Improving GARCH Volatility Forecasts with Regime-Switching GARCH

Klaassen, F.J.G.M.

Publication date
2002

Published in
Advances in Markov-Switching Models

Improving GARCH Volatility Forecasts with Regime-Switching GARCH

by Franc Klaassen *

Department of Economics
University of Amsterdam

July 31, 2001
Forthcoming in *Empirical Economics*

Abstract

Many researchers use GARCH models to generate volatility forecasts. Using data on three major U.S. dollar exchange rates we show that such forecasts are too high in volatile periods. We argue that this is due to the high persistence of shocks in GARCH forecasts. To obtain more flexibility regarding volatility persistence, this paper generalizes the GARCH model by distinguishing two regimes with different volatility levels; GARCH effects are allowed within each regime. The resulting Markov regime-switching GARCH model improves on existing variants, for instance by making multi-period-ahead volatility forecasting a convenient recursive procedure. The empirical analysis demonstrates that the model resolves the problem with the high single-regime GARCH forecasts and that it yields significantly better out-of-sample volatility forecasts.

Key words: GARCH, Markov-switching, variance, forecasting, exchange rates.
JEL classification: C52, C53, F31.

---

*Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands; tel: +31-20-5254191; fax: +31-20-5254254; E-mail: klaassen@fee.uva.nl.
I thank Harry Huizinga, Frank de Jong, Michael McAleer, Bertrand Melenberg, Theo Nijman, Arthur van Soest, Kenneth West, the editors of this issue, James Hamilton and Baldev Raj, and two referees for very useful and constructive comments.

1 Introduction

Volatility of financial returns is an important aspect of many financial decisions. For example, volatility of exchange rates is a determinant for pricing currency options used for risk management. Hence, there is a need for good volatility forecasts.

Such forecasts are often based on the fact that volatility is time-varying in high-frequency data and that periods of high volatility tend to cluster. To capture this, many authors use autoregressive conditional heteroskedasticity (ARCH) models, as introduced by Engle (1982) and extended to generalized ARCH (GARCH) in Bollerslev (1986); see Bollerslev, Chou and Kroner (1992) for an overview of the GARCH literature. Such models usually improve the fit a lot compared with a constant variance model and, as Andersen and Bollerslev (1998) claim, GARCH models provide good volatility forecasts.

This paper shows that GARCH forecasts are, nevertheless, too high in volatile periods, using almost twenty years of daily data on U.S. dollar exchange rates versus the British pound, German mark and Japanese yen. This suggests that better volatility forecasts can be obtained by solving that problem. The goal of this paper is to adapt the GARCH model in order to obtain such better forecasts.

The reason for the excessive GARCH forecasts in volatile periods may be the well-known high persistence of individual shocks in those forecasts. Lamoureux and Lastrapes (1990), among others, show that this persistence may originate from structural changes in the variance process. For example, if the variance is high but constant for some time and low but constant otherwise, persistence of such high- and low-volatility homoskedastic periods already results in volatility persistence (see also Timmermann (2000)). A GARCH model, which cannot capture persistence of such periods, puts all volatility persistence in the persistence of individual shocks. This idea is similar to Perron's (1989) work on the mean equation, as he finds that structural breaks in the mean make it more difficult to reject the null of a unit-root, that is, permanent persistence of shocks in the mean.

One possibility to allow for periods with different unconditional variances is, of course, by introducing deterministic shifts into the variance process, but this is rather ad hoc. A popular approach to endogenize changes in the data generating process is the Markov regime-switching model. Hamilton (1989) introduces this model to describe the U.S. business cycle, which is characterized by periodic shifts from recessions to expansions and vice versa. In our context of exchange rate volatility, a Markov process can be used to govern the switches between regimes with different variances. See Kaufmann and Scheicher (1996) for a survey on Markov-switching models.

To solve the problem of the excessive GARCH forecasts in volatile periods, we therefore generalize the GARCH model by allowing for regimes with different volatility levels. We use two regimes and do not consider models with more regimes, because we want to explore whether the introduction of regimes helps solve the problem with the GARCH forecasts and it turns out that two regimes are sufficient for that. Within each regime we use GARCH models to govern the variance. Hence, it is a regime-switching GARCH model. The persistence of both regimes yields an extra source of volatility persistence compared to standard, single-regime GARCH, thereby enhancing the flexibility in describing the volatility persistence of shocks.

The regime-switching GARCH model we develop differs from existing variants. First, it allows for GARCH dynamics, thereby generalizing the regime-switching ARCH models of Cai (1994) and Hamilton and Susmel (1994), and also used by Fong (1998). Because for our data the conditional heteroskedasticity within regimes cannot always be captured by a moderate number of ARCH terms, we need a GARCH term for parsimony. In addition, the data reveal that the variance dynamics differ across regimes. Our model allows for that, which is not the case for their models.

A second difference between our model and existing variants concerns Gray's (1996a) regime-switching GARCH specification, also used in Ang and Bekaert (1998). For that variant we are unable to compute the multi-period-ahead volatility forecasts needed for the detailed forecasting analysis we want to do in the current paper. In contrast, for our version such forecasts can be obtained through a convenient first-order recursive procedure. The idea behind this difference is that, when integrating out the unobserved regimes, we use all available information, whereas Gray uses only part of it; this also explains the better fit for our model. The development of our convenient regime-switching GARCH model is the main contribution of the paper from a theoretical point of view.

The main empirical results are that regime-switching GARCH resolves the problem that standard GARCH forecasts are significantly too high in volatile periods and that regime-switching GARCH forecasts significantly outperform GARCH forecasts in terms of mean squared error. These results hold out-of-sample and for both forecast horizons we examine, namely the one-day and ten-day horizons. This provides evidence for the conjecture raised by West and Cho (1995) that it will be productive to explore models that explicitly account for movement in the variance generating process, for instance, by regime switches.

The next section introduces the regime-switching GARCH model and discusses its properties. Section 3 describes the data used in the empirical application and presents

the empirical results. It also contains the out-of-sample forecasting exercise that yields the most important empirical results of the paper. Section 4 concludes.

2 Regime-Switching GARCH

In this section we introduce the regime-switching GARCH model, with which we try to improve on the standard, one-regime GARCH volatility forecasts. We describe the model, discuss its properties and relate the model to existing regime-switching ARCH and GARCH models.

2.1 The Model

We use the following notation. Let S_t denote the logarithm of a spot exchange rate at time t, that is, the domestic currency price of one unit of foreign currency. We concentrate on the exchange rate change $s_t = 100(S_t - S_{t-1})$, so that s_t is the percentage depreciation of the domestic currency from time $t-1$ to t.

The regime-switching GARCH model consists of four elements, namely the mean, regime process, variance and distribution. Two of them, the regime process and variance are crucial for interpreting the empirical results, as they are directly related to the difference between our model and standard, one-regime GARCH models.

