 1989.
2. In rows (1)-(4), (13) and (14), heteroskedasticity and autocorrelation consistent asymptotic standard errors are in parentheses.
3. Rows (5) to (7) and (15) to (17) contain Ljung-Box statistics of order given in the header to the row. In rows (5) to (7), the statistics are computed as in equation (2-4) to allow for possible conditional heteroskedasticity in e_t. The p-values of the asymptotic chi-squared statistics are given in the lower halves of the rows.

Table 2

Models

(1) Model	(2) Formula for h_t	(3) Acronym
Homoskedastic Model		
1. Homoskedastic	$h_t = \omega$	homo
GARCH Models		
2. GARCH(1,1)	$h_t = \omega + \alpha e_t^2 + \beta h_{t-1}$	(1,1)
3. IGARCH(1,1)	$h_t = \alpha e_t^2 + (1-\alpha) h_{t-1}$	ig
Autoregressive models		
4. AR(12) in e_t^2	$h_t = \omega + \Sigma_{i=1}^{12} \alpha_i e_{t-i+1}^2$	e2AR
5. AR(12) in $\|e_t\|$	$h_t = (\pi/2)(E_t\|e_{t+1}\|)^2$; $E_t\|e_{t+1}\| = \omega + \Sigma_{i=1}^{12} \alpha_i \|e_{t-i+1}\|$	|e|AR
Nonparametric Model		
6. Gaussian kernel	$h_{t,j} = E(e_{t+j}^2 \| e_t)$; $\hat{h}_{t,j} = \Sigma_{t=1}^{N-j} w_{tN,j} e_{t+j}^2$, $w_{tN,j} = c_{tN,j} / \Sigma_{s=1}^{N-j} c_{sN,j}$, $c_{tN,j} = \exp[-.5(e_N - e_t)^2/b^2]$, b= bandwidth defined in text	nonp

Table 3

A. GARCH (1,1) Estimates, Sample = 3/14/73 to 6/17/81

	ω (x10^5)	α	β
1. Canada	0.5	0.26	0.54
	(0.1)	(0.02)	(0.06)
2. France	1.3	0.35	0.61
	(0.2)	(0.05)	(0.04)
3. Germany	2.0	0.30	0.61
	(0.4)	(0.05)	(0.04)
4. Japan	0.07	0.05	0.94
	(0.03)	(0.01)	(0.01)
5. U.K.	1.9	0.11	0.73
	(0.7)	(0.05)	(0.10)

B. Population Root Mean Square Prediction Errors, Homoskedastic Relative to GARCH (1,1)

Horizon	(1) α=.26,β=.54	(2) α=.35,β=.61	(3) α=.30,β=.61	(4) α=.05,β=.94	(5) α=.11,β=.73
1	1.09	1.60	1.23	1.06	1.02
12	1.00	1.15	1.02	1.05	1.00
24	1.00	1.05	1.00	1.04	1.00

Notes:
1. The numbers in parentheses in panel A are asymptotic standard errors.
2. Panel B presents the ratio of RMSPEs for the indicated horizons, computed assuming that the data are driven by a GARCH(1,1) model with the indicated parameters, and abstracting from sampling error in estimation of the model parameters. The ratio is invariant to ω. The RMSPE for the homoskedastic model is constant for all horizons. The ratio asymptotes to 1 as the horizon approaches infinity, for each pair of α and β.

Table 4

Root Mean Squared Prediction Errors

A. One Week Horizon

	Canada		France		Germany		Japan		U.K.	
	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE
homo	5	0.714	2	5.167	2	4.704	3	4.380	4	5.745
(1,1)	1	0.702	6	5.351	5	4.783	1	4.323	2	5.632
ig	3	0.706	1	5.161	1	4.695	2	4.343	1	5.563
e2AR	4	0.712	5	5.273	6	4.925	5	4.411	5	6.033
|e|AR	2	0.704	3	5.200	4	4.767	4	4.388	3	5.726
nonp	6	0.737	4	5.201	3	4.724	6	4.442	6	6.537
H_A	9.70 [0.084]		8.91 [0.113]		8.23 [0.144]		6.42 [0.268]		3.76 [0.584]	
H_B	1.24 [0.265]		0.01 [0.918]		0.01 [0.912]		0.77 [0.380]		1.52 [0.217]	
H_C	4.24 [0.237]		8.99 [0.029]		4.13 [0.247]		2.86 [0.413]		3.59 [0.310]	

B. Twelve Week Horizon

	Canada		France		Germany		Japan		U.K.	
	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE
homo	1	0.695	1	5.219	1	4.754	3	4.435	5	5.794
(1,1)	2	0.697	6	5.696	6	4.831	5	4.451	4	5.756
ig	6	0.731	5	5.268	5	4.817	6	4.454	2	5.692
e2AR	3	0.700	3	5.251	4	4.796	2	4.433	3	5.726
|e|AR	4	0.701	4	5.267	3	4.785	1	4.430	1	5.674
nonp	5	0.704	2	5.250	2	4.762	4	4.447	6	5.841
H_A	16.71 [0.005]		15.06 [0.010]		7.32 [0.198]		1.08 [0.956]		8.49 [0.131]	
H_B	n.a.		n.a.		n.a.		0.01 [0.921]		1.29 [0.256]	
H_C	5.96 [0.114]		13.60 [0.004]		5.83 [0.120]		0.39 [0.943]		5.66 [0.129]	

C. Twenty Four Week Horizon

	Canada		France		Germany		Japan		U.K.	
	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE
homo	1	0.695	3	5.094	3	4.500	2	4.424	4	5.770
(1,1)	5	0.703	6	5.694	1	4.490	6	4.498	1	5.708
ig	6	0.743	1	5.060	5	4.509	5	4.483	5	5.834
e2AR	2	0.695	2	5.087	2	4.498	1	4.422	2	5.721
|e|AR	3	0.697	4	5.109	4	4.505	4	4.441	3	5.729
nonp	4	0.702	5	5.131	6	4.535	3	4.436	6	5.943
H_A	18.95 [0.002]		18.08 [0.003]		3.80 [0.578]		6.05 [0.301]		5.82 [0.324]	
H_B	n.a.		0.25 [0.619]		0.11 [0.741]		0.12 [0.728]		0.07 [0.789]	
H_C	3.35 [0.340]		17.64 [0.001]		1.07 [0.785]		5.82 [0.121]		1.95 [0.583]	

Notes:

1. The "RMSPE" columns present the out of sample root mean squared error in predicting e^2_{t+j} for horizon j (j=1, 12 or 24) and the indicated country and model. The "Rank" columns index the relative size of the RMSPEs for a given country and horizon, 1 indicating the smallest RMSPE, 6 the largest.

2. The H_A, H_B and H_C rows present χ^2 statistics (asymptotic p-values in brackets) for the following hypothesis: A: equality of MSPEs of all 6 models ($\chi^2(5)$); B: equality of MSPEs of best and homo models ($\chi^2(1)$); C: equality of MSPEs from homo, (1,1), e2AR and |e|AR models ($\chi^2(3)$). The statistics are computed as in equation (2-5).

Table 5

Regression Tests of Efficiency, One Week Horizon

	b_0	b_1	R^2	$\chi^2(2)$	b_0	b_1	R^2	$\chi^2(2)$
	Canada				France			
homo	1.34 (0.53)	-2.99* (1.62)	0.018	8.07 [0.018]	3.92 (1.13)	-0.71 (0.50)	0.004	12.15 [0.002]
(1,1)	0.15 (0.06)	0.60** (0.17)	0.044	6.98 [0.031]	2.26 (0.40)	0.09 (0.14)	0.0009	45.45 [0.000]
ig	0.17 (0.05)	0.55** (0.17)	0.037	10.04 [0.007]	2.23 (0.67)	0.10 (0.24)	0.0004	13.53 [0.001]
e2AR	0.20 (0.09)	0.48* (0.28)	0.019	4.90 [0.086]	2.86 (0.41)	-0.15 (0.13)	0.001	82.32 [0.000]
|e|AR	0.16 (0.09)	0.66** (0.29)	0.031	5.44 [0.066]	2.38 (0.48)	0.05 (0.22)	0.0001	25.05 [0.000]
nonp	0.28 (0.05)	0.27** (0.13)	0.012	31.66 [0.000]	2.86 (0.57)	-0.16 (0.22)	0.0008	27.63 [0.000]
	Germany				Japan			
homo	2.65 (0.74)	-0.06 (0.35)	0.00003	14.52 [0.001]	3.40 (1.37)	-0.59 (0.70)	0.001	7.46 [0.024]
(1,1)	1.95 (0.44)	0.23 (0.21)	0.005	19.85 [0.000]	0.89 (0.79)	0.60 (0.37)	0.023	1.25 [0.537]
ig	1.56 (0.58)	0.37 (0.27)	0.008	7.93 [0.019]	1.02 (0.63)	0.60** (0.28)	0.011	2.66 [0.264]
e2AR	2.45 (0.28)	0.03 (0.11)	0.0001	92.85 [0.000]	1.56 (0.47)	0.32 (0.22)	0.008	11.09 [0.004]
|e|AR	2.18 (0.54)	0.15 (0.27)	0.001	19.73 [0.000]	1.43 (0.42)	0.40* (0.21)	0.012	11.39 [0.003]
nonp	2.48 (0.59)	0.02 (0.25)	0.00001	17.94 [0.000]	2.51 (0.73)	-0.14 (0.32)	0.0006	12.44 [0.002]
	U.K.							
homo	3.55 (0.81)	-0.42 (0.36)	0.002	19.18 [0.000]				
(1,1)	1.25 (0.48)	0.53** (0.22)	0.045	7.20 [0.027]				
ig	0.95 (0.61)	0.69** (0.30)	0.042	3.90 [0.142]				
e2AR	2.40 (0.44)	0.12 (0.13)	0.003	45.07 [0.000]				
|e|AR	1.84 (0.51)	0.36 (0.23)	0.010	13.08 [0.001]				
nonp	2.78 (0.32)	-0.03 (0.04)	0.0003	943.63 [0.000]				

Notes:
1. This reports results of the regression $e_{t+1}^2 = b_0 + b_1\hat{h}_{mt} + \epsilon_t$. For b_0 and b_1, heteroskedasticity and autocorrelation consistent standard errors are in parentheses. The $\chi^2(2)$ tests H_0: $b_0=0$, $b_1=1$, with asymptotic p-value in brackets.
2. For b_1, "**" denotes significance at the 5 percent level, "*" at the 10 percent level.

Table 6

Subsample Statistics on e_t

A. Standard Deviation

	Canada	France	Germany	Japan	U.K.
(1)Standard Deviation, 3/14/73-6/17/81	0.499 (0.049)	1.185 (0.097)	1.309 (0.117)	1.174 (0.102)	1.093 (0.076)
(2)Standard Deviation, 6/24/81-9/20/89	0.600 (0.050)	1.603 (0.153)	1.609 (0.149)	1.526 (0.146)	1.663 (0.159)
(3) (2)-(1)	0.101 (0.070)	0.418 (0.181)	0.299 (0.190)	0.352 (0.178)	0.570 (0.177)

B. Modified Range Scale Tests for Constancy of Unconditional Variance

Sample	No. of Obs.	Canada	France	Germany	Japan	U.K.
(1)3/14/73-9/20/89	863	1.428	2.238**	1.901**	1.874**	1.753**
(2)3/14/73-6/17/81	432	1.413	1.659*	1.540	1.892**	1.046
(3)4/24/77-7/31/85	432	1.242	1.725*	1.365	1.852**	1.317
(4)6/24/81-9/20/89	431	1.260	1.226	1.326	1.561	1.667

Notes:

1. In panel A, heteroskedasticity and autocorrelation consistent asymptotic standard errors are in parentheses.
2. In panel B, let $x_t=(e_t-\bar{e})^2$, where $\bar{e}$ is the mean of e_t in the sample in question, and let $\bar{x}$ be the corresponding mean of x_t. Let $\psi(r) = [\Sigma_{t=1}^{r}(x_t-\bar{x})] / (T\hat{s})^{1/2}$, where $1 \le r \le T$, $T=431,432$ or 863 is the sample size and $\hat{s}$ is an estimate of the asymptotic variance of $T^{-1/2}\Sigma_{t=1}^{T}[(e_t-Ee_t)^2 - E(e_t-Ee_t)^2]$. The table reports the difference between the maximum and minimum of $\psi(r)$.
3. "**" means significant at the .05 level, "*" at the .10 level, according to Table 1a in Haubrich and Lo (1989).