JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **23**: 65–90 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.976

STRUCTURAL BREAKS AND GARCH MODELS OF EXCHANGE RATE VOLATILITY

DAVID E. RAPACH* AND JACK K. STRAUSS
Department of Economics, Saint Louis University, Saint Louis, Missouri, USA

SUMMARY

We investigate the empirical relevance of structural breaks for GARCH models of exchange rate volatility using both in-sample and out-of-sample tests. We find significant evidence of structural breaks in the unconditional variance of seven of eight US dollar exchange rate return series over the 1980–2005 period—implying unstable GARCH processes for these exchange rates—and GARCH(1,1) parameter estimates often vary substantially across the subsamples defined by the structural breaks. We also find that it almost always pays to allow for structural breaks when forecasting exchange rate return volatility in real time. Combining forecasts from different models that accommodate structural breaks in volatility in various ways appears to offer a reliable method for improving volatility forecast accuracy given the uncertainty surrounding the timing and size of the structural breaks. Copyright © 2008 John Wiley & Sons, Ltd.

Received 30 November 2005; Revised 1 November 2006

1. INTRODUCTION

Spurred by the seminal contributions of Engle (1982) and Bollerslev (1986), an extensive literature models time-varying volatility in asset returns using generalized autoregressive conditional heteroskedastic (GARCH) processes. A number of studies use GARCH models—especially GARCH(1,1) models—to characterize and forecast time-varying volatility in exchange rate returns; see, for example, Engle and Bollerslev (1986), Baillie and Bollerslev (1989, 1991), Bollerslev and Engle (1993), West and Cho (1995), Neely (1999), and Hansen and Lunde (2005). Accurately modeling and forecasting time-varying volatility in exchange rate returns have important implications for financial decision-making, including the pricing of derivatives and portfolio risk management in an international setting. In addition, a large body of theoretical research ties exchange rate volatility to trade and welfare.[1]

Researchers often assume (explicitly or implicitly) that a stable GARCH process governs conditional exchange rate return volatility, so that the unconditional variance of exchange rate returns is constant. However, international financial markets are periodically subject to sudden 'large' shocks, such as the Exchange Rate Mechanism crisis in Europe in the early 1990s and the East Asian crisis of the late 1990s. These types of shocks can cause abrupt breaks in the unconditional variance of exchange rate returns and are equivalent to structural breaks in the parameters of the GARCH processes governing the conditional volatility of exchange rate returns.

* Correspondence to: David E. Rapach, Department of Economics, St Louis University, 3674 Lindell Boulevard, St Louis, MO 63108-3397, USA. E-mail: rapachde@slu.edu

Contract/grant sponsor: John Cook School of Business, St Louis University.

[1] See Clark *et al.* (2004) for a recent survey.

Recent theoretical studies show that structural breaks in unconditional variance have potentially important implications for estimated GARCH models of exchange rate return volatility. Building on insights from Diebold (1986), Hendry (1986), and Lamoureux and Lastrapes (1990), recent research by Mikosch and Stărică (2004) and Hillebrand (2005) demonstrates that neglected structural breaks in the parameters of GARCH processes induce upward biases in estimates of the persistence of GARCH processes.[2] This is a relevant concern, as estimated GARCH(1,1) models of exchange rate return volatility in the extant literature are typically highly persistent. More generally, failure to account for structural breaks can obviously lead to poor estimates of the unconditional volatility of exchange rate returns for GARCH processes.

Structural breaks also have potentially important implications for forecasts of exchange rate return volatility. In out-of-sample volatility forecasting exercises, the use of an expanding data window (or a fixed data window to reduce computational costs) to estimate GARCH forecasting models is common and appropriate under the assumption of a stable GARCH process. However, this approach may not perform well in the presence of sudden breaks in volatility. Along this line, West and Cho (1995) posit that the forecasting performance of GARCH(1,1) models of exchange rate return volatility could be improved by allowing for structural breaks in the unconditional variance of exchange rate returns. In addition, recent research by Stărică *et al.* (2005) shows that long-horizon forecasts of stock return volatility generated by GARCH(1,1) models assuming parameter stability are often inferior to forecasts that allow for frequent changes in the unconditional variance of stock returns.

In light of the above considerations, the present paper investigates the empirical relevance of structural breaks for GARCH(1,1) models of exchange rate return volatility using both in-sample and out-of-sample tests. In our in-sample analysis, we employ a modified version of the Inclán and Tiao (1994) iterated cumulative sum of squares algorithm that allows for dependent processes. We use the algorithm to test for (potentially multiple) structural breaks in the unconditional variance of daily US dollar exchange rate returns *vis-à-vis* the currencies of seven OECD countries, as well as daily returns for a trade-weighted US dollar exchange rate. We find significant evidence of structural breaks in the unconditional variance for seven of the eight exchange rate return series, and inspection of the estimated GARCH(1,1) processes across the subsamples defined by the structural breaks often reveals sharp differences in parameter estimates.

In our out-of-sample analysis, we consider various methods of accommodating potential structural breaks when forming exchange rate return volatility forecasts in real time. Three popular forecasting models estimated using an expanding window—the GARCH(1,1), RiskMetrics, and fractionally integrated GARCH(1,d,1) models—serve as natural benchmarks. We compare forecasts of daily exchange rate return volatility generated by the three benchmark models to forecasts generated by five competing models, each of which makes some type of adjustment, typically to the estimation window, in order to account for potential structural breaks in the unconditional variance of exchange rate returns.

The first two competing models use rolling windows with sizes equal to one-half and one-quarter of the length of the in-sample period when estimating the GARCH(1,1) forecasting

[2] Mikosch and Stărică (2004) show this for Whittle estimators, while Hillebrand (2005) shows this for a wide class of estimators, including maximum likelihood and quasi maximum likelihood estimators, the most common estimators of GARCH models in the literature.

model.[3] In out-of-sample volatility forecasting exercises, researchers sometimes apply a rolling estimation window to allow the parameters of GARCH processes to evolve over time. The third competing model uses an expanding window in conjunction with a weighted maximum likelihood procedure that places less weight on observations in the more distant past when estimating the parameters of the GARCH(1,1) forecasting model. Mittnik and Paolella (2000) suggest this model as a way of handling structural instabilities in GARCH parameters. The fourth competing model applies the modified iterative cumulative sum of squares algorithm to the observations available at the time of forecast formation to determine the estimation window for the GARCH(1,1) forecasting model. The final competing model simply averages the daily squared returns over the previous 250 days, thereby allowing for a frequently changing unconditional variance. Stărică *et al.* (2005) find that this model often outperforms a GARCH(1,1) model estimated using an expanding window with respect to forecasting daily stock return volatility in a large number of industrialized countries. Considering loss functions based on mean square forecast error (MSFE) and value-at-risk (VaR), as well as forecast horizons of 1, 20, 60, and 120 days, we find that it almost always pays to accommodate potential structural breaks when forming out-of-sample volatility forecasts. Overall, structural breaks appear to be an empirically relevant property of exchange rate volatility which should be accounted for when using GARCH models (and presumably other types of models) to forecast exchange rate volatility.

The rest of the paper is organized as follows. Section 2 outlines our econometric methodology. Section 3 presents the empirical results for the in-sample and out-of-sample tests. Section 4 concludes.

2. ECONOMETRIC METHODOLOGY

2.1. In-Sample Tests

Let $e_t = 100\log(E_t/E_{t-1})$, where E_t is the nominal exchange rate at the end of period t, so that e_t is the percent return for the exchange rate from period $t-1$ to period t. Following West and Cho (1995), we treat the unconditional and conditional mean of e_t as zero. (The eight exchange rate return series we consider in Section 3 below satisfy this assumption.) Suppose we observe e_t for $t = 1, \ldots, T$ and are interested in testing whether the unconditional variance of e_t is constant over the available sample. A structural break in the unconditional variance implies a structural break in the GARCH process governing conditional volatility. Inclán and Tiao (1994) develop a cumulative sum of squares statistic to test the null hypothesis of a constant unconditional variance against the alternative hypothesis of a break in the unconditional variance. The Inclán and Tiao (1994) statistic is given by

$$IT = \sup_k |(T/2)^{0.5} D_k| \tag{1}$$

where $D_k = (C_k/C_T) - (k/T)$ and $C_k = \sum_{t=1}^{k} e_t^2$ for $k = 1, \ldots, T$. The value of k that maximizes $|(T/2)^{0.5} D_k|$ is the estimate of the break date. When e_t is distributed i.i.d. $N(0, \sigma_e^2)$, Inclán and

[3] The choices of window size are admittedly somewhat arbitrary. Note that in situations with fewer in-sample observations than ours it may be necessary to select a 'small' estimation window with a size greater than one-quarter of the length of the in-sample period.

DOI: 10.1002/jae

Tiao (1994) show that the asymptotic distribution of the *IT* statistic is given by $\sup_r |W^*(r)|$, where $W^*(r) = W(r) - rW(1)$ is a Brownian bridge and $W(r)$ is standard Brownian motion. Simulation methods can be used to generate finite-sample critical values.

The *IT* statistic is designed for i.i.d. processes, and Andreou and Ghysels (2002), de Pooter and van Dijk (2004), and Sansó *et al.* (2004) show that the *IT* statistic can be substantially oversized when e_t follows a dependent process, such as a GARCH process.[4] A nonparametric adjustment can be applied to the *IT* statistic that allows e_t to obey a wide class of dependent processes, including GARCH processes, under the null hypothesis (Kokoszka and Leipus, 1999; Lee and Park, 2001; Sansó *et al.*, 2004). We use a nonparametric adjustment based on the Bartlett kernel, and the adjusted statistic is given by

$$AIT = \sup_k |T^{-0.5} G_k| \tag{2}$$

where $G_k = \hat{\lambda}^{-0.5}[C_k - (k/T)C_T]$, $\hat{\lambda} = \hat{\gamma}_0 + 2\sum_{l=1}^{m}[1 - l(m+1)^{-1}]\hat{\gamma}_l$, $\hat{\gamma}_l = T^{-1}\sum_{t=l+1}^{T}(e_t^2 - \hat{\sigma}^2)(e_{t-l}^2 - \hat{\sigma}^2)$, $\hat{\sigma}^2 = T^{-1}C_T$, and the lag truncation parameter m is selected using the procedure in Newey and West (1994). Under general conditions, the asymptotic distribution of *AIT* is also given by $\sup_r |W^*(r)|$, and finite-sample critical values can again be generated via simulation.

To test for multiple breaks in unconditional volatility, Inclán and Tiao (1994) develop an iterated cumulative sum of squares (ICSS) algorithm based on the *IT* statistic; see Steps 0–3 in Inclán and Tiao (1994: 916). Alternatively, the ICSS algorithm can be based on the *AIT* statistic to avoid the size distortions that plague the *IT* statistic when e_t follows a dependent process. In our applications in Section 3 below, we use the ICSS algorithm based on the *AIT* statistic (the 'modified ICSS algorithm') and the 5% significance level to test for multiple breaks in the unconditional variance of eight daily US dollar exchange rate return series. If we detect significant evidence of at least one structural break, we estimate GARCH(1,1) models over the different regimes defined by the significant structural breaks and compare these models to a GARCH(1,1) model estimated over the full sample.

The canonical GARCH(1,1) model for e_t with mean zero (conditional and unconditional) takes the form

$$e_t = h_t^{0.5}\varepsilon_t \tag{3}$$

$$h_t = \omega + \alpha e_{t-1}^2 + \beta h_{t-1} \tag{4}$$

where ε_t is i.i.d. with mean zero and unit variance. In order to ensure that the conditional variance, h_t, is positive, we require $\omega > 0$ and $\alpha, \beta \geq 0$. The GARCH(1,1) process specified in equations (3) and (4) is stationary if $\alpha + \beta < 1$; when $\alpha + \beta = 1$, we have the integrated GARCH(1,1) or IGARCH(1,1) model of Engle and Bollerslev (1986). For a stationary GARCH(1,1) process, the unconditional variance for e_t is given by $\omega/(1 - \alpha - \beta)$. Note that when $\alpha = 0$ in equation (4), β is unidentified (and set to zero), so that $h_t = \omega$ and e_t is characterized by conditional homoskedasticity. The GARCH(1,1) model is often estimated using quasi maximum likelihood estimation (QMLE), where the likelihood function corresponding

[4] The size distortions tend to increase with the sample size.

DOI: 10.1002/jae

to $\varepsilon_t \sim N(0,1)$ is used and the restrictions $\omega > 0$ and $\alpha, \beta \geq 0$ are imposed, and we follow this in the present paper. The QMLE parameter estimates are consistent and asymptotically normal; see, for example, Berkes *et al.* (2003), Jensen and Rahbek (2004), or Straumann (2005).[5]

2.2. Out-of-Sample Tests

We compare out-of-sample forecasts of exchange rate return volatility generated by three benchmark forecasting models and five competing forecasting models. The first benchmark model is a GARCH(1,1) model estimated using an expanding window ('GARCH(1,1) expanding window' model). More specifically, we divide the sample for a given exchange rate return series into in-sample and out-of-sample portions, where the in-sample portion spans the first R observations and the out-of-sample portion the last P observations. In order to generate the first out-of-sample forecast at the 1-period horizon, we estimate the GARCH(1,1) model given by equations (3) and (4) using QMLE and data from the first observation through observation R. The initial forecast is given by $\hat{h}_{R+1|R,\mathrm{EXP}} = \hat{\omega}_{R,\mathrm{EXP}} + \hat{\alpha}_{R,\mathrm{EXP}} e_R^2 + \hat{\beta}_{R,\mathrm{EXP}} \hat{h}_{R,\mathrm{EXP}}$, where $\hat{\omega}_{R,\mathrm{EXP}}$, $\hat{\alpha}_{R,\mathrm{EXP}}$, $\hat{\beta}_{R,\mathrm{EXP}}$, and $\hat{h}_{R,\mathrm{EXP}}$ are the estimates of ω, α, β, and h_R, respectively, in equation (4) obtained using data from the first observation through observation R. We then expand the estimation window by one observation in order to form a forecast for period $R+2$ using data from the first observation through observation $R+1$, $\hat{h}_{R+2|R+1,\mathrm{EXP}}$. We proceed in this manner through the end of the available out-of-sample period, leaving us with a series of P out-of-sample forecasts, $\{\hat{h}_{t|t-1,\mathrm{EXP}}\}_{t=R+1}^{T}$. The GARCH(1,1) expanding window model is a natural benchmark that is appropriate for forecasting when the data are generated by a stable GARCH(1,1) process.

The second benchmark model we consider is the RiskMetrics model based on an expanding window, a popular model often included in studies of out-of-sample volatility forecasting exercises. The RiskMetrics forecast is given by the exponential weighted moving average formula, $\hat{h}_{t+1|t} = (1-\lambda)\sum_{k=0}^{t-1} \lambda^k e_{t-k}^2$, where following the recommendation of the RiskMetrics Group (1996) for daily data we use $\lambda = 0.94$. Note that the RiskMetrics model does not involve the estimation of any parameters, making it easy to implement. We denote the forecasts for the RiskMetrics model by $\{\hat{h}_{t|t-1,\mathrm{RM}}\}_{t=R+1}^{T}$.

The third benchmark model is the fractionally integrated GARCH(1,d,1) or FIGARCH(1,d,1) model of Baillie *et al.* (1996) estimated using an expanding window ('FIGARCH(1,d,1) expanding window' model). This is a relevant benchmark, as it is well known that the autocorrelations of squared (or absolute) returns for many financial assets decay slower than exponentially (as implied by GARCH models), so that conditional heteroskedasticity may be better described by a long-memory process as captured by the FIGARCH(1,d,1) specification.[6] Furthermore, Diebold and Inoue (2001: 157) opine that even if structural change

[5] We focus on the GARCH(1,1) model specification in the present paper and do not consider GARCH(p, q) specifications where p and/or q are greater than one, as the GARCH(1,1) specification is the canonical specification in the literature on exchange rate volatility modeling. The Lagrange multiplier tests of Bollerslev (1986) generally do not indicate the presence of higher-order ARCH or GARCH terms for the exchange rate return series that we consider.

[6] The FIGARCH(1,d,1) specification is given by $h_t = \omega + \beta h_{t-1} + [1 - \beta L - (1-\phi L)(1-L)^d]e_t^2$, where L is the lag operator and $(1-L)^d = 1 - dL - [d(1-d)/2]L^2 - [d(1-d)(2-d)/6]L^3 - \ldots$ is the fractional differencing operator. We estimate the parameter vector, $(\omega, \beta, \phi, d)'$, using QMLE. Following Baillie *et al.* (1996), we assume that $\omega > 0$, $0 \leq d \leq 1 - 2\phi$, and $0 \leq \beta \leq \phi + d$ to ensure that the conditional variance is positive. We thank Richard Baillie for providing GAUSS code that was used to compute the log-likelihood function for the FIGARCH(1,d,1) model.

is a true characteristic of the underlying data-generating process, long memory can be a 'convenient shorthand description, which may remain very useful for tasks such as prediction'.[7] We denote the forecasts for the FIGARCH(1,d,1) expanding window model by $\{\hat{h}_{t|t-1,\mathrm{FI}}\}_{t=R+1}^{T}$.

The five competing forecasting models all make adjustments, usually to the estimation window, in order to accommodate potential changes in the unconditional variance of exchange rate returns. The first competing model is a GARCH(1,1) model estimated using a rolling window with size equal to one-half of the length of the in-sample period ('GARCH(1,1) 0.50 rolling window' model).[8] We denote the forecasts for the GARCH(1,1) 0.50 rolling window model by $\{\hat{h}_{t|t-1,\mathrm{ROLL}(0.5)}\}_{t=R+1}^{T}$. A rolling window with size equal to one-half of the length of the in-sample period represents a compromise between having a relatively long estimation window to accurately estimate the parameters of the GARCH(1,1) process and not relying too extensively on data from potentially separate regimes. We also consider a GARCH(1,1) model estimated using a shorter rolling window with size equal to one-quarter of the length of the in-sample period (the second competing model; 'GARCH(1,1) 0.25 rolling window' model). By using a shorter estimation window, this forecasting model has fewer observations available for estimating the parameters of the GARCH(1,1) process,[9] but it is less likely to use data from different regimes. We denote the forecasts generated by the GARCH(1,1) 0.25 rolling window model by $\{\hat{h}_{t|t-1,\mathrm{ROLL}(0.25)}\}_{t=R+1}^{T}$.

The third competing model is a GARCH(1,1) model estimated using an expanding window and a weighted maximum likelihood procedure, following Mittnik and Paolella (2000; 'GARCH(1,1) weighted ML' model). The GARCH(1,1) weighted ML model assigns declining weights to observations in the more distant past when forming the likelihood function used to estimate the GARCH(1,1) model parameters.[10] Similar to the rolling window models, the GARCH(1,1) weighted ML model allows the GARCH(1,1) parameters to evolve over time, and Mittnik and Paolella (2000) recommend this forecasting model as a way of handling structural instabilities in GARCH parameters. We denote the forecasts generated by the GARCH(1,1) weighted ML model by $\{\hat{h}_{t|t-1,\mathrm{WML}}\}_{t=R+1}^{T}$.

The fourth competing model uses the modified ICSS algorithm to select the estimation window for the GARCH(1,1) model ('GARCH(1,1) with breaks' model). We first apply the modified ICSS algorithm to observations one through R. Suppose we find significant evidence of one or more structural breaks according to the modified ICSS algorithm and that the final break is estimated to occur at time T_B. We then estimate a GARCH(1,1) model using observations $T_B + 1$ through R

[7] The FIGARCH(1,d,1) model is also a relevant benchmark in light of Mikosch and Stărică (2003, 2004) and Perron and Qu (2006), who show that structural breaks can give rise to spurious evidence of long-range dependence or long memory in financial volatility data.

[8] That is, the first forecast uses estimates of equation (4) based on observations $0.5R + 1$ through R, the second forecast uses estimates based on observations $0.5R + 2$ through $R + 1$, and so on.

[9] In our applications in Section 3 below, this leaves us with approximately 1500 observations for estimating the GARCH(1,1) 0.25 rolling window model. This should be enough observations to obtain reasonably reliable estimates of the GARCH parameters; see, for example, Hwang and Valls Pereira (2006) and Straumann (2005).

[10] More specifically, when forming the initial out-of-sample forecast using data through observation R, a weight of ρ^{R-t} is attached to observation t ($t = 1, \ldots, R$) in the log-likelihood function used to estimate the GARCH(1,1) parameters. The window is then expanded by one observation, and a weight of ρ^{R+1-t} is attached to observation t ($t = 1, \ldots, R + 1$) in the log-likelihood function used to estimate the GARCH(1,1) parameters that generate the second out-of-sample forecast. We continue in this manner through the end of the available out-of-sample period. We use $\rho = 0.994$, as Mittnik and Paolella (2000) generally find that this value works well in out-of-sample volatility forecasting exercises involving a number of exchange rate return series.

DOI: 10.1002/jae

to form an estimate of h_{R+1}. If there is no significant evidence of a structural break according to the modified ICSS algorithm, we estimate a GARCH(1,1) model using observations one through R to form an estimate of h_{R+1}. To compute the second out-of-sample forecast, we apply the modified ICSS algorithm to observations one through $R+1$ and proceed as described above. Continuing in this manner through the end of the available out-of-sample period, we generate a series of forecasts corresponding to the GARCH(1,1) with breaks model, $\{\hat{h}_{t|t-1,\mathrm{BREAKS}}\}_{t=R+1}^{T}$. Note well that there is no 'look-ahead' bias involved in the generation of the forecasts for the GARCH(1,1) with breaks model, as the modified ICSS algorithm that determines the size of the estimation window only uses data available at the time of the forecast formation. A potential drawback to this forecasting model is that a relatively short sample will be available for estimating the GARCH(1,1) parameters when a break is detected relatively close to the forecast date.[11]

The final forecasting model is a simple moving average model that uses the average of the squared returns over the previous 250 days to form the volatility forecast for day t : $\hat{h}_{t|t-1,\mathrm{MA}} = (1/250)\sum_{i=1}^{250} e_{t-i}^2$. This model allows the unconditional variance to change frequently over time. As noted in the introduction, Stărică *et al.* (2005) find that this model often outperforms a GARCH(1,1) model estimated using an expanding window at longer horizons when forecasting daily stock return volatility in industrialized countries.

In our applications in Section 3 below, we consider forecast horizons of 1, 20, 60, and 120 days. In general, for model i ($i =$ EXP,RM,FI,ROLL(0.5),ROLL(0.25),WML,BREAKS,MA), we denote the model i forecast of h_t formed at period $t-s$ (and thus based on information available at period $t-s$) by $\hat{h}_{t|t-s,i}$, leaving us with a series of $P-(s-1)$ s-step-ahead out-of-sample forecasts, $\{\hat{h}_{t|t-s,i}\}_{t=R+s}^{T}$. We generate $\hat{h}_{t|t-s,i}$ for $s>1$ using equation (4.117) in Franses and van Dijk (2000) to iterate forward using the fitted GARCH(1,1) processes. Following convention, the RiskMetrics and moving average forecasts ($\hat{h}_{t|t-s,\mathrm{RM}}$ and $\hat{h}_{t|t-s,\mathrm{MA}}$, respectively) for $s>1$ are equal to the 1-step-ahead forecast. We emphasize that our out-of-sample forecasting exercise mimics the situation of a forecaster in real time.

In order to compare forecasts across models, we consider two loss functions. We first follow Stărică and Granger (2005) and Stărică *et al.* (2005) and use an aggregated version of the familiar MSFE metric:

$$\mathrm{MSFE}_{s,i}^{*} = [P-(s-1)]^{-1}\sum_{t=R+s}^{T}(\tilde{e}_t^2 - \tilde{\hat{h}}_{t|t-s,i})^2 \tag{5}$$

where $\tilde{e}_t^2 = \sum_{j=1}^{s} e_{t-(j-1)}^2$ and $\tilde{\hat{h}}_{t|t-s,i} = \sum_{j=1}^{s}\hat{h}_{t-(j-1)|t-s,i}$. As emphasized by Andersen and Bollerslev (1998), squared returns are a very noisy proxy for the latent volatility, and aggregating helps to reduce the idiosyncratic noise in squared returns at horizons beyond one period, thereby providing a more useful metric for comparing volatility forecasts. Awartani and Corradi (2004) and Hansen and Lunde (2006) show that the MSFE loss function produces a consistent empirical ranking of forecasting models when squared returns serve as a proxy for the latent volatility. Patton (2006) also shows that the MSFE provides a consistent ranking of models when using squared

[11] Our strategy of comparing volatility forecasts generated by GARCH(1,1) models that use an expanding window, long and short rolling windows, and a window selected by applying a structural break test to the data available at the time of forecast formation is similar in spirit to an out-of-sample forecasting exercise in Pesaran and Timmermann (2005). They consider forecasts of macro variables in the G7 countries generated by linear autoregressive models estimated with windows of different sizes.

DOI: 10.1002/jae

returns as a volatility proxy, while a number of other popular loss functions, such as absolute error, fail to provide consistent rankings.

The second loss function we consider is the González-Rivera *et al.* (2004) VaR loss function. Let $\text{VaR}^{0.05}_{t|t-s,i}$ be the forecast of the 0.05 quantile of the cumulative distribution function for the cumulative return, $\tilde{e}_t = \sum_{j=1}^{s} e_{t-(j-1)}$, generated by model i and formed at time $t-s$. We follow González-Rivera *et al.* (2004) and evaluate the forecasting models with respect to VaR using the following mean loss function:

$$\text{MVaR}_{s,i} = [P-(s-1)]^{-1} \sum_{t=R+s}^{T} (0.05 - d^{0.05}_{t,i})(\tilde{e}_t - \text{VaR}^{0.05}_{t|t-s,i}) \tag{6}$$

where $d^{0.05}_{t,i} = 1(\tilde{e}_t < \text{VaR}^{0.05}_{t|t-s,i})$ and $1(\cdot)$ is the indicator function that takes a value of unity when the argument is satisfied. This is an asymmetric loss function that is more sophisticated than simple 'hit or no hit' VaR-based loss functions. In particular, when the actual return, $\tilde{e}_t$, is less than $\text{VaR}^{0.05}_{t|t-s,i}$, the loss function attaches a large weight of 0.95 to the (absolute value of) the difference between $\tilde{e}_t$ and $\text{VaR}^{0.05}_{t|t-s,i}$, reflecting a relatively high cost associated with large losses. The loss function attaches a smaller weight of 0.05 to the difference between $\tilde{e}_t$ and $\text{VaR}^{0.05}_{t|t-s,i}$ when $\tilde{e}_t$ is greater than $\text{VaR}^{0.05}_{t|t-s,i}$. While the weight attached to the magnitude of the difference between $\tilde{e}_t$ and $\text{VaR}^{0.05}_{t|t-s,i}$ is smaller when $\tilde{e}_t$ is greater than $\text{VaR}^{0.05}_{t|t-s,i}$, it is still positive, thereby allowing the loss function to reflect the opportunity cost of the capital held to cover the potential losses indicated by $\text{VaR}^{0.05}_{t|t-s,i}$. The $\text{MVaR}_{s,i}$ criterion has the advantage that it does not require observations of the latent volatility, h_t, and is motivated by the importance of VaR as a risk management tool. We generate $\text{VaR}^{0.05}_{t|t-s,i}$ by assuming $\varepsilon_t \sim N(0,1)$, simulating a sequence of returns based on the estimates of the conditional volatility process available at the time of forecast formation ($\{e^*_{t-(j-1)}\}_{j=1}^{s}$), and computing the simulated cumulative return ($\tilde{e}^*_t = \sum_{j=1}^{s} e^*_{t-(j-1)}$). We repeat this process 2000 times, leaving us with an empirical distribution of simulated cumulative returns. $\text{VaR}^{0.05}_{t|t-s,i}$ is the 100th element of the ordered simulated cumulative returns.[12]

In addition to using the $\text{MSFE}^*_{s,i}$ and $\text{MVaR}_{s,i}$ loss functions to rank the forecasting models, we employ the White (2000) 'reality check' to test whether the expected loss of the forecasts generated by at least one of the five competing models is significantly less than the expected loss of the forecasts generated by a given benchmark model. Define the loss at time t for forecasting model j relative to benchmark model i as $f_{t,i,j} = L_{t,i} - L_{t,j}$, where $L_{t,\cdot}$ is given for each loss function by the expression after the summation operator in equation (5) or (6), and let $\overline{f}_{i,j} = [P-(s-1)]^{-1}\sum_{t=R+s}^{T} f_{t,i,j}$. The White (2000) statistic is given by

$$\overline{V}_l = \max_{k=1,\ldots,l}[P-(s-1)]^{0.5}(\overline{f}_{i,1}, \ldots, \overline{f}_{i,l}) \tag{7}$$

where l is the number of competing models ($l = 5$ in our applications). The null hypothesis is that none of the competing models has superior predictive ability in terms of expected loss over the benchmark model, whereas the one-sided (upper-tail) alternative hypothesis is that at least one of the competing models has superior predictive ability over the benchmark model. Following White

[12] In order to relax the assumption that ε_t is normally distributed, we also generate $\text{VaR}^{0.05}_{t|t-s,i}$ using a bootstrap procedure. The bootstrap procedure and results are described below in Section 3.3.

DOI: 10.1002/jae

(2000), a p-value corresponding to $\overline{V}_l$ is generated using the stationary bootstrap of Politis and Romano (1994).

We perform the White (2000) reality check with the GARCH(1,1) expanding window, RiskMetrics, and FIGARCH(1,d,1) expanding window models serving in turn as the benchmark model. The reality check allows us to test whether any of the five methods of accommodating structural breaks in the unconditional variance of exchange rate returns improves real-time volatility forecasting performance relative to a given benchmark model. The reality check helps to control for data mining when considering a multiple number of competing models. We also compute the Hansen (2005) studentized version of the $\overline{V}_l$ statistic, T_n^{SPA}, where we again generate the corresponding p-value using the stationary bootstrap of Politis and Romano (1994).[13] The Hansen (2005) version of the White (2000) reality check is designed to be a more powerful test of superior predictive ability.

A word of caution is in order with respect to the use of the White (2000) and Hansen (2005) statistics and the stationary bootstrap. Recent research shows that making inferences concerning relative predictive accuracy across forecasting models can be tricky and depends on a number of factors, such as the size of the in-sample period relative to the out-of-sample period (P/R), type of estimation window used (expanding, rolling, or fixed), and whether the models being compared are nested or non-nested.[14] We recognize that, strictly speaking, all of the conditions required for the validity of the stationary bootstrap are not necessarily satisfied in our applications, and we report bootstrapped p-values for the White (2000) $\overline{V}_l$ and Hansen (2005) T_n^{SPA} statistics as a rough guide to assessing statistical significance.

3. ESTIMATION RESULTS

3.1. Data

We use daily nominal exchange rate data from Global Financial Data to compute the daily return of the US dollar against the currencies of Canada, Denmark, Germany, Japan, Norway, Switzerland, and the UK for January 2 1980 to August 31 2005. We also consider the daily return corresponding to the US trade-weighted exchange rate for the same period. Table I reports summary statistics for the eight exchange rate return series. Heteroskedastic and autocorrelation consistent standard errors for the mean, standard deviation, skewness, and excess kurtosis are included in the table and computed as in West and Cho (1995). Results in Panel A of Table I indicate that none of the means is significantly different from zero and that Japan and Switzerland are the only currencies that exhibit significant skewness. All of the currencies display significant excess kurtosis, a well-known stylized fact of dollar exchange rate returns. The West and Cho (1995) modified Ljung-Box statistics reported in Table I are robust to conditional heteroskedasticity, and they give no significant evidence of autocorrelation in any of the exchange rate return series. With respect to the squared returns in Panel B of Table I, the Ljung-Box statistics give clear indication of serial correlation, and the Engle (1982) Lagrange multiplier statistics offer significant evidence of

[13] Hansen (2005) discusses the generation of three bootstrapped p-values (consistent, lower bound, and upper bound) corresponding to the T_n^{SPA} statistic. We report the consistent p-value in Section 3.3 below.

[14] A partial list of relevant studies includes West (1996), West and McCracken (1998), Clark and McCracken (2001, 2005), McCracken (2004), Corradi and Swanson (2007), and Giacomini and White (2006). See Corradi and Swanson (2006) for an informative review of some of these issues.

DOI: 10.1002/jae

ARCH effects. Overall, the results in Table I provide support for modeling US dollar exchange rate returns as GARCH processes, helping to explain the popularity of GARCH models in the literature.

3.2. In-Sample Test Results

We apply the modified ICSS algorithm to the eight exchange rate return series for January 2 1980 to August 31 2005.[15] Figure 1 shows the break dates identified by the modified ICSS

Table I. Summary statistics, US dollar exchange rate returns, January 2 1980 to August 31 2005

A. Exchange rate returns

	Canada	Denmark	Germany	Japan
Mean	0.000 (0.004)	0.002 (0.009)	−0.001 (0.008)	−0.012 (0.009)
Standard deviation	0.325 (0.005)	0.691 (0.014)	0.684 (0.008)	0.697 (0.012)
Skewness	−0.017 (0.090)	0.050 (0.364)	−0.115 (0.071)	−0.560 (0.197)
Excess kurtosis	2.954 (0.287)	8.687 (3.360)	1.779 (0.228)	5.583 (1.534)
Minimum	−1.997	−7.813	−4.141	−6.953
Maximum	1.917	8.130	3.490	4.154
Modified Ljung-Box ($r = 20$)	13.300 [0.864]	18.089 [0.582]	14.162 [0.822]	25.008 [0.201]

	Norway	Switzerland	UK	US (trade-weighted)
Mean	0.004 (0.008)	−0.004 (0.009)	−0.003 (0.008)	0.000 (0.007)
Standard deviation	0.630 (0.010)	0.747 (0.009)	0.639 (0.009)	0.563 (0.007)
Skewness	0.347 (0.231)	−0.145 (0.066)	−0.022 (0.115)	−0.095 (0.075)
Excess kurtosis	5.064 (2.098)	1.584 (0.210)	3.213 (0.501)	1.978 (0.278)
Minimum	−4.527	−4.408	−4.390	−3.546
Maximum	6.818	3.837	4.589	2.861
Modified Ljung-Box ($r = 20$)	18.917 [0.527]	10.318 [0.962]	15.915 [0.722]	18.741 [0.539]

B. Squared exchange rate returns

	Canada	Denmark	Germany	Japan
Ljung-Box ($r = 20$)	2405.817 [0.000]	1095.640 [0.000]	515.872 [0.000]	409.644 [0.000]
ARCH Lagrange multiplier ($q = 2$)	229.357 [0.000]	1052.623 [0.000]	58.058 [0.000]	176.947 [0.000]
ARCH Lagrange multiplier ($q = 10$)	570.912 [0.000]	1134.201 [0.000]	208.522 [0.000]	217.110 [0.000]

	Norway	Switzerland	UK	US (trade-weighted)
Ljung-Box ($r = 20$)	354.405 [0.000]	422.218 [0.000]	1312.358 [0.000]	652.178 [0.000]
ARCH Lagrange multiplier ($q = 2$)	84.320 [0.000]	48.902 [0.000]	147.154 [0.000]	73.438 [0.000]
ARCH Lagrange multiplier ($q = 10$)	160.419 [0.000]	189.723 [0.000]	389.903 [0.000]	232.752 [0.000]

Note: Returns are defined as 100 times the log-differences of daily US dollar exchange rates. Heteroskedastic and autocorrelation consistent standard errors for the mean, standard deviation, skewness, and excess kurtosis are given in parentheses. Ljung-Box statistics correspond to a test of the null hypothesis that the first r autocorrelations are zero. Modified Ljung-Box statistics are robust to conditional heteroskedasticity. ARCH Lagrange multiplier statistics correspond to a test of the null hypothesis of no ARCH effects from lag orders 1 through q. P-values are given in brackets; 0.000 indicates less than 0.0005.

[15] We implement the modified ICSS algorithm using the GAUSS procedures available from Andreu Sansó's web page at http://www.uib.es/depart/deaweb/personal/profesores/personalpages/andreusanso/we. The original code computes critical

DOI: 10.1002/jae

algorithm, as well as three-standard-deviation bands for each of the regimes defined by the structural breaks. The modified ICSS algorithm selects no structural breaks in the unconditional variance of exchange rate returns for Japan; a single structural break for Denmark, Germany, Norway, and the US; two structural breaks for Switzerland and the UK; and three structural breaks for Canada. Overall, the modified ICSS algorithm identifies one or more variance breaks—and thus breaks in the GARCH processes governing volatility—for seven of the eight exchange rate return series.[16]

A number of the variance breaks detected by the modified ICSS algorithm appear to be associated with significant economic events.[17] The single break for Germany occurs at the end of 1996 and signals a decrease in volatility. This break likely represents a decrease in uncertainty resulting from final discussions concerning the debut of the euro and modeling of the

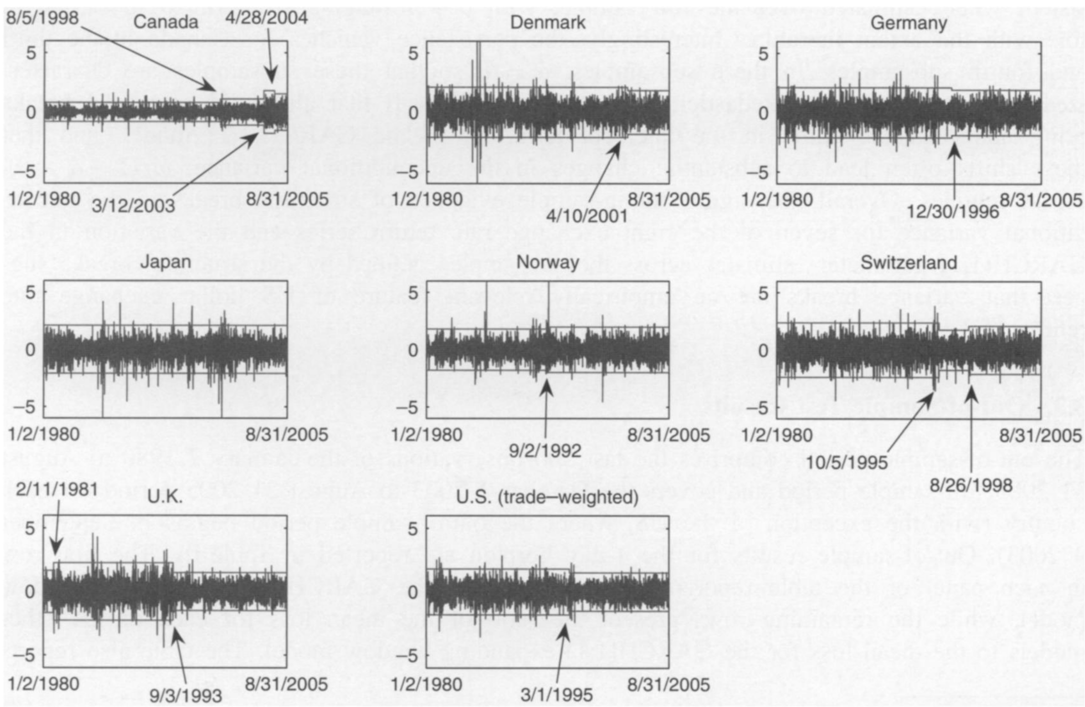

Figure 1. Exchange rate returns and three-standard-deviation bands for the regimes defined by the structural breaks identified by the modified ICSS algorithm. *Note*: Estimated break dates are indicated by the text arrows

values using a response surface that is appropriate for sample sizes up to 1000 observations. We estimate an additional response surface to generate critical values that are appropriate for samples sizes up to 7000 observations.

[16] We also find evidence of GARCH parameter non-constancy for all countries but Japan using the Lagrange multiplier test of Lundberg and Teräsvirta (2002). This procedure is designed to test for parameters that change smoothly over time, although Lundberg and Teräsvirta (2002) find that it performs well in Monte Carlo experiments in detecting discrete breaks. We implement the Lundberg and Teräsvirta (2002) test using the GAUSS procedure available from Dick van Dijk's web page at http://people.few.eur.nl/djvandijk/nltsmef/nltsmef.htm.

[17] The discussion in this paragraph is meant to be suggestive, as *a posteriori* one can probably always find some event that is relatively close to a detected structural break that could conceivably have caused the break.

DOI: 10.1002/jae

proposed European Central Bank after the German Bundesbank. The first break for Switzerland in October of 1995 signifies a decrease in volatility and corresponds to an abandonment of monetary targeting by the Swiss central bank (Rich, 2003). The second break for Switzerland in August of 1998, associated with an increase in volatility, occurs near the euro's successful debut, which challenged the leading role played by the Swiss franc as a 'safe haven' currency (Rich, 1998). The second break for the UK in September of 1993 represents a decrease in volatility that roughly corresponds to the end of the Exchange Rate Mechanism crisis in Europe.

In Table II, we report full-sample QMLE GARCH(1,1) parameter estimates for the eight exchange rate return series, as well as QMLE GARCH(1,1) parameter estimates for each of the subsamples defined by the structural breaks identified by the modified ICSS algorithm.[18] Inspection of the parameter estimates reveals that the GARCH(1,1) processes are quite persistent when estimated over the full sample, with $\hat{\alpha} + \hat{\beta}$ ranging from 0.976 to 0.998, in line with the extant literature. Interestingly, the persistence vanishes for Canada in the third and fourth subsamples. In these subsamples, $\hat{\alpha} = 0$, so that these subsamples are characterized by conditional homoskedasticity. We see from Table II that all of the structural breaks bring about sizable shifts in the intercept term, $\hat{\omega}$, of the GARCH(1,1) model, and that these shifts often lead to substantial changes in the unconditional variance, $\hat{\omega}/(1 - \hat{\alpha} - \hat{\beta})$, across regimes. Overall, the significant in-sample evidence of structural breaks in the unconditional variance for seven of the eight exchange rate return series and the variation in the GARCH(1,1) parameter estimates across the subsamples defined by the structural breaks suggest that variance breaks are an empirically relevant feature of US dollar exchange rate returns.[19]

3.3. Out-of-Sample Test Results

The out-of-sample period comprises the last 500 observations of the January 2 1980 to August 31 2005 full-sample period and covers the October 1 2003 to August 31 2005 period for each country (with the exception of the US, where the out-of-sample period begins in September 9 2003). Out-of-sample results for the 1-day horizon are reported in Table III. The first row in each panel of the table reports the mean loss for the GARCH(1,1) expanding window model, while the remaining rows present the ratio of the mean loss for each of the other models to the mean loss for the GARCH(1,1) expanding window model. The table also reports

[18] We use the GAUSS module Constrained Maximum Likelihood 2.0 to obtain the QMLE GARCH(1,1) parameter estimates. Within the module, we use the Newton-Raphson algorithm and impose the restrictions that $\omega > 0$ and $\alpha, \beta \geq 0$. We considered different starting values for the parameters, and the reported parameter estimates in Table II appear to correspond to the global maximum.

[19] Stărică (2003), Stărică and Granger (2005), and Stărică *et al.* (2005) consider modeling stock return volatility as a conditionally homoskedastic process subject to potentially frequently occurring breaks in unconditional volatility. Our results for the last two regimes for Canada provide evidence of this type of behavior, and as we discuss in Section 2.2 above, this modeling approach motivates our inclusion of the moving average model in our out-of-sample forecasting exercises. Note that we find strong evidence of ARCH effects according to the Lagrange multiplier test in nearly all of the other regimes defined by the modified ICSS algorithm and that the estimates of α and β are significant in Table II in all of the other regimes. There is thus significant evidence of GARCH dynamics for most regimes. We detect substantially more breaks (and find less evidence of GARCH dynamics) using the 'unmodified' version of the ICSS algorithm, which assumes conditional homoskedasticity under the null. However, this effectively assumes away any GARCH dynamics, and as we note in Section 2.1 above, the unmodified algorithm is plagued by size distortions if stable regimes contain GARCH dynamics.

DOI: 10.1002/jae

Table II. Quasi maximum likelihood estimation results for GARCH(1,1) models

	Canada	Denmark	Germany	Japan
A. GARCH(1,1) full sample estimation results, January 2 1980 to August 31 2005				
$\hat{\omega}$	0.001 (0.0001)	0.010 (0.002)	0.010 (0.002)	0.011 (0.002)
$\hat{\alpha}$	0.066 (0.005)	0.060 (0.006)	0.050 (0.005)	0.051 (0.006)
$\hat{\beta}$	0.932 (0.005)	0.920 (0.010)	0.929 (0.008)	0.927 (0.009)
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.242 (0.181)	0.501 (0.040)	0.479 (0.031)	0.497 (0.032)
B. GARCH(1,1) estimation results for the sub-samples defined by the structural breaks				
Subsample 1	Jan 2 1980–Aug 5 1998	Jan 2 1980–Apr 10 2001	Jan 2 1980–Dec 30 1996	Jan 2 1980–Aug 31 2005
$\hat{\omega}$	0.001 (0.0002)	0.012 (0.003)	0.012 (0.002)	0.011 (0.002)
$\hat{\alpha}$	0.079 (0.008)	0.072 (0.008)	0.068 (0.008)	0.051 (0.006)
$\hat{\beta}$	0.913 (0.008)	0.906 (0.011)	0.910 (0.010)	0.927 (0.009)
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.113 (0.036)	0.542 (0.054)	0.534 (0.053)	0.497 (0.032)
Subsample 2	Aug 6 1998–Mar 12 2003	Apr 11 2001–Aug 31 2005	Dec 31 1996–Aug 31 2005	
$\hat{\omega}$	0.004 (0.002)	0.007 (0.005)	0.005 (0.002)	
$\hat{\alpha}$	0.044 (0.011)	0.018 (0.008)	0.019 (0.005)	
$\hat{\beta}$	0.926 (0.022)	0.962 (0.016)	0.970 (0.009)	
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.130 (0.014)	0.377 (0.031)	0.404 (0.033)	
Subsample 3	Mar 13 2003–Apr 28 2004			
$\hat{\omega}$	0.362 (0.029)			
$\hat{\alpha}$	0			
$\hat{\beta}$	—			
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.362 (0.029)			
Subsample 4	Apr 29 2004–Aug 31 2005			
$\hat{\omega}$	0.264 (0.021)			
$\hat{\alpha}$	0			
$\hat{\beta}$	—			
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.264 (0.021)			

Note: Standard errors are given in parentheses.

p-values corresponding to the White (2000) $\overline{V}_l$ and Hansen (2005) T_n^{SPA} statistics with the GARCH(1,1) expanding window, RiskMetrics, and FIGARCH(1,d,1) expanding window models serving in turn as the benchmark model and the two GARCH(1,1) rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models serving as the competing models.

DOI: 10.1002/jae

Table II. (continued)

	Norway	Switzerland	UK	US (trade-weighted)
A. GARCH(1,1) full sample estimation results, January 2 1980 to August 31 2005				
$\hat{\omega}$	0.006 (0.001)	0.013 (0.003)	0.003 (0.001)	0.004 (0.001)
$\hat{\alpha}$	0.077 (0.007)	0.041 (0.005)	0.038 (0.004)	0.045 (0.005)
$\hat{\beta}$	0.911 (0.008)	0.935 (0.009)	0.955 (0.005)	0.942 (0.006)
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.494 (0.080)	0.560 (0.028)	0.413 (0.050)	0.328 (0.029)
B. GARCH(1,1) estimation results for the sub-samples defined by the structural breaks				
Subsample 1	Jan 2 1980–Sep 2 1992	Jan 2 1980–Oct 5 1995	Jan 2 1980–Feb 11 1981	Jan 2 1980–Mar 1 1995
$\hat{\omega}$	0.007 (0.002)	0.019 (0.005)	0.010 (0.009)	0.010 (0.002)
$\hat{\alpha}$	0.104 (0.011)	0.053 (0.007)	0.034 (0.021)	0.063 (0.008)
$\hat{\beta}$	0.881 (0.012)	0.917 (0.012)	0.936 (0.035)	0.913 (0.011)
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.487 (0.129)	0.643 (0.043)	0.318 (0.062)	0.373 (0.034)
Subsample 2	Sep 3 1992–Aug 31 2005	Oct 6 1995–Aug 26 1998	Feb 13 1981–Sep 3 1993	Mar 2 1995–Aug 31 2005
$\hat{\omega}$	0.005 (0.002)	0.003 (0.002)	0.012 (0.004)	0.001 (0.001)
$\hat{\alpha}$	0.049 (0.007)	0.012 (0.006)	0.056 (0.010)	0.027 (0.005)
$\hat{\beta}$	0.942 (0.009)	0.979 (0.010)	0.923 (0.014)	0.967 (0.006)
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$	0.470 (0.080)	0.338 (0.043)	0.578 (0.058)	0.256 (0.042)
Subsample 3		Aug 27 1998–Aug 31 2005	Sep 6 1993–Aug 31 2005	
$\hat{\omega}$		0.011 (0.006)	0.004 (0.001)	
$\hat{\alpha}$		0.017 (0.006)	0.025 (0.005)	
$\hat{\beta}$		0.960 (0.015)	0.960 (0.008)	
$\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$		0.478 (0.028)	0.250 (0.018)	

The results in Table III show that none of the benchmark models—the GARCH(1,1) expanding window, RiskMetrics, FIGARCH(1,d,1) expanding window—delivers the lowest mean loss for any country using the $\text{MSFE}^*_{s,i}$ loss function. Among the competing models, the two GARCH(1,1) rolling window models almost always have a lower $\text{MSFE}^*_{s,i}$ than the three benchmark models. This is the first evidence that allowing for instabilities in GARCH(1,1) models leads to out-of-sample forecasting gains. The GARCH(1,1) weighted ML model outperforms the GARCH(1,1) expanding window and RiskMetrics models in at least half the cases using the $\text{MSFE}^*_{s,i}$ criterion. The GARCH(1,1) with breaks model typically outperforms all three benchmark models for the $\text{MSFE}^*_{s,i}$ criterion, and the moving average model outperforms the GARCH(1,1) expanding window model according to the $\text{MSFE}^*_{s,i}$ metric for half of the countries. The five competing models, all of which make adjustments to accommodate potential structural breaks, typically lead to reductions in $\text{MSFE}^*_{s,i}$ of 1–4% relative to the benchmark models, and there is often significant evidence of superior predictive ability relative to the

DOI: 10.1002/jae

Table III. Out-of-sample exchange rate return volatility forecasting results, $s = 1$

Model	Canada	Denmark	Germany	Japan
A. $\text{MSFE}^*_{s,i}$				
GARCH(1,1) expanding window	0.199 [0.02] {0.00}	0.351 [0.21] {0.04}	0.344 [0.24] {0.06}	0.327 [0.38] {0.09}
RiskMetrics	0.994 [0.03] {0.06}	1.001 [0.22] {0.02}	1.005 [0.22] {0.04}	1.002 [0.39] {0.18}
FIGARCH(1,*d*,1) expanding window	1.009 [0.01] {0.00}	1.001 [0.18] {0.02}	0.994 [0.48] {0.36}	0.998 [0.54] {0.41}
GARCH(1,1) 0.50 rolling window	0.982	0.985	**0.988**	0.997
GARCH(1,1) 0.25 rolling window	0.986	**0.984**	0.990	**0.992**
GARCH(1,1) weighted ML	0.981	0.996	1.000	0.998
GARCH(1,1) with breaks	0.984	0.992	0.995	1.000
Moving average	**0.964**	0.995	1.001	1.014
B. $\text{MVaR}_{s,i}$				
GARCH(1,1) expanding window	0.060 [0.03] {0.05}	0.065 [0.39] {0.37}	0.066 [0.23] {0.17}	**0.062** [1.00] {1.00}
RiskMetrics	0.993 [0.04] {0.08}	1.011 [0.22] {0.17}	1.001 [0.26] {0.23}	1.010 [0.51] {0.48}
FIGARCH(1,*d*,1) expanding window	1.004 [0.02] {0.04}	1.007 [0.25] {0.19}	0.991 [0.49] {0.55}	1.010 [0.56] {0.53}
GARCH(1,1) 0.50 rolling window	0.980	0.992	0.986	1.007
GARCH(1,1) 0.25 rolling window	0.988	**0.988**	0.985	1.002
GARCH(1,1) weighted ML	0.979	0.999	0.992	1.011
GARCH(1,1) with breaks	0.959	0.990	**0.982**	1.000
Moving average	**0.951**	1.008	0.998	1.001

Model	Norway	Switzerland	UK	US (trade-weighted)
C. $\text{MSFE}^*_{s,i}$				
GARCH(1,1) expanding window	0.524 [0.32] {0.23}	0.595 [0.34] {0.04}	0.209 [0.73] {0.46}	0.179 [0.41] {0.27}
RiskMetrics	0.999 [0.36] {0.05}	1.012 [0.14] {0.02}	1.004 [0.54] {0.46}	1.009 [0.22] {0.05}
FIGARCH(1,*d*,1) exp. win.	0.998 [0.34] {0.31}	0.998 [0.42] {0.14}	1.006 [0.60] {0.28}	0.999 [0.45] {0.36}
GARCH(1,1) 0.50 rolling win.	0.990	**0.992**	**0.998**	**0.993**
GARCH(1,1) 0.25 rolling win.	**0.985**	0.994	1.007	0.995
GARCH(1,1) weighted ML	0.994	1.000	1.023	1.006
GARCH(1,1) with breaks	0.991	0.993	1.005	0.994
Moving average	0.992	0.999	1.054	1.003
D. $\text{MVaR}_{s,i}$				
GARCH(1,1) expanding window	0.071 [0.18] {0.07}	0.076 [0.36] {0.28}	0.058 [0.59] {0.39}	0.056 [0.35] {0.24}
RiskMetrics	1.001 [0.19] {0.18}	1.018 [0.13] {0.10}	**0.990** [1.00] {1.00}	1.015 [0.15] {0.05}
FIGARCH(1,*d*,1) exp. win.	0.990 [0.50] {0.47}	0.997 [0.46] {0.38}	1.010 [0.29] {0.03}	0.991 [0.63] {0.59}
GARCH(1,1) 0.50 rolling win.	1.002	1.000	0.992	**0.986**
GARCH(1,1) 0.25 rolling win.	**0.980**	**0.989**	1.006	0.988
GARCH(1,1) weighted ML	0.988	0.997	1.012	0.996
GARCH(1,1) with breaks	1.001	0.994	1.009	0.995
Moving average	1.007	0.990	1.074	0.998

Note: Entries for the GARCH(1,1) expanding window model give the mean loss for this model. Entries for the other models give the ratio of the mean loss for each model to the mean loss for the GARCH(1,1) expanding window model. Bold entries denote the model with the smallest mean loss among all of the models. *P*-values for the White (2000) $\overline{V}_l$ (Hansen, 2005, T_n^{SPA}) statistics are given in brackets (curly brackets) and correspond to a test of the null hypothesis that none of the five competing models (two GARCH(1,1) rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models) has a lower expected loss than the benchmark model indicated on the left against the one-sided (upper-tail) alternative hypothesis that at least one of the competing models has a lower expected loss than the benchmark model; 0.00 indicates less than 0.005.

DOI: 10.1002/jae

benchmark models according to the p-values corresponding to the Hansen (2005) T_n^{SPA} statistics.[20]

With respect to the $\text{MVaR}_{s,i}$ loss function in Table III, the competing models again often outperform the three benchmark models, and there are only two instances where any of the benchmark models produce the lowest mean loss (the GARCH(1,1) expanding window model for Japan and the RiskMetrics model for the UK). The GARCH(1,1) 0.50 and 0.25 rolling window models have a lower $\text{MVaR}_{s,i}$ than the three benchmark models for most countries. The GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models often have a lower $\text{MVaR}_{s,i}$ than the GARCH(1,1) expanding window model, and there are also a number of cases where these competing models have a lower $\text{MVaR}_{s,i}$ than the other two benchmark models.[21]

Results for horizons of 20 and 60 days are reported in Tables IV and V, respectively, while the complete results for the 120-day horizon are not reported, to conserve space.[22] Similar to the results in Table III, there are very few cases where any of the benchmark models deliver the lowest mean loss for either loss function at longer horizons. Note that compared to the results for $s = 1$ in Table III, we see more sizable reductions in mean loss relative to the benchmark models for $s = 20$ and $s = 60$ in Tables IV and V. At the 20-day horizon in Table IV, the best-performing competing models reduce the $\text{MSFE}^*_{s,i}$ ($\text{MVaR}_{s,i}$) by approximately 5–15% (2–10%) relative to the GARCH(1,1) expanding window benchmark model. At the 60-day horizon in Table V, the best-performing competing models attain mean loss reductions of approximately 8–25% (up to nearly 14%) for the $\text{MSFE}^*_{s,i}$ ($\text{MVaR}_{s,i}$) loss function. The pattern of results for the 120-day horizon is similar to that for the 60-day horizon, and following the trend in Tables III–V the best-performing competing models reduce the mean loss up to a very sizable 38% (17%) relative to the GARCH(1,1) expanding window benchmark model for the $\text{MSFE}^*_{s,i}$ ($\text{MVaR}_{s,i}$) loss function at the 120-day horizon.[23]

Table VI records the number of times each model has the lowest mean loss and provides summary statistics for the loss function ratios at horizons of 1, 20, 60, and 120 days. For the $\text{MSFE}^*_{s,i}$ loss function in Panel A, the performance of the GARCH(1,1) 0.50 rolling window model stands out. The 0.50 (0.25) rolling window model has the lowest $\text{MSFE}^*_{s,i}$ in 17 (6) cases, and the reduction in $\text{MSFE}^*_{s,i}$ relative to the GARCH(1,1) expanding window model averages 9%

[20] There is more evidence of superior predictive ability according to the Hansen (2005) T_n^{SPA} statistic than the White (2000) $\overline{V}_l$ statistic. This is not surprising, as the T_n^{SPA} statistic is designed to be more powerful than the $\overline{V}_l$ statistic.

[21] The empirical coverage frequencies for the 1-day-ahead 5% VaR forecasts are typically very close to 5% for each model, and the null hypothesis of correct (unconditional) coverage is only rejected at the 5% significance level using the Christoffersen (1998) likelihood-ratio statistic for the GARCH(1,1) 0.25 rolling window and GARCH(1,1) with breaks models for the USA; in these cases, the empirical frequencies are still relatively near 5% (7.6% and 7.0%, respectively). In order to conserve space, the complete results for the empirical coverage frequencies are not reported. They are available in Table A1 of a not-for-publication appendix of additional results available at http://pages.slu.edu/faculty/rapachde.

[22] The complete results for the 120-day horizon are reported in Table A2 of the not-for-publication appendix.

[23] Stărică *et al.* (2005) emphasize that as the forecast horizon increases, the volatility forecast for the GARCH(1,1) model approaches the unconditional variance of the estimated GARCH(1,1) model. By failing to allow for breaks in the unconditional variance, the forecasting performance of the GARCH(1,1) expanding window model can thus suffer at longer horizons relative to a model that allows for a changing unconditional variance. The empirical coverage frequencies for the 5% VaR forecasts at horizons of 20, 60, and 120 days are reported in Table A1 of the not-for-publication appendix. The empirical frequencies are typically close to 5% at the 20- and 60-day horizons, while a number of the empirical frequencies are zero at the 120-day horizon. This latter result is likely due to the reduced number of out-of-sample observations available at the 120-day horizon. Note that the $\text{MVaR}_{s,i}$ loss function is still informative when the empirical frequencies are zero, as it takes into account the opportunity cost of the capital held to cover the potential losses predicted by the 5% VaR forecast (recall the discussion of the $\text{MVaR}_{s,i}$ loss function in Section 2.2 above).

DOI: 10.1002/jae

Table IV. Out-of-sample exchange rate return volatility forecasting results, $s = 20$

Model	Canada	Denmark	Germany	Japan
A. $\mathrm{MSFE}^*_{s,i}$				
GARCH(1,1) expanding window	4.035 [0.29] {0.31}	10.712 [0.23] {0.02}	9.777 [0.31] {0.03}	10.454 [0.40] {0.06}
RiskMetrics	0.976 [0.38] {0.49}	1.050 [0.20] {0.01}	1.087 [0.24] {0.04}	1.034 [0.48] {0.25}
FIGARCH(1,*d*,1) exp. win.	0.933 [0.54] {0.65}	0.955 [0.31] {0.06}	0.976 [0.37] {0.04}	0.947 [0.76] {0.55}
GARCH(1,1) 0.50 rolling win.	0.945	**0.840**	**0.881**	0.976
GARCH(1,1) 0.25 rolling win.	0.966	0.865	0.923	**0.893**
GARCH(1,1) weighted ML	0.942	0.967	1.010	1.011
GARCH(1,1) with breaks	1.004	1.078	1.062	1.000
Moving average	**0.853**	1.040	1.128	1.222
B. $\mathrm{MVaR}_{s,i}$				
GARCH(1,1) expanding window	0.218 [0.14] {0.00}	0.246 [0.12] {0.01}	0.244 [0.24] {0.00}	0.252 [0.40] {0.01}
RiskMetrics	0.994 [0.11] {0.00}	1.008 [0.19] {0.12}	1.025 [0.14] {0.09}	1.020 [0.31] {0.30}
FIGARCH(1,*d*,1) exp. win.	0.969 [0.19] {0.08}	0.982 [0.45] {0.41}	0.998 [0.19] {0.19}	0.982 [0.71] {0.79}
GARCH(1,1) 0.50 rolling win.	0.967	0.974	0.976	0.993
GARCH(1,1) 0.25 rolling win.	0.953	0.977	0.987	0.977
GARCH(1,1) weighted ML	0.960	**0.969**	0.985	0.985
GARCH(1,1) with breaks	0.940	0.985	**0.975**	0.996
Moving average	**0.903**	0.991	0.993	**0.967**

Model	Norway	Switzerland	UK	US (trade-weighted)
C. $\mathrm{MSFE}^*_{s,i}$				
GARCH(1,1) expanding window	20.940 [0.17] {0.00}	15.876 [0.30] {0.01}	7.710 [0.60] {0.49}	4.869 [0.56] {0.49}
RiskMetrics	1.055 [0.14] {0.01}	1.199 [0.13] {0.04}	1.060 [0.49] {0.25}	1.185 [0.13] {0.01}
FIGARCH(1,*d*,1) exp. win.	0.973 [0.17] {0.20}	0.977 [0.38] {0.00}	1.023 [0.63] {0.25}	0.999 [0.59] {0.50}
GARCH(1,1) 0.50 rolling win.	0.905	**0.894**	**0.960**	0.961
GARCH(1,1) 0.25 rolling win.	**0.831**	0.965	1.099	0.999
GARCH(1,1) weighted ML	0.944	1.053	1.276	1.090
GARCH(1,1) with breaks	0.913	0.942	1.071	**0.952**
Moving average	0.851	1.079	1.589	1.128
D. $\mathrm{MVaR}_{s,i}$				
GARCH(1,1) expanding window	0.316 [0.17] {0.12}	0.298 [0.14] {0.02}	0.266 [0.45] {0.34}	0.224 [0.52] {0.44}
RiskMetrics	1.000 [0.20] {0.19}	1.071 [0.11] {0.07}	1.058 [0.13] {0.06}	0.992 [0.29] {0.06}
FIGARCH(1,*d*,1) exp. win.	0.962 [0.45] {0.53}	0.991 [0.22] {0.18}	0.983 [0.85] {0.85}	0.993 [0.85] {0.82}
GARCH(1,1) 0.50 rolling win.	0.978	0.990	0.997	**0.991**
GARCH(1,1) 0.25 rolling win.	0.955	0.973	0.992	1.018
GARCH(1,1) weighted ML	0.937	1.003	1.032	1.003
GARCH(1,1) with breaks	0.976	0.974	**0.979**	1.003
Moving average	**0.935**	**0.960**	0.995	1.028

Note: Entries for the GARCH(1,1) expanding window model give the mean loss for this model. Entries for the other models give the ratio of the mean loss for each model to the mean loss for the GARCH(1,1) expanding window model. Bold entries denote the model with the smallest mean loss among all of the models. *P*-values for the White (2000) $\overline{V}_l$ (Hansen, 2005, T_n^{SPA}) statistics are given in brackets (curly brackets) and correspond to a test of the null hypothesis that none of the five competing models (two GARCH(1,1) rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models) has a lower expected loss than the benchmark model indicated on the left against the one-sided (upper-tail) alternative hypothesis that at least one of the competing models has a lower expected loss than the benchmark model; 0.00 indicates less than 0.005.

(5%). The two rolling window models have the two lowest average ratios in Panel A. The rolling window models also have two of the lowest standard deviations for the ratios in Panel A, indicating

DOI: 10.1002/jae

Table V. Out-of-sample exchange rate return volatility forecasting results, $s = 60$

Model	Canada	Denmark	Germany	Japan
A. $\text{MSFE}^*_{s,i}$				
GARCH(1,1) exp. win.	28.118 [0.44] {0.46}	65.167 [0.20] {0.21}	60.037 [0.32] {0.28}	100.764 [0.35] {0.16}
RiskMetrics	1.035 [0.31] {0.37}	1.155 [0.18] {0.02}	1.195 [0.20] {0.02}	1.169 [0.18] {0.08}
FIGARCH(1,*d*,1) exp. win.	0.804 [0.94] {0.66}	0.936 [0.29] {0.39}	0.968 [0.39] {0.31}	**0.767** [1.00] {1.00}
GARCH(1,1) 0.50 rolling win.	1.045	**0.727**	**0.790**	0.987
GARCH(1,1) 0.25 rolling win.	1.081	0.805	0.897	0.820
GARCH(1,1) weighted ML	1.061	1.003	1.080	0.806
GARCH(1,1) with breaks	0.953	1.169	1.059	1.000
Moving average	**0.766**	1.144	1.273	0.869
B. $\text{MVaR}_{s,i}$				
GARCH(1,1) exp. win.	0.465 [0.17] {0.41}	0.469 [0.13] {0.13}	0.466 [0.03] {0.04}	0.431 [0.00] {0.00}
RiskMetrics	0.992 [0.14] {0.26}	1.082 [0.16] {0.12}	1.086 [0.13] {0.16}	0.968 [0.14] {0.13}
FIGARCH(1,*d*,1) exp. win.	0.975 [0.22] {0.51}	0.987 [0.25] {0.31}	0.995 [0.08] {0.10}	0.948 [0.00] {0.00}
GARCH(1,1) 0.50 rolling win.	0.999	1.024	1.019	0.973
GARCH(1,1) 0.25 rolling win.	0.990	1.001	0.997	0.952
GARCH(1,1) weighted ML	0.978	0.973	0.971	0.881
GARCH(1,1) with breaks	**0.871**	**0.929**	**0.930**	0.999
Moving average	0.888	0.964	0.980	**0.867**

Model	Norway	Switzerland	UK	US (trade-weighted)
C. $\text{MSFE}^*_{s,i}$				
GARCH(1,1) exp. win.	116.263 [0.23] {0.01}	70.668 [0.42] {0.08}	69.568 [0.56] {0.50}	25.286 [0.74] {0.69}
RiskMetrics	1.351 [0.11] {0.01}	1.358 [0.05] {0.00}	1.254 [0.16] {0.04}	1.605 [0.00] {0.00}
FIGARCH(1,*d*,1) exp. win.	0.983 [0.23] {0.08}	0.974 [0.42] {0.03}	0.961 [0.71] {0.52}	1.028 [0.75] {0.54}
GARCH(1,1) 0.50 rolling win.	0.884	**0.764**	**0.925**	1.030
GARCH(1,1) 0.25 rolling win.	**0.763**	0.993	1.074	1.105
GARCH(1,1) weighted ML	0.884	1.265	1.375	1.345
GARCH(1,1) with breaks	0.910	0.956	1.029	**0.995**
Moving average	0.850	1.369	1.440	1.360
D. $\text{MVaR}_{s,i}$				
GARCH(1,1) exp. win.	0.546 [0.19] {0.13}	0.509 [0.09] {0.24}	0.475 [0.40] {0.47}	0.446 [0.68] {0.73}
RiskMetrics	1.016 [0.14] {0.16}	1.122 [0.12] {0.10}	1.035 [0.17] {0.03}	1.084 [0.30] {0.21}
FIGARCH(1,*d*,1) exp. win.	0.955 [0.24] {0.14}	0.981 [0.23] {0.45}	0.985 [0.61] {0.70}	0.998 [0.86] {0.88}
GARCH(1,1) 0.50 rolling win.	1.004	1.008	0.996	1.036
GARCH(1,1) 0.25 rolling win.	0.964	0.993	**0.971**	1.046
GARCH(1,1) weighted ML	0.912	0.966	1.000	1.007
GARCH(1,1) with breaks	0.990	0.981	0.979	1.028
Moving average	**0.876**	**0.933**	0.999	**0.994**

Note: Entries for the GARCH(1,1) expanding window model give the mean loss for this model. Entries for the other models give the ratio of the mean loss for each model to the mean loss for the GARCH(1,1) expanding window model. Bold entries denote the model with the smallest mean loss among all of the models. *P*-values for the White (2000) $\overline{V}_l$ (Hansen, 2005, T_n^{SPA}) statistics are given in brackets (curly brackets) and correspond to a test of the null hypothesis that none of the five competing models (two GARCH(1,1) rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models) has a lower expected loss than the benchmark model indicated on the left against the one-sided (upper-tail) alternative hypothesis that at least one of the competing models has a lower expected loss than the benchmark model; 0.00 indicates less than 0.005.

that these models perform consistently well. The weighted ML model never achieves the lowest $\text{MSFE}^*_{s,i}$, has an average ratio above unity, and has a relatively large standard deviation for its ratios. The GARCH(1,1) with breaks model produces the lowest $\text{MSFE}^*_{s,i}$ in two cases and has

DOI: 10.1002/jae

an average ratio close to unity. The moving average model has the lowest $\mathrm{MSFE}^*_{s,i}$ in four cases (these are all for Canada), while its average ratio is well above unity.[24] The last three columns show that the performances of the GARCH(1,1) weighted ML and moving average models are inconsistent: sometimes the models outperform the GARCH(1,1) expanding window benchmark model by a sizable margin, but at other times they do substantially worse than the benchmark. Overall, one of the five competing models delivers the lowest $\mathrm{MSFE}^*_{s,i}$ in 29 of 32 cases in Panel A of Table VI.[25]

It is interesting that the rolling window models often perform the best among the competing models that accommodate breaks in Panel A of Table VI, as one might suspect that selecting the estimation window by explicitly testing for the most recent break, as in the GARCH(1,1) with breaks model, would perform better. Recent research by Clark and McCracken (2004) and Pesaran and Timmermann (2007) helps to explain the relatively good performance of the rolling window models. Suppose we wish to forecast a variable, and we know that a structural break has occurred at a specific date in the recent past. At first blush, it may seem optimal to use only post-break data when estimating the forecasting model in order to avoid biased model parameter estimates and forecasts. However, as noted by Clark and McCracken (2004: 1), 'there is a balance between using too much or too little data to estimate model parameters'. While discarding data from the pre-break period helps to reduce biases, it also leads to greater variances in the estimates of the forecasting

Table VI. Summary statistics for the mean loss ratios

Model	# best	Mean	Median	SD	Min.	Max.
A. $\mathrm{MSFE}^*_{s,i}$						
GARCH(1,1) expanding window	1					
RiskMetrics	0	1.221	1.121	0.287	0.976	2.137
FIGARCH(1,*d*,1) expanding window	2	0.958	0.977	0.130	0.562	1.333
GARCH(1,1) 0.50 rolling window	17	0.914	0.961	0.124	0.552	1.090
GARCH(1,1) 0.25 rolling window	6	0.946	0.975	0.115	0.680	1.200
GARCH(1,1) weighted ML	0	1.093	1.009	0.244	0.629	1.778
GARCH(1,1) with breaks	2	0.992	0.995	0.072	0.817	1.169
Moving average	4	1.120	1.067	0.276	0.624	1.781
B. $\mathrm{MVaR}_{s,i}$						
GARCH(1,1) expanding window	1					
RiskMetrics	2	1.006	1.005	0.053	0.862	1.122
FIGARCH(1,*d*,1) expanding window	0	0.985	0.989	0.023	0.934	1.063
GARCH(1,1) 0.50 rolling window	3	0.986	0.992	0.026	0.917	1.036
GARCH(1,1) 0.25 rolling window	8	0.975	0.980	0.031	0.876	1.046
GARCH(1,1) weighted ML	1	0.973	0.985	0.047	0.842	1.056
GARCH(1,1) with breaks	6	0.967	0.982	0.047	0.796	1.028
Moving average	11	0.967	0.991	0.062	0.795	1.081

Note: '# best' is the number of times that the model has the lowest mean loss.

[24] The in-sample results provide insight into why the moving average model performs the best for Canada. The results in Table II show that the last two regimes, which are contained in the out-of-sample period, are characterized by conditional homoskedasticity. As put forth by Stărică (2003) and Stărică *et al.* (2005), the moving-average model is a convenient way of capturing a conditionally homoskedastic process with relatively frequent breaks.

[25] The FIGARCH(1,*d*,1) model demonstrates the best overall performance among the benchmark models in Panel A of Table VI. As noted in Section 2.2 above, Diebold and Inoue (2001) posit that long-memory models could be useful for prediction purposes in the presence of structural breaks.

DOI: 10.1002/jae

model's parameters, thereby creating a bias-efficiency tradeoff between using pre- and post-break data.

Clark and McCracken (2004) and Pesaran and Timmermann (2007) derive the optimal least squares estimation window size for a linear regression forecasting model for the conditional mean that minimizes MSFE. Importantly, they show that it can be optimal to include pre-break data in the estimation window, and that the optimal amount of pre-break data to include depends on a number of factors, such as the sizes and directions of the changes in the parameters of the model and the exact timing of the structural break. Unfortunately, given that research on the optimal estimation window size for forecasting models is in its early stages, results are not available for GARCH processes, limiting what we can conclude with respect to the optimal window size for GARCH(1,1) forecasting models of exchange rate return volatility. Nevertheless, recent research provides intuition as to why the GARCH(1,1) rolling window models—which can frequently include data prior to the most recent break when estimating the GARCH(1,1) forecasting model—often display the best performance for the $\text{MSFE}^*_{s,i}$ loss function in our out-of-sample forecasting exercises.[26]

Panel B of Table VI presents summary statistics for the $\text{MVaR}_{s,i}$ loss function. The mean and median ratios are below unity for all of the competing models. The 0.50 (0.25) rolling window model delivers the lowest $\text{MVaR}_{s,i}$ in 3 (8) of 32 cases, so that the 0.50 rolling window model produces the lowest mean loss far less frequently than in Panel A. The number of times that the GARCH(1,1) with breaks and moving average models perform the best increases from Panel A to Panel B, and these two models have the lowest average ratios in Panel B. It is interesting to observe that methods accommodating more frequent breaks, such as the moving average, GARCH(1,1) with breaks, and GARCH(1,1) 0.25 rolling window models, perform relatively better for the $\text{MVaR}_{s,i}$ than the $\text{MSFE}^*_{s,i}$ loss function.

We briefly describe the results of several robustness checks of the out-of-sample forecasting results.[27] We employ a bootstrap procedure to forecast the 5% quantile used to compute the $\text{MVaR}_{s,i}$ loss function for the different forecasting models at the 60-day horizon. The bootstrap procedure allows us to relax the assumption that the standardized residuals, ε_t, have a standard normal distribution, thereby allowing for, among other things, leptokurtosis in the standardized residuals; see, for example, Andersen *et al.* (2006). More specifically, at the time of forecast formation, we resample (with replacement) from the 500 most recent estimated standardized residuals from the model (instead of sampling from a standard normal distribution) in order to simulate return series to generate the empirical distribution used to estimate the 5% quantile. Using this bootstrap procedure, we obtain results that are very similar with respect to relative forecasting performance to those reported in Table V for the $\text{MVaR}_{s,i}$ loss function.

We also consider forecasts generated by the asymmetric GJR-GARCH(1,1) model of Glosten *et al.* (1993) estimated using an expanding window. This model almost always has a mean loss that is higher (often considerably so) than the GARCH(1,1) expanding window benchmark model.[28] In addition, we consider forecasts generated by a Markov-switching GARCH(1,1) model estimated

[26] The rolling window models may also be better able to capture a smoothly evolving data-generating volatility process than the GARCH(1,1) with breaks model, as the latter model explicitly treats breaks as abrupt.

[27] In order to conserve space, the complete results are not reported. The complete results for the robustness checks discussed in the next three paragraphs are reported in Tables A3–A5 of the not-for-publication appendix.

[28] This is in line with the results in Hansen and Lunde (2005). With respect to forecasting return volatility for the German mark-US dollar exchange rate, they find that forecasts generated by a large number of GARCH model specifications,

DOI: 10.1002/jae

using an expanding window, in light of Klaassen (2002) and Haas *et al.* (2004). Instead of viewing structural breaks as deterministic, the Markov-switching GARCH(1,1) model treats breaks as governed by a two-state Markov process. The Markov-switching GARCH(1,1) model does not consistently outperform the GARCH(1,1) expanding window benchmark model, and it has a mean loss that is often much higher than the GARCH(1,1) expanding window model at longer horizons.[29]

Finally, we consider 60-day-ahead forecasts for an earlier out-of-sample period comprising the 500 observations prior to the out-of-sample period used in Tables III–V. Overall, the results again show that there are typically forecasting gains associated with accommodating structural breaks in the unconditional volatility of exchange rate returns.

3.4. Combination Forecasts

Tables III–VI show that the best-performing forecasting model can vary across countries, forecast horizons, and loss functions, making it difficult to identify a priori the 'best' individual forecasting model in a given situation in the presence of volatility breaks. As discussed above, even if one somehow knows that a structural break has occurred in the recent past, selecting the optimal estimation window size for the forecasting model can be quite complicated, as it depends on the exact size, timing, and direction of the break. In practice, as emphasized by Pesaran and Timmermann (2007), there is thus substantial uncertainty surrounding the optimal estimation window for GARCH(1,1) forecasting models. In addition, there can be situations where the moving average model, which ignores any GARCH dynamics and allows for a frequently changing unconditional volatility, is a useful forecasting model.

Recent research suggests that combination forecasts provide a practical way of dealing with this uncertainty. Clark and McCracken (2004) and Pesaran and Timmermann (2007) consider combining the forecasts of individual models that use estimation windows of various sizes. Using Monte Carlo simulations based on linear regression models, they show that combination forecasts can work well in the presence of structural breaks in comparison to forecasting models that ignore potential structural breaks by relying on an expanding estimation window. We consider combining the forecasts generated by the GARCH(1,1) expanding window model and the five competing models (GARCH(1,1) 0.50 rolling window, GARCH(1,1) 0.25 rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models).[30] We use two simple methods for combining the forecasts generated by the GARCH(1,1) expanding window and five

including the GJR-GARCH, are not superior to forecasts generated by a GARCH(1,1) model. Given the focus of Hansen and Lunde (2005), they do not consider methods specifically designed to accommodate structural breaks in volatility (as in the present paper).

[29] Klaassen (2002), with respect to exchange rate return volatility, and Marcucci (2005), with respect to US stock return volatility, find some forecasting gains associated with the Markov-switching GARCH model at shorter horizons. They do not consider longer horizons, as in the present paper, and we find that the forecasting performance of the Markov-switching GARCH model deteriorates significantly at longer horizons. Klaassen (2002) considers horizons of 1 and 10 days; Marcucci (2005) considers horizons of 1–22 days and finds that the Markov-switching GARCH model offers forecasting gains at horizons of 1 and 5 days.

[30] Subject to a minimum window length requirement, Pesaran and Timmermann (2007) consider combining the forecasts from individual models estimated using all possible window sizes. While this is feasible for linear regression models and samples of moderate size, it is prohibitively expensive in our applications, as the computation of the combination forecast at each forecast date would entail the estimation of thousands of GARCH(1,1) models. Nevertheless, it would be interesting in future research to explore combination forecasts based on models estimated using a greater variety of window sizes.

DOI: 10.1002/jae

other models. The first combining method simply takes the mean of the six individual forecasts, while the second combining method uses the trimmed mean (the mean of the four individual forecasts remaining after discarding the highest and lowest individual forecasts). The forecast combination literature finds that simple combining methods, such as the mean and trimmed mean, often perform well in out-of-sample forecasting exercises; see, for example, Stock and Watson (2003).

Table VII reports the ratios of the mean loss for the mean and trimmed mean combination forecasts—denoted by $\text{MSFE}^*_{s,m}$ and $\text{MVaR}_{s,m}$ ($\text{MSFE}^*_{s,tm}$ and $\text{MVaR}_{s,tm}$) for the mean (trimmed mean) forecasts—to the mean loss for the GARCH(1,1) expanding window model, which again serves as a benchmark that is appropriate in the absence of structural breaks. From Panel A of Table VII, we see that the $\text{MSFE}^*_{s,m}$ for the mean combination forecast is almost always less than or equal to the $\text{MSFE}^*_{s,i}$ for the GARCH(1,1) expanding window benchmark model. The reduction in mean loss is often sizable (up to 36%) and tends to increase as the forecast horizon increases. We see from Panel B of Table VII that the $\text{MVaR}_{s,m}$ ratios are again almost all less than unity. This further indicates that there are often forecasting gains associated with the mean combination forecasts relative to the benchmark model forecasts. The reduction in mean loss in Panel B for the mean combination forecast typically ranges from 2% to 5% (reaching up to 11%) and there is a slight tendency for the reduction to increase with the forecast horizon. Results for the trimmed mean combination forecasts and the two loss functions are given in Panels C and D of Table VII, and the results are quite similar to those in Panels A and B. Overall, Table VII shows that the combination forecasts perform consistently well and help to avoid the inconsistencies in forecasting performance that can plague individual models. From a practical standpoint, the results in Table VII suggest that taking the average of the forecasts generated by a GARCH(1,1) model estimated using an expanding window and models that accommodate structural breaks in various ways provides a reasonably dependable method for forecasting exchange rate volatility in the presence of structural breaks.[31]

4. CONCLUSION

Our results reveal that structural breaks are an empirically relevant phenomenon for GARCH models of US dollar exchange rate return volatility. In-sample application of the modified ICSS algorithm uncovers significant evidence of structural breaks in the unconditional variance of seven of eight daily US dollar exchange rate return series. This implies structural breaks in the GARCH(1,1) processes governing the conditional volatility for these exchange rates, and the GARCH(1,1) parameter estimates often vary significantly across the subsamples defined by the structural breaks. We also find that accommodating structural breaks in the unconditional variance of exchange rate returns often improves out-of-sample forecasts of exchange rate return volatility. Finally, from a practical standpoint, combining forecasts across models that accommodate structural breaks in various ways appears to offer a reasonably

[31] It would be interesting in future research to investigate the usefulness of Bayesian methods to select the optimal window size and/or combining weights for GARCH(1,1) forecasting models. See Clark and McCracken (2004) and Pesaran *et al.* (2006) on the use of Bayesian methods to construct forecasts in the presence of potential structural breaks in linear regression models.

DOI: 10.1002/jae

Table VII. Out-of-sample exchange rate return volatility forecasting results, mean and trimmed mean combination forecasts

	Canada	Denmark	Germany	Japan	Norway	Switzerland	UK	US (trade-weighted)
A. $\mathrm{MSFE}^*_{s,m}$								
$h = 1$	0.978	0.987	0.944	0.991	0.985	0.993	1.004	0.995
$h = 20$	0.836	0.905	0.939	0.912	0.838	0.943	1.064	0.965
$h = 60$	0.782	0.891	0.941	0.791	0.775	0.950	1.024	1.038
$h = 120$	0.684	0.807	0.890	0.644	0.853	1.016	0.912	1.172
B. $\mathrm{MVaR}_{s,m}$								
$h = 1$	0.972	0.994	0.987	0.999	0.990	0.991	1.006	0.987
$h = 20$	0.941	0.975	0.979	0.979	0.952	0.977	0.993	0.998
$h = 60$	0.948	0.974	0.972	0.943	0.951	0.969	0.989	1.010
$h = 120$	0.889	0.950	0.972	0.934	1.010	0.974	0.931	0.966
C. $\mathrm{MSFE}^*_{s,tm}$								
$h = 1$	0.979	0.987	0.991	0.994	0.988	0.994	1.004	0.995
$h = 20$	0.843	0.917	0.956	0.926	0.868	0.932	1.067	0.966
$h = 60$	0.786	0.906	0.951	0.831	0.808	0.912	1.044	1.045
$h = 120$	0.678	0.830	0.896	0.714	0.858	0.946	0.938	1.200
D. $\mathrm{MVaR}_{s,tm}$								
$h = 1$	0.973	0.995	0.986	0.998	0.993	0.994	1.002	0.990
$h = 20$	0.945	0.975	0.979	0.982	0.960	0.983	0.993	1.001
$h = 60$	0.954	0.973	0.974	0.949	0.960	0.976	0.984	1.020
$h = 120$	0.892	0.950	0.972	0.941	1.008	0.969	0.924	0.964

Note: Entries report the ratio of the mean loss for the mean and trimmed mean combination forecasts to the mean loss for the GARCH(1,1) expanding window model. $\mathrm{MSFE}^*_{s,m}$ and $\mathrm{MVaR}_{s,m}$ correspond to the mean combination forecasts based on the average of the individual forecasts generated by the GARCH(1,1) expanding window, two GARCH(1,1) rolling window, GARCH(1,1) weighted ML, GARCH(1,1) with breaks, and moving average models. $\mathrm{MSFE}^*_{s,tm}$ and $\mathrm{MVaR}_{s,tm}$ correspond to the trimmed mean combination forecasts that omit the highest and lowest forecasts before taking the average.

reliable method for generating more accurate real-time exchange rate return volatility forecasts.

ACKNOWLEDGEMENTS

The results reported in this paper were generated using GAUSS 6.0. The GAUSS programs are available at http://pages.slu.edu/faculty/rapachde/. Grant sponsor was the John Cook School of Business at Saint Louis University. We thank three referees and a Co-Editor (Tim Bollerslev) for very helpful comments. We also thank seminar participants at the 2005 Midwest Econometrics Group meetings and 2006 Society for Nonlinear Dynamics and Econometrics conference, especially Richard Baillie, Eric Hillebrand, and James Morley, as well as Mike McCracken, Chris Neely, and Mark Wohar, for insightful comments on earlier drafts. The usual disclaimer applies.

REFERENCES

Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885–905.

Andersen TG, Bollerslev T, Christoffersen PF, Diebold FX. 2006. Volatility and correlation forecasting. In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermann A (eds). Elsevier: Amsterdam; 778–878.

Andreou E, Ghysels E. 2002. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* **17**: 579–600.

Awartani BMA, Corradi V. 2004. Predicting the volatility of the S&P 500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting* **21**: 167–183.

Baillie RT, Bollerslev T. 1989. The message in daily exchange rates: a conditional-variance tale. *Journal of Business and Economic Statistics* **7**: 297–305.

Baillie RT, Bollerslev T. 1991. Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies* **58**: 565–585.

Baillie RT, Bollerslev T, Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3–30.

Berkes I, Horváth L, Kokoszka P. 2003. GARCH processes: structure and estimation. *Bernoulli* **9**: 201–227.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Bollerslev T, Engle RF. 1993. Common persistence in conditional variances. *Econometrica* **61**: 167–186.

Christoffersen PF. 1998. Evaluating interval forecasts. *International Economic Review* **39**: 841–862.

Clark P, Tamirisa N, Wei S-J with Sadikov A, Zeng L. 2004. A new look at exchange rate volatility and trade flows. International Monetary Fund Occasional Paper No. 235.

Clark TE, McCracken MW. 2001. Tests of forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**: 85–110.

Clark TE, McCracken MW. 2004. Improving forecast accuracy by combining recursive and rolling forecasts. Federal Reserve Bank of Kansas City Research Working Paper 04-10.

Clark TE, McCracken MW. 2005. Evaluating direct multistep forecasts. *Econometric Reviews* **24**: 369–404.

Corradi V, Swanson NR. 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review* **48**: 67–109.

Corradi V, Swanson NR. 2006. Predictive density evaluation. In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermann A (eds). Elsevier: Amsterdam; 197–284.

de Pooter M, van Dijk D. 2004. Testing for changes in volatility in heteroskedastic time series: a further examination. Econometric Institute Research Report 2004-38/A, Erasmus School of Economics, Rotterdam.

Diebold FX. 1986. Modelling the persistence of conditional variance: a comment. *Econometric Reviews* **5**: 51–56.

Diebold FX, Inoue A. 2001. Long memory and regime switching. *Journal of Econometrics* **105**: 131–159.

Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of U.K. inflation. *Econometrica* **50**: 987–1008.

Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variance. *Econometric Reviews* **5**: 1–50.

Franses PH, van Dijk D. 2000. *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press: Cambridge, UK.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.

González-Rivera G, Lee T-H, Mishra S. 2004. Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* **20**: 629–645.

Haas M, Mittnik S, Paolella MS. 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* **2**: 493–530.

Hansen PR. 2005. A test for superior predictive ability. *Journal of Business and Economic Statistics* **23**: 365–380.

Hansen PR, Lunde A. 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* **20**: 873–889.

Hansen PR, Lunde A. 2006. Consistent ranking of volatility models. *Journal of Econometrics* **131**: 97–121.

Hendry DF. 1986. An excursion into conditional variance land. *Econometric Reviews* **5**: 63–69.

Hillebrand E. 2005. Neglecting parameter changes in GARCH models. *Journal of Econometrics* **129**: 121–138.

Hwang S, Valls Pereira PS. 2006. Small sample properties of GARCH estimates and persistence. *European Journal of Finance* **12**: 473–494.
Inclán C, Tiao GC. 1994. Use of cumulative sums of squares for retrospective detection of changes in variance. *Journal of the American Statistic Association* **89**: 913–923.
Jensen ST, Rahbek A. 2004. Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**: 1203–1226.
Klaassen F. 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* **27**: 363–394.
Kokoszka P, Leipus R. 1999. Testing for parameter changes in ARCH models. *Lithuanian Mathematic Journal* **39**: 182–195.
Lamoureux CG, Lastrapes WD. 1990. Persistence in variance, structural change and the GARCH model. *Journal of Business and Economic Statistics* **8**: 225–234.
Lee S, Park S. 2001. The CUSUM of squares test for scaled changes in infinite order moving average processes. *Scandinavian Journal of Statistics* **28**: 625–644.
Lundberg S, Teräsvirta T. 2002. Evaluating GARCH models. *Journal of Econometrics* **110**: 417–435.
Marcucci J. 2005. Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics* **9**: Article 6.
McCracken MW. 2004. Asymptotics for out of sample tests of Granger causality. Working paper, University of Missouri at Columbia.
Mikosch T, Stărică C. 2003. Long range dependence effects and ARCH modeling. In *Theory and Applications of Long Range Dependence*, Doukham P, Oppenheim G, Taqqu M (eds). Birkhäuser: Boston, MA; 439–460.
Mikosch T, Stărică C. 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* **86**: 378–390.
Mittnik S, Paolella MS. 2000. Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting* **19**: 313–333.
Neely C. 1999. Target zones and conditional volatility: the role of realignments. *Journal of Empirical Finance* **6**: 177–192.
Newey WK, West KD. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**: 631–654.
Patton AJ. 2006. Volatility forecast evaluation and comparison using imperfect volatility proxies. Working paper, London School of Economics.
Perron P, Qu Z. 2006. An analytical evaluation of the log-periodogram estimate in the presence of level shifts and its implications for stock return volatility. Working paper, Boston University.
Pesaran MH, Timmermann A. 2005. Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* **129**: 183–217.
Pesaran MH, Timmermann A. 2007. Selection of estimation window in the presence of structural breaks. *Journal of Econometrics* **137**: 134–161.
Pesaran MH, Pettenuzzo D, Timmermann A. 2006. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* **73**: 1057–1084.
Politis D, Romano J. 1994. The stationary bootstrap. *Journal of the American Statistical Association* **89**: 1303–1313.
Rich G. 1998. The Euro and Swiss monetary policy. Speech delivered to the Norwegian-Swiss Euro Seminar on September 18, 1998.
Rich G. 2003. Swiss monetary targeting 1994–1996: the role of internal policy analysis. European Central Bank Working paper no. 236.
RiskMetrics Group. 1996. *RiskMetrics: Technical Document*. J. P. Morgan/Reuters: New York.
Sansó A, Arragó V, Carrion JL. 2004. Testing for change in the unconditional variance of financial time series. *Revista de Economiá Financiera* **4**: 32–53.
Stărică C. 2003. Is GARCH(1,1) as good a model as the Nobel accolades would imply? Working paper, Chalmers University of Technology, Göteborg.
Stărică C, Granger CWJ. 2005. Nonstationarities in stock returns. *Review of Economics and Statistics* **87**: 503–522.
Stărică C, Herzel S, Nord T. 2005. Why does the GARCH(1,1) model fail to provide sensible longer-horizon volatility forecasts? Working paper, Chalmers University of Technology, Göteborg.

Stock JH, Watson MW. 2003. Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature* **41**: 788–829.
Straumann D. 2005. *Lecture Notes in Statistics: Estimation in Conditionally Heteroskedastic Time Series Models*, Vol. 181. Springer: New York.
West KD. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.
West KD, Cho D. 1995. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics* **69**: 367–391.
West KD, McCracken MW. 1998. Regression-based tests of predictive ability. *International Economic Review* **39**: 817–840.
White H. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.

DOI: 10.1002/jae