The mean of exchange rate processes is often modeled by a random walk (with drift). For instance, Meese and Rogoff (1983) and MacDonald and Taylor (1992) stress the empirical quality of the random walk over structural models of exchange rate determination, particularly in the short run. We follow this simple but reasonable approach, also because the focus of our paper is on the volatility rather than the mean:

$$s_t = \mu + \varepsilon_t. \tag{1}$$

The innovation ε_t has zero mean conditional on the information set of the data generating process to be defined below. Thus, μ is the constant conditional mean of s_t. (It is possible to incorporate, for example, autoregressive terms in the conditional mean without making the formulas that follow essentially different.)

As argued in the introduction, the purpose of the regimes with different volatility levels is to explain part of the volatility persistence. This requires that regimes can be persistent. To model this, let $r_t \in \{1, 2\}$ be the (unobserved) variance regime at time t, where the first regime is identified as the low-variance one. Let $p_{t-1}(r_t \,|\, \widetilde{r}_{t-1}) = p(r_t \mid I_{t-1}, \widetilde{r}_{t-1})$ denote the probability of going to regime r_t at time t conditional on the information set of the data generating process, which consists of two parts.

The first part, I_{t-1}, denotes the information observed by the econometrician, that is $(s_{t-1}, s_{t-2}, \dots)$. The second part, $\widetilde{r}_{t-1}$, is the regime path $(r_{t-1}, r_{t-2}, \dots)$, which is not observed by the econometrician. Note that the subscript $t-1$ below an operator (probability, expectation or variance) is short-hand notation for conditioning on I_{t-1}.

As in Hamilton (1989), we assume that r_t follows a first-order Markov process with constant staying probabilities

$$p_{t-1}(r_t \,|\, \widetilde{r}_{t-1}) = p(r_t \,|\, r_{t-1}) = \begin{cases} p_{11} & \text{if } r_t = r_{t-1} = 1 \\ p_{22} & \text{if } r_t = r_{t-1} = 2. \end{cases} \tag{2}$$

If p_{11} and p_{22} are high, this specification results in the regime persistence required above.

The specification of the conditional variance, the third element of the model, represents the main difference between this paper and earlier ones on regime-switching ARCH and GARCH. Using the law of iterated expectations and (1), the conditional variance $V_{t-1}\{s_t\}$ equals $E_{t-1}[V_{t-1}\{\varepsilon_t \mid \widetilde{r}_t\}]$, so that we concentrate on $V_{t-1}\{\varepsilon_t \mid \widetilde{r}_t\}$. Four specifications of the latter variance will be discussed, where the final one turns out to be the most convenient. For the sake of exposition, we confine ourselves to models with only one ARCH and one GARCH term; including more ARCH and GARCH terms is straightforward.

The first specification of the conditional variance is a direct application of the GARCH(1,1) model in a regime-switching context:

$$V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\} = \omega_{r_t} + \alpha_{r_t}\varepsilon_{t-1}^2 + \beta_{r_t} V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\}, \tag{3}$$

where $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ denotes the variance of ε_t conditional on observable information I_{t-1} and on the regime path $\widetilde{r}_t$. The current regime only determines the parameters, that is, the intercept ω_{r_t}, the ARCH parameter α_{r_t} and the GARCH parameter β_{r_t}.

This specification, however, appears practically infeasible when estimating the model. This is due to the fact that $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ in (3) depends on the entire regime path $\widetilde{r}_t$, because it depends on r_t and $V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\}$, which depends on r_{t-1} and $V_{t-3}\{\varepsilon_{t-2} \,|\, \widetilde{r}_{t-2}\}$, which depends on r_{t-2} and $V_{t-4}\{\varepsilon_{t-3} \,|\, \widetilde{r}_{t-3}\}$, and so on. Since the number of possible regime paths grows exponentially with t, this leads to an enormous number of paths to t. The econometrician, who does not observe regimes, has to integrate out all possible paths when computing the sample likelihood. This renders estimation intractable. The remaining specifications of the conditional variance are ways to avoid this problem of path dependence.

The second specification is based on Cai (1994) and Hamilton and Susmel (1994). They essentially remove the GARCH term, which is the cause of the path dependence,

and thus use only an ARCH term in (3). Since $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ then only depends on the current regime r_t, there is no problem of path dependence. (More precisely, Cai (1994) and Hamilton and Susmel (1994) use slightly different models in which $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ not only depends on the current but also on a few recent regimes. The essential point is that the conditional variance depends only on a small number of regimes, which can be integrated out in the likelihood quite easily.)

The third specification of the conditional variance comes from Gray (1996a). He argues that the problem of path dependence can be solved without giving up the potentially important persistence effects of a GARCH term, as has been done in the second specification. The basic idea of Gray is to integrate out the unobserved regime path $\widetilde{r}_{t-1}$ directly in the source of the path dependence, $V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\}$ in (3), instead of only in the likelihood. This makes $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ only depend on the current regime r_t, not on the path $\widetilde{r}_{t-1}$, as is clear from the explanation of the path dependency problem below (3). As Gray shows, this is very convenient from an estimation point of view, because the likelihood can then be computed in a first-order recursive way, which speeds up the estimation process considerably. Since Gray uses the information observable at time $t-2$ when integrating out, he actually assumes that

$$V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\} = \omega_{r_t} + \alpha_{r_t}\varepsilon_{t-1}^2 + \beta_{r_t} E_{t-2}\left[V_{t-2}\{\varepsilon_{t-1}|\widetilde{r}_{t-1}\}\right], \tag{4}$$

where the expectation on the right-hand-side is across the regime path $\widetilde{r}_{t-1}$, conditional on information I_{t-2}. Note that this is equivalent to integrating out only the single regime r_{t-1}, as the lag of (4) implies that $V_{t-2}\{\varepsilon_{t-1}|\widetilde{r}_{t-1}\}$ is independent of $\widetilde{r}_{t-2}$.

The main benefit of specification (4) is that there is no path dependence problem any more, although GARCH effects are still allowed. There is, however, one important inconvenience, especially regarding our focus of volatility forecasting: generating multi-period-ahead variance forecasts such as $V_{t-1}\{s_{t+1}\}$ turns out to be very complicated. This motivates our search for another specification that makes multi-period-ahead forecasting more convenient while preserving the attractive features of Gray's model.

Our specification of $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ differs from Gray's (1996a) model in two ways. First, as the expectation in (4) shows, Gray integrates out the regime r_{t-1} at time $t-2$. We postpone this till $t-1$, the time at which the conditional variance $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ is really needed. This allows us to use more observable information when integrating out the previous regime. This extra data embodies information about previous regimes and is thus useful.

The second difference is that, when integrating out the regime r_{t-1}, Gray does not use the information that the regime at time t is in the conditioning information of $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$. Particularly if regimes are highly persistent, r_t gives much information

about r_{t-1}. In contrast to Gray, we do use this information.

In formula, our regime-switching GARCH(1,1) model is described by

$$V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\} = \omega_{r_t} + \alpha_{r_t}\varepsilon_{t-1}^2 + \beta_{r_t} E_{t-1}\left[V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\} \middle| \, r_t \right], \tag{5}$$

where the expectation on the right-hand-side is across the regime path $\widetilde{r}_{t-1}$, conditional on information I_{t-1} and r_t. Note that this is equivalent to integrating out only the single regime r_{t-1}, as the lag of (5) implies that $V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\}$ is independent of $\widetilde{r}_{t-2}$. By construction, $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ only depends on the current variance regime r_t, so that $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\} = V_{t-1}\{\varepsilon_t \,|\, r_t\}$. Hence, there is no problem of path dependence. To complete the specification of the conditional variance, we impose $\omega_{r_t} > 0$ and α_{r_t}, $\beta_{r_t} \geqslant 0$ to ensure positivity of $V_{t-1}\{\varepsilon_t \,|\, \widetilde{r}_t\}$ for all t, just as for single-regime GARCH.

The final element of the regime-switching GARCH model is the conditional distribution. We assume that, conditional on I_{t-1} and $\widetilde{r}_t$, the innovation ε_t has a t-distribution with ν degrees of freedom, where ν is assumed to be independent of the conditioning information, and with mean zero and variance $V_{t-1}\{\varepsilon_t \,|\, r_t\}$:

$$\varepsilon_t \,|\, I_{t-1}, \widetilde{r}_t \quad \sim \quad t\left(\nu \,,\, 0 \,,\, V_{t-1}\{\varepsilon_t \,|\, r_t\}\right). \tag{6}$$

The use of a t-distribution instead of a normal one is quite popular in the standard, single-regime GARCH literature (see Bollerslev, Chou and Kroner (1992)). For regime-switching models, a t-distribution can be extra useful. After all, in case of normality, a large innovation in the low-volatility period will lead to a switch to the high-volatility regime earlier, even if it is a single outlier in an otherwise tranquil period. Allowing for a t-distribution will thus enhance the stability of the regimes. Note that the t-distribution includes the normal distribution as the limiting case where the degrees of freedom go to infinity.

In summary, equations (1), (2), (5) and (6) describe our regime-switching GARCH model. It contains the standard, one-regime GARCH(1,1) model as a special case, since that model results when all regime-specific parameters are equal across regimes.

2.2 Properties of the Model

The model just described has several interesting properties. We first show the increased flexibility regarding the volatility persistence of shocks. After that we present the convenient procedure for multi-period-ahead volatility forecasting, which we need to examine the forecast quality of regime-switching GARCH compared with single-regime GARCH, the focus of the paper. Then the unconditional variance, the estimation procedure, and inference about the unobserved regime are discussed.

2.2.1 Flexibility Regarding Volatility Persistence

As motivated in the introduction, the reason to generalize the single-regime GARCH model by introducing regimes is to enhance the flexibility of the model to capture the persistence of shocks in volatility. One example is that in regime-switching GARCH a shock can be followed by a volatile period not only because of GARCH effects, but also because of a switch to the high-volatility regime.

The flexibility with respect to volatility persistence is further improved by the allowance for different ARCH and GARCH parameters across regimes. For instance, if shocks are more persistent in periods of high than in periods of low volatility, this can be captured by the regime specific parameters in (5). This has consequences for capturing the "pressure relieving" effect of some large shocks, that is, some shocks are not persistent at all but are followed by a tranquil period. Any regime-switching model can capture this to some extent by a shift from the high-volatility to the low-volatility regime. However, our regime-switching model with different parameters across regimes has a second source of neglecting large recent shocks. After all, if the low-variance regime is also the low-persistence regime, the large shock will be out of the market very soon after the switch to the low-variance regime. In this respect, our model generalizes the models in Hamilton and Susmel (1994) and Cai (1994), as their regime variances only differ by a multiplicative or additive constant, respectively, not by differences in the ARCH parameters.

2.2.2 Recursive Volatility Forecasting

Suppose we need the variance of the exchange rate change over a horizon h, conditional on information available at time $t-1$. Let $s_{t,h}$ denote the h-period change, that is, $s_{t,1} = s_t$ and $s_{t,h} = s_t + ... + s_{t-1+h}$ for $h > 1$. The variance of interest is thus $V_{t-1}\{s_{t,h}\}$. Because of the absence of serial correlation in the one-period changes (see below (1)),

$$V_{t-1}\{s_{t,h}\} = \sum_{\tau=t}^{t-1+h} V_{t-1}\{s_\tau\}. \tag{7}$$

Each variance on the right-hand-side is equal to

$$V_{t-1}\{s_\tau\} = \sum_{r_\tau=1,2} p_{t-1}(r_\tau) \cdot V_{t-1}\{\varepsilon_\tau \,|\, r_\tau\}, \tag{8}$$

where $p_{t-1}(r_\tau)$ is the probability that the regime at time τ is r_τ conditional on I_{t-1}. Note that we use the same symbol p_{t-1} for several probabilities (for instance, see (2)

and (8)). The specific meaning of p_{t-1} is uniquely determined by the symbols in its argument. This results in a concise notation.

An important implication of our way of modeling the conditional variance in (5) is that $V_{t-1}\{\varepsilon_\tau \,|\, r_\tau\}$ in (8) can be computed in a first-order recursive manner using a formula analogous to the one Engle and Bollerslev (1986) have derived for the standard, one-regime GARCH model. Starting from $V_{t-1}\{\varepsilon_t \,|\, r_t\}$, appendix A shows that one can compute $V_{t-1}\{\varepsilon_\tau \,|\, r_\tau\}$ for $\tau > t$ by iterating forward on

$$V_{t-1}\{\varepsilon_{t+i} \,|\, r_{t+i}\} = \omega_{r_{t+i}} + (\alpha_{r_{t+i}} + \beta_{r_{t+i}}) \cdot E_{t-1}\left[V_{t-1}\{\varepsilon_{t+i-1} \,|\, r_{t+i-1}\} \middle| r_{t+i} \right] \qquad (9)$$

for $i = 1, ..., \tau - t$. This simplifies the computation of $V_{t-1}\{s_{t,h}\}$ in (7) substantially and represents one of the main advantages of our regime-switching GARCH model over Gray's (1996a) model.

2.2.3 Unconditional Variance

In Appendix B we derive the following results for the "unconditional" error variance $V\{\varepsilon_t \mid r_t\}$. First, if $V\{\varepsilon_t \mid r_t = i\}$ exists for both $i = 1, 2$ and both ω_1, ω_2 and is independent of t, denoted by σ_i^2, then

$$\begin{bmatrix} \sigma_1^2 \\ \\ \sigma_2^2 \end{bmatrix} = (I_2 - A)^{-1} \cdot \begin{bmatrix} \omega_1 \\ \\ \omega_2 \end{bmatrix}, \qquad (10)$$

where I_2 is the identity matrix of order two and the (i, j)-th element of A is $A_{ij} = P\{r_{t-1} = j \,|\, r_t = i\}(\alpha_i + \beta_i)$ (appendix B gives expressions for the probabilities in A_{ij}).

Second, necessary conditions for the existence of both variances are $A_{11}, A_{22} < 1$ and $\det(I_2 - A) > 0$. So, given the definition of A_{ii}, a probability times the sum of the regime-specific ARCH and GARCH coefficients must be less than one for both regimes. Moreover, there is some restriction on a combination of the ARCH and GARCH coefficients across regimes.

To get a better understanding of these results, let us look at standard GARCH(1,1). There the unconditional variance is $\sigma^2 = (1 - \alpha - \beta)^{-1}\omega$ and the necessary (and sufficient) condition for its existence is $\alpha + \beta < 1$. Hence, we again see a correspondence between single-regime GARCH and regime-switching GARCH.

2.2.4 Recursive Estimation

The regime-switching GARCH model can be estimated by maximum likelihood (ML). The likelihood function is derived in appendix C, using similar techniques as Gray

(1996a). As for Gray's model, this likelihood has a first-order recursive structure, similar to that of single-regime GARCH. This speeds up the estimation process.

2.2.5 Recursive Regime Inference

Although regimes are not observed, one can estimate the probability that the process is in a particular regime at a specific time. This is, for instance, useful if one wants to classify a series into periods of low and high volatility.

Following Gray (1996a), we use two types of regime probabilities, namely ex ante and smoothed probabilities. The ex ante probability of regime r_t at time t, $p_{t-1}(r_t)$, is the conditional probability that the process is in that regime at time t using only information available to the econometrician at time $t-1$, that is, I_{t-1}. The smoothed regime probability $p_T(r_t)$, on the other hand, uses the complete data set I_T, thereby smoothing the ex ante probabilities. Hence, it gives the most informative answer to the question which regime the process was likely in at time t. The ex ante probabilities are computed during estimation (see appendix C). The smoothed probabilities can be calculated in a recursive manner starting from the ex ante probabilities, as appendix D shows using an algorithm based on Gray (1996b).

3 Empirical Results

So far, we have generalized single-regime GARCH to regime-switching GARCH to obtain more flexibility regarding the volatility persistence of shocks. In this section we estimate both models and examine whether that generalization pays off empirically in terms of improved volatility forecasts, the central issue of the paper.

3.1 Data

We consider three major U.S. dollar exchange rates, namely, the dollar price of the British pound, the German mark and the Japanese yen. We have 4,982 daily observations for the exchange rate change s_t from January 3, 1978 to July 23, 1997. All rates have been obtained from Datastream.

Panel A of figures 1, 2 and 3 gives an indication of the volatility clustering of the three exchange rates under consideration over the sample period. As usual, all three plots show substantial volatility clustering. This is confirmed by (not reported) Box-Pierce tests for serial correlation in the squared exchange rate changes, as these are significant at any reasonable significance level.

The plots also demonstrate that shocks sometimes have a long effect on subsequent volatility, but that shocks can also be followed by a period of low volatility. For instance, in figure 1A the large peak in the squared change plot for the British pound on March 27, 1985 was followed by about half a year of substantial volatility. On the other hand, the G-5 Plaza announcement on September 22, 1985 to bring about a dollar depreciation had a sharp effect on the dollar the next day, as the second largest peak in the figure makes clear, but was followed by a period of low instead of high volatility. Therefore, at first sight the extra flexibility regarding volatility persistence that is present in regime-switching GARCH seems worthwhile.

3.2 Estimation Results

This subsection presents the estimation results for the regime-switching GARCH model. Let GARCH($P_1, Q_1; P_2, Q_2$) denote a regime-switching model with Q_1 (Q_2) ARCH and P_1 (P_2) GARCH terms in the first (second) regime. These are obvious variants of the GARCH(1,1;1,1) model developed in subsection 2.1. The models for the pound contain an AR(1) term in mean equation (1) to correct for the small first-order autocorrelation found in the data.

For comparison, we also estimate five other models. Two of them are single-regime models, namely the constant variance model and the popular GARCH(1,1) model. Two other models belong to the regime-switching ARCH class. The ARCH(0;0) model has zero ARCH terms, so constant variance, in both regimes, as in Dewachter (1997) and Scheicher (1999); this model is used to analyze the effect of introducing only regimes. The other ARCH-type model, ARCH(Q_1;4) with Q_1 determined below, is in the spirit of Cai (1994) and Hamilton and Susmel (1994). It is, however, somewhat more general in the sense that the regime-specific ARCH models are allowed to vary across regime in an unrestricted way, whereas in Cai (1994) and Hamilton and Susmel (1994) the difference between the low- and high-variance regime ARCH models is just an additive or multiplicative constant, respectively. The final model for comparison is Gray's (1996a) variant of regime-switching GARCH.

Table 1 presents the maximum likelihood estimation results. The inverse of the degrees of freedom ν of the t-distribution (see (6)) is presented; testing for conditional normality then boils down to testing whether ν^{-1} differs significantly from zero. Moreover, for easier comparison of the models, we report the regime-specific unconditional variances $\sigma_r^2\,(r = 1, 2)$ instead of the intercepts ω_r in the conditional variance formula (5); see (10) for the computation of σ_r^2. Finally, the last column of table 1 reports the log-likelihood using GARCH(1,1) as a reference, so that the values for the other

models are differences with respect to GARCH(1,1). Note that one should be careful when interpreting differences in log-likelihoods in terms of likelihood ratio tests. First, not all models are nested. Second, testing the null of a single-regime against a regime-switching model involves unidentified parameters (the regime-staying probabilities) under the null, so that the asymptotic distribution of the likelihood ratio is not the usual χ^2-distribution (see Hansen (1992)). In this paper we do not formally test for the significance of the second regime, because the focus of the paper concerns forecasting quality, so that we concentrate on the effects of regimes on that.

3.2.1 Single-regime GARCH

As is typically found, the standard, one-regime GARCH(1,1) model provides a much better fit than the constant variance model. For instance, the increase in log-likelihood of the GARCH model over the constant variance model is 244.34 for the British pound, so that ARCH and GARCH effects are statistically very important. GARCH(1,1) is also the preferred model within the class of GARCH(P, Q) models, as the likelihood ratios of GARCH(1,1) versus GARCH(2,1) and GARCH(1,2) are 1.12 and 0.00, respectively, for the pound, 0.92 and 0.00 for the mark, and 1.90 and 0.00 for the yen, which are all insignificant. This is in accordance with Bollerslev et al. (1992), who state that in most applications $P = Q = 1$ is sufficient.

As usual, the estimated sum of the ARCH and GARCH parameters ($\alpha + \beta$) is large for all three series, pointing at high volatility persistence of individual shocks. This may indicate parameter instability, as argued in the introduction. We estimate regime-switching models to analyze whether the high volatility persistence is indeed spurious.

3.2.2 Regime-switching ARCH(0;0)

Let us first consider the regime-switching ARCH(0;0) model, in which persistence of regimes is the only source of volatility clustering. Table 1 shows that for the three rates there is a distinction between a low- and a high-volatility regime, where the unconditional variance in the latter is three to four times as large.

The variance regimes are also persistent, since the staying probabilities p_{11} and p_{22} are all above 0.975. To get a better idea about the amount of persistence that such staying probabilities imply, we compute the expected duration of the high-variance

regime. Conditional on being in this regime ($r_t=2$), this is (see Hamilton (1989))

$$\sum_{h=1}^{\infty} h \cdot P\{r_t=2,\ldots,r_{t+h-1}=2, r_{t+h}=1 \,|\, r_t=2\} = \sum_{h=1}^{\infty} h \cdot (p_{22})^{h-1}(1-p_{22}) = (1-p_{22})^{-1}. \tag{11}$$

For a typical ARCH(0;0) staying probability of 0.98, this implies an expected duration of 50 (working) days, which is about 2.5 months.

The log-likelihood gives a first idea of whether the regime persistence is an important source of volatility clustering. For the pound and mark the log-likelihood is lower than for GARCH(1,1), but for the yen it is higher. Hence, regimes can be an important mechanism to capture volatility clustering.

This is confirmed by table 2, which gives tests for autocorrelation in the squared normalized residuals (see the notes below the table for the computation of the normalized residuals). The first-order autocorrelations ρ_1 and the Box-Pierce tests Q_{10} show that the conditional heteroskedasticity in the normalized residuals is greatly reduced when going from the constant variance model to the regime-switching model with constant regime-specific variances. However, the conditional heteroskedasticity tests also make clear that there is still heteroskedasticity left (we use a significance level of 5% throughout the paper). Apparently, there is also volatility clustering within a regime.

3.2.3 Regime-switching ARCH(Q_1;4)

To capture the remaining conditional heteroskedasticity, we first add only ARCH terms to the model, so no GARCH terms yet. To get some insight into the magnitude of volatility clustering across the regimes, we start with a model with several ARCH terms in both regimes. For parsimony, we restrict the number of ARCH terms to four in both regimes, that is, ARCH(4;4). We find that four ARCH terms is too much for the low-variance regime: for the pound two ARCH terms suffice (likelihood ratio of ARCH(2;4) versus ARCH(4;4) is 0.74, which is insignificant because the p-value is [0.69]), for the mark zero terms suffice (1.87 [0.76]) and for the yen one term (3.79 [0.29]). In contrast, the high-volatility regime keeps its four ARCH terms, as reducing that number to the number of ARCH terms in the first regime yields likelihood ratios of 23.71 [0.00] for the pound, 19.76 [0.00] for the mark, and 7.06 [0.07] for the yen. We thus obtain ARCH(2;4) for the pound, ARCH(0;4) for the mark, and ARCH(1;4) for the yen, thereby highlighting that there is more volatility clustering in the high- than in the low-variance regime for our data.

The latter result is supported by Chaudhuri and Klaassen (2000), who find for weekly data on East Asian stock index returns that there is more conditional het-

eroskedasticity in the high- than in the low-volatility regime. Our evidence, however, is in contrast with the models in Cai (1994) and Hamilton and Susmel (1994). Their regime-specific ARCH models only differ by an additive or multiplicative parameter, respectively, so that, for instance, the number of ARCH terms is the same across regimes. Since we find evidence of longer volatility persistence in the high-volatility regime, we prefer our asymmetric approach for the data in this paper.

The usefulness of the regime-switching ARCH approach appears from the tests in table 2. For the yen there is no remaining conditional heteroskedasticity after estimation of ARCH(1;4). For the other two exchange rates, however, the regime-switching ARCH models are insufficient. The remaining conditional heteroskedasticity can be attributed to the high-variance regime, as the likelihood ratios given above show that higher-order ARCH estimates are insignificant for the low-volatility regime.

3.2.4 Regime-switching GARCH

The residual conditional heteroskedasticity can be modeled by adding ARCH terms to the high-volatility part of ARCH(Q_1;4). However, that increases the number of parameters substantially. For reasons of parsimony it is better to use a GARCH term in the high-variance regime. This leads to regime-switching GARCH(0,Q_1;1,1).

Table 2 shows that the evidence of residual conditional heteroskedasticity that was present for regime-switching ARCH on the pound and mark has disappeared when using regime-switching GARCH. Moreover, table 1 demonstrates that the log-likelihood increases a lot after the introduction of GARCH terms, namely 38.90 for the pound and 23.41 for the mark. Remarkably, this increase is achieved by using fewer instead of more parameters. After all, the regime-switching GARCH models have two parameters less than regime-switching ARCH, and the difference becomes even larger when the regime-switching ARCH models are extended to capture the residual volatility clustering. For Japan, with no residual conditional heteroskedasticity after ARCH(1;4), it is not surprising the increase in the log-likelihood is negligible (0.24). However, also there GARCH(0,1;1,1) has fewer parameters than ARCH(1;4). We thus find that GARCH terms can be important to capture volatility persistence. Subsection 3.2.2 has shown that regimes are also important. The advantage of regime-switching GARCH models is that they allow for both.

The outperformance of regime-switching GARCH over regime-switching ARCH also holds for the fourth-order regime-switching ARCH variants in Cai (1994) and Hamilton and Susmel (1994). First, regime-switching GARCH removes the residual volatility clustering that is present for their models for the pound and mark. Second, the in-

crements in the log-likelihood are 40.46 (pound), 24.93 (mark) and -0.34 (yen) for the Cai version, and 43.12 (pound), 23.14 (mark) and 1.10 (yen) for the Hamilton-Susmel model. Third, regime-switching GARCH is more parsimonious: for the pound, mark, and yen the number of parameters is 0, 2, and 1 lower, respectively, than for their models and the difference becomes larger when the Cai and Hamilton-Susmel models are extended to account for the residual volatility clustering.

Next, we relate our version (5) of regime-switching GARCH to Gray's (1996a) variant (4). As table 1 shows, the parameter estimates for Gray's specification are roughly the same as for our specification. However, the log-likelihood for Gray's specification is lower, namely 10.10, 11.73 and 3.93 for the three rates. This is because Gray's model makes less efficient use of the conditioning information when integrating out regimes (see below (4)). This is perhaps also the reason why there is some conditional heteroskedasticity left in the normalized residuals for Gray's model. Besides the theoretical advantages, as given in section 2, we thus also find empirical support for our model over Gray's variant.

Figures 1B, 2B, and 3B provide some additional insight into our regime-switching GARCH model. They plot the estimated smoothed probabilities of being in the high-volatility regime, as defined in subsection 2.2.5. The two European currencies have experienced fewer regime shifts than the Japanese yen. Apparently, sudden shifts in the variance are more important for the description of the yen than for the European currencies, where the conditional variance is governed more by smooth transitions (GARCH effects) from high-volatility periods to low ones. This supports the conclusion given above that both regimes and GARCH terms can be important.

An issue closely related to the persistence of regimes is the allowance for extra leptokurtosis by a t-distribution, as in (6). Without this, the persistence of the, for example, low-volatility regime would have been lower, since then a large sudden change in the exchange rate would have been considered earlier as a shift to the high-volatility regime. This is illustrated by figure 1C, which gives the smoothed regime probabilities of the regime-switching GARCH model for the British pound under the restriction of normality: more regime-switches occur.

3.2.5 Comparison of regime-switching with single-regime GARCH

Though our regime-switching GARCH model outperforms regime-switching ARCH and Gray's regime-switching GARCH, the main reason to introduce the model was to improve on single-regime GARCH. Using the log-likelihoods in table 1, we indeed document an increased fit of 19.34 (UK), 8.44 (Germany) and 16.52 (Japan). Because

the regime-switching GARCH methodology generalizes single-regime GARCH, this improvement is presumably not surprising, even though the GARCH(0,Q_1;1,1) variants used here do not strictly encompass GARCH(1,1).

It is, however, interesting to find out where the improved fit originates from, so as to derive the key differences between the two models. We do this in two stages. First, we examine for which kinds of observations regime-switching GARCH outperforms GARCH. Then we show which model differences are responsible for that.

Since both models focus on volatility, any difference in fit is presumably related to the volatility. Therefore, to find out when regime-switching GARCH is better, we regress the log-likelihood contribution of an observation for regime-switching GARCH minus that for GARCH, dl_t, on a simple measure of past volatility, s_{t-1}^2, and its square. We correct the standard errors for autocorrelation and heteroskedasticity using the Newey and West (1987) asymptotic covariance matrix. (Following West and Cho (1995), we take Bartlett weights and use the same data-dependent automatic lag selection rule. This rule has certain asymptotic optimality properties and was introduced by Newey and West (1994).) The regression results (not reported) show that, although both slope estimates are positive for all three countries, they are all insignificant.

A potential reason for this insignificance is that both dl_t and s_{t-1}^2 are very volatile and may contain much noise. For instance, s_{t-1}^2 is sometimes low even in an otherwise volatile period. To reduce the effect of both sources of noise, we transform dl_t into the binary variable $1[dl_t > 0]$, which is one if regime-switching GARCH is better, and proxy past volatility by the logarithm of the average of, say, ten past squared changes $s_{t-1}^2, ..., s_{t-10}^2$. The regression model thus now tries to explain the probability of outperformance from past volatility. The estimates for past volatility and its square are again positive for all three countries, but now they are clearly significant (t-values between 6 and 9). A plot of the parabolic dependence of the estimated probability of outperformance on past volatility shows that regime-switching GARCH outperforms single-regime GARCH particularly in tranquil and volatile periods.

Next, we analyze the reasons for this. Because the main difference between the two models concerns the variance specification, a difference in the log-likelihood contributions is very likely caused by a difference in the variance estimates, $d\widehat{V}_{t-1}\{s_t\}$. Indeed, a graph of $d\widehat{V}_{t-1}\{s_t\}$ against past volatility demonstrates that regime-switching GARCH has lower variance estimates in tranquil as well as volatile periods. We thus conclude that the improved fit originates from lower variance estimates in both tranquil and volatile periods.

To explain this in terms of the model differences, we first consider the volatile pe-

riods. The regime-switching model is then mainly in the high-volatility regime 2, so that the difference with GARCH likely originates from the differences between the estimated second regime parameters and the estimated standard GARCH parameters. Indeed, if we reestimate the regime-switching GARCH model under the restriction $(\alpha_{12}, \beta_2) = (\widehat{\alpha}, \widehat{\beta})$ and then again regress $1[dl_t > 0]$ on past volatility and its square, the outperformance of regime-switching GARCH in volatile periods disappears. (This may be surprising, as the differences between $(\widehat{\alpha}_{12}, \widehat{\beta}_2)$ and $(\widehat{\alpha}, \widehat{\beta})$ in table 1 are small at first sight. Nevertheless, a likelihood ratio test rejects the restriction $(\widehat{\alpha}_{12}, \widehat{\beta}_2) = (\widehat{\alpha}, \widehat{\beta})$ for Germany and Japan (not for the UK). Hence, the small differences are relevant.) Since restricting both α_{12} and β_2 is also necessary to remove the outperformance, the difference between the estimates of (α_{12}, β_2) and (α, β) is the reason for the outperformance in volatile periods. From table 1 we see that according to the regime-switching model shocks have a smaller direct effect on the volatility estimates $(\widehat{\alpha}_{12} < \widehat{\alpha})$ and their subsequent impact is also lower $(\widehat{\alpha}_{12} + \widehat{\beta}_2 < \widehat{\alpha} + \widehat{\beta})$; see Lamoureux and Lastrapes (1990) for this interpretation of α and β. We thus conclude that the outperformance in volatile periods is due to the smaller effect of shocks on variance estimates.

Next, we explain the outperformance in tranquil periods, which is caused by the lower regime-switching variance estimates. To abstract from the outperformance in volatile periods, this paragraph uses the restricted regime-switching GARCH model introduced above, so $(\alpha_{12}, \beta_2) = (\widehat{\alpha}, \widehat{\beta})$. This model also outperforms standard GARCH in tranquil periods and yields lower variance estimates there. The main difference between regime-switching and single-regime GARCH in tranquil periods is that the former has a separate regime for such periods. This regime is relevant for the outperformance if the latter depends on the probability of being in that regime, $P_{t-1}\{r_t = 1\}$. Therefore, we add the estimated probability as a regressor to the model that explains $1[dl_t > 0]$ from past volatility and its square. The estimated effect of the probability is significantly positive for all three countries (t-values between 7 and 11). Moreover, the effect of past volatility has disappeared. Hence we conclude that the outperformance in tranquil periods is obtained by the use of the low-variance regime.

The low-variance regime is used in two ways. First, it can explain why volatility is low for a long time. Second, as explained in subsection 2.2.1, it helps describe that several shocks are "pressure relieving," that is, are followed by a tranquil instead of volatile period. Figure 1D clarifies this by visualizing the impact of two particular shocks. It contains the conditional variance estimates of both GARCH(1;1) and GARCH(0,2;1,1) for the British pound over 1985 only. The persistent effect of the first shock on subsequent volatility is captured by both models (though the regime-switching GARCH

variances are less affected by the shock, in line with our argument above). On the other hand, the pressure relieving effect of the second shock, which is the sharp fall in the dollar one day after the G-5 Plaza announcement on September 22, 1985, is better described by the regime-switching model. The reason is a temporary switch to the low-volatility regime, which helps reduce the variance estimates rather quickly.

3.3 Forecasting Performance

So far, we have developed a regime-switching GARCH model to obtain more flexibility in capturing the persistence of shocks in volatility. We have shown that this is worthwhile from an in-sample point of view. In this subsection we analyze whether regime-switching models can also improve on the out-of-sample performance of single-regime GARCH.

The volatility forecasts of interest are the forecasts at time $t-1$ of the variance of the exchange rate change over a h-day horizon, that is, $\widehat{V}_{t-1}\{s_{t,h}\}$. They follow from subsection 2.2.2 after substitution of the estimation results of table 1. We analyze two forecast horizons, namely one day ($h = 1$) and ten days ($h = 10$).

To get some insight into the generality of the results, we need an extensive out-of-sample period. Therefore, we split the sample into two parts of both 2,491 days; the second half starts at October 20, 1987. As usual, we reestimate the models using the first half and, keeping the parameters fixed to save on estimation time, use the observations of the second half to generate the forecasts $\widehat{V}_{t-1}\{s_{t,h}\}$. We also do the reverse, that is, estimate the parameters on the second half and use the first half for forecasting.

To investigate the quality of the volatility forecasts, we need some measure of "observed volatility." Since $V_{t-1}\{s_{t,h}\} = E_{t-1}\{(s_{t,h} - h \cdot \mu)^2\}$, an obvious candidate is the (mean adjusted) squared change $(s_{t,h} - h \cdot \mu)^2$. However, one can obtain a more accurate measure following an idea advocated by Merton (1980) and Schwert (1989) and formalized by Andersen and Bollerslev (1998). They argue that the single squared change, though unbiased, is a noisy indicator for the latent volatility in the period, because the idiosyncratic component of a single change is large. The noise is reduced by taking the sum of all squared intra-period changes, and the smaller the subperiods, the larger the noise reduction. Since the highest frequency available to us is daily data, this idea results in the sum of squared daily changes over the h days in the forecast period: $\sum_{\tau=t}^{t-1+h}(s_\tau - \mu)^2$. This measure is unbiased, just as the single squared change, but it is more accurate (for $h > 1$; for $h = 1$ both measures are equivalent). Therefore, we prefer this measure. Substituting the estimate $\widehat{\mu}$ for μ, we thus define observed

volatility $v_{t,h}$ over the h days $t, ..., t-1+h$ as

$$v_{t,h} = \sum_{\tau=t}^{t-1+h} (s_\tau - \widehat{\mu})^2. \tag{12}$$

As stated in the introduction, the paper is motivated by the claim that single-regime GARCH forecasts are too high in volatile periods. This claim is based on the standard forecast efficiency regression

$$v_{t,h} = \gamma_0 + \gamma_1 \widehat{V}_{t-1}\{s_{t,h}\} + \eta_\tau \tag{13}$$

(see also Pagan and Schwert (1990)). If the mean and variance forecasts are (conditionally) unbiased, that is, $\widehat{\mu} = E_{t-1}\{s_t\}$ and $\widehat{V}_{t-1}\{s_{t,h}\} = V_{t-1}\{s_{t,h}\}$, then regression (13) implies $\gamma_0 = 0$ and $\gamma_1 = 1$. To test both implications we estimate (13) by OLS and correct the standard errors for autocorrelation and heteroskedasticity following Newey and West (1987), as explained in subsection 3.2.5. We also correct the standard errors for the uncertainty originating from the fact that the parameters used to compute the forecasts are not known but are estimated. This correction is based on West and McCracken (1998). As we keep the parameters fixed over the forecasting period, we have what they call the "fixed sampling scheme". Because in our study the in-sample and out-of-sample periods have the same number of observations, West and McCracken show that we have to multiply the Newey-West standard errors by $\sqrt{2}$.

The results are in table 3. For each model and horizon we have two estimates for both γ_0 and γ_1; the left one is based on the usual procedure of estimating the parameters from the first half of the sample and obtaining forecasts from the second half, while the right one is computed from the reverse procedure. We see for the GARCH(1,1) model that in eight out of twelve cases both implications $\gamma_0 = 0$ and $\gamma_1 = 1$ are significantly rejected (an asterisk for the estimate of γ_1 means that it is significantly different from one, not zero). For all twelve cases the estimate of γ_0 is larger than zero and the estimate of γ_1 is smaller than one. This is in line with the results of West and Cho (1995), among others.

The finding of $\gamma_0 > 0$ and $\gamma_1 < 1$ suggests that low GARCH(1,1) forecasts underestimate the true volatility or that high forecasts overestimate volatility, or both. To distinguish between the two cases we reestimate (13), but now allowing for a break in the regression line at, say, the median forecast (allowing for more breaks does not alter the conclusion). That is, one pair (γ_0^-, γ_1^-) is relevant for forecasts below the median and another pair (γ_0^+, γ_1^+) for forecasts above the median. The results (not tabulated) show that the estimates of (γ_0^-, γ_1^-) are close to (0, 1) (average estimate is (-0.00, 1.02)

for the one-day and (-0.38, 1.09) for the ten-day horizon) and that they are nowhere significantly different from (0, 1). The estimates of (γ_0^+, γ_1^+), however, differ substantially from (0, 1) (averages are (0.23, 0.55) and (2.96, 0.43)) and in nine out of twelve cases the difference is significant. Therefore, high GARCH forecasts generally overestimate the true variance, while low GARCH forecasts do not underestimate volatility. This is in line with the in-sample result that regime-switching GARCH improves on GARCH by reducing the high GARCH forecasts in volatile periods and by reducing instead of increasing the low GARCH forecasts in tranquil periods (see subsection 3.2.4). We thus conclude that single-regime GARCH volatility forecasts are too high in volatile periods.

To compare the regime-switching models to GARCH in this respect, we return to the standard forecast efficiency regression (13), so without the break. Table 3 shows that the regime-switching models do better than GARCH(1,1), as $\gamma_0 = 0$ and $\gamma_1 = 1$ are generally not rejected. Apparently, the excessive GARCH forecasts are sufficiently reduced by the regime-switching models. In subsection 3.2.5 we have shown that this is caused by the smaller persistence of shocks in volatility. Hence, allowing for more flexibility in volatility persistence by using regimes is worthwhile to improve the standard GARCH forecasts in the sense of regression (13).

Another way to compare the forecasts is by using the mean squared error (MSE) defined as the mean of $(v_{t,h} - \widehat{V}_{t-1}\{s_{t,h}\})^2$ over the out-of-sample period. Table 3 gives the MSE for GARCH(1,1) and the difference in MSE with respect to GARCH(1,1) for the other models. The standard errors are the heteroskedasticity and autocorrelation consistent standard errors from a regression of $(v_{t,h} - \widehat{V}_{t-1}\{s_{t,h}\})^2$ obtained from GARCH(1,1) (or the difference with respect to GARCH(1,1) for other models) on a constant; these standard errors need no further correction for estimation uncertainty (West (1996)).

Table 3 shows that in 11 out of 12 cases our regime-switching GARCH forecasts are better (lower MSE) than those from single-regime GARCH. Moreover, in 6 cases the outperformance is significant. Hence, also for the MSE criterion regime-switching GARCH improves on single-regime GARCH in terms of volatility forecasting. This does, of course, not mean that single-regime GARCH forecasts are bad. After all, Andersen and Bollerslev (1998) show that GARCH(1,1) yields good volatility forecast. We only conclude that regime-switching GARCH forecasts are better.

For the two variants of regime-switching ARCH we find an improvement over single-regime GARCH in 15 out of 24 cases (4 significant). These improvements are all for the mark and yen, as for the pound GARCH is the best in all 8 cases (1 significant). This is partly in line with our conclusion from the in-sample fit in subsection 3.2 that

GARCH gives a better fit for the pound and regime-switching ARCH yields a better fit for the yen; for the mark the GARCH fit is better but the forecasts are worse. Hence, as in subsection 3.2.4, both regimes and GARCH effects can be important to model volatility, which is another argument for using regime-switching GARCH.

Table 3 also shows that there is some preference of our model over Gray's (1996a) variant of regime-switching GARCH for the one-day horizon, as the MSE for our model is lower in 5 out of 6 cases. For the ten-day horizon we cannot make such a comparison, since we are unable to forecast more than one day ahead with Gray's model.

A final means to compare the volatility forecasts is to analyze the coefficient of determination, R^2, of the forecast efficiency regression (13). The standard R^2 measures the explanatory quality of a linear combination, $\gamma_0 + \gamma_1 \widehat{V}_{t-1}\{s_{t,h}\}$, of the forecast. However, one is interested in the quality of the forecast itself, not a linear combination of it. Therefore, we prefer the R^2 under the restriction $\gamma_0 = 0$ and $\gamma_1 = 1$,

$$\widetilde{R}^2 = 1 - \frac{V\left\{v_{t,h} - \widehat{V}_{t-1}\{s_{t,h}\}\right\}}{V\{v_{t,h}\}}. \tag{14}$$

This forecasting statistic is similar to the R^2-type measure used by Gray (1996a). It is generally smaller than the standard (unrestricted) R^2 and it can be negative.

The values of $\widetilde{R}^2$ in table 3 confirm the conclusions obtained from the MSE above. In particular, our regime-switching GARCH forecasts are better (higher $\widetilde{R}^2$) than those from single-regime GARCH in 11 out of 12 cases. The average improvements are 0.007 for the 1-day horizon and 0.057 for the 10-day horizon.

These improvements may seem low. However, the average $\widetilde{R}^2$ are also low (0.034 for $h = 1$ and 0.099 for $h = 10$ for single-regime GARCH). This does not mean that the forecasts are bad, as Andersen and Bollerslev (1998) show. The primary reason for the low R^2 (and thus low $\widetilde{R}^2$ and low improvements in $\widetilde{R}^2$) is the noise in the observed volatility measure $v_{t,h}$. As discussed above (12), this noise can be reduced by taking the sum of squared changes over smaller subperiods. To give an indication of the magnitude of the effect of this noise reduction on R^2, Andersen and Bollerslev compute the R^2 for a GARCH(1,1) model on daily mark/dollar and yen/dollar exchange rates both using a single squared daily change and using the sum of 288 squared five-minute changes in a day. The R^2 increases from on average 0.036 to 0.436. They conclude that GARCH does provide good volatility forecasts despite the low R^2 that one typically obtains using a single squared change. (Note that their argument also explains why for our series $\widetilde{R}^2$ is higher for the ten-day than for the one-day horizon, as we have reduced the noise in the ten-day observed volatility by using ten instead of one squared changes.)

To put the $\widetilde{R}^2$ improvements given above into perspective, we divide them by the average $\widetilde{R}^2$. We conclude that the relative outperformance of regime-switching GARCH over single-regime GARCH is 22% ($h = 1$) and 58% ($h = 10$), which is quite substantial.

4 Conclusion

This paper is based on the observation that forecasts from the widely-used GARCH model are significantly too high in volatile periods. We argue that this is due to the well-known high degree of persistence of individual shocks in volatility according to GARCH estimates. Therefore, we develop a model with more flexibility regarding volatility persistence, that is, not all shocks have to be highly persistent. The model generalizes GARCH by using two regimes with different levels of volatility and regime-specific GARCH formulas to describe the variance within the regimes. It is thus a regime-switching GARCH model.

Its specification shares the attractive features of Gray's (1996a) version of regime-switching GARCH, such as the recursiveness of the likelihood function. Our model, however, is preferable in other respects. For instance, multi-period-ahead volatility forecasting is a recursive procedure similar to standard GARCH and is much more convenient than with Gray's variant. Moreover, our model makes better use of the conditioning information to integrate out the unobserved regimes, which translates into a better fit.

In the empirical part of the paper we estimate the model using about twenty years of daily data on three U.S. dollar exchange rates (British pound, German mark and Japanese yen). The out-of-sample study shows that the problem of excessive single-regime GARCH forecasts in volatile periods disappears when using regime-switching GARCH. Moreover, regime-switching GARCH yields significantly better volatility forecasts than single-regime GARCH. We quantify the relative outperformance by 22% for the one-day and 58% for the ten-day horizon.

The other empirical results can be summarized as follows. First, including regimes is important, since even a regime-switching model with constant regime-specific variances sometimes outperforms standard GARCH. Second, there is conditional heteroskedasticity within regimes. Thus there is a need for ARCH or GARCH terms. Third, the heteroskedasticity differs across regimes, as more ARCH terms are needed in high-volatility regimes. This is in contrast with the regime-switching ARCH models of Cai (1994) and Hamilton and Susmel (1994). Fourth, GARCH terms, which are not allowed in regime-switching ARCH models, are important, because even single-regime GARCH is sometimes better than regime-switching models with several ARCH terms. Finally,

a t-distribution instead of a normal one for the error term helps make the regimes more stable.

There are a number of other possible applications of the model. For example, the proposed technique of averaging out unobserved regimes to avoid path-dependence of the likelihood function may also be useful in models that combine switches in the mean with a GARCH variance specification (see Klaassen (1999)). Moreover, regime-switching GARCH volatility forecasts can be used to analyze the effect of volatility on stock returns and to price options, for which volatility assessments are crucial.

Our results also yield several methodological suggestions for future research. Regime-switching GARCH generalizes standard GARCH by making the volatility persistence of shocks more flexible. Of course, it does so in a specific way, which is presumably not the optimal one. Moreover, we have not tried to rationalize the timing of the estimated regime switches in our data from an economic point of view, for instance, in terms of financial market liberalization, changes in exchange rate policies, or oil shocks. The regimes are only used as a technical means to obtain more flexibility regarding volatility persistence, just as the standard GARCH model is only a technical means to model volatility persistence. This appears, nevertheless, worthwhile, as we do find that the regime-switching GARCH forecasts are better than the GARCH ones. Hence, the paper suggests that it is promising to study volatility persistence in more detail, including the economic mechanisms behind it, to improve volatility forecasts even further.

Appendices

A Volatility Forecasting

In this appendix we give an expression for $p_{t-1}(r_\tau)$ in the volatility forecasting formula (8) and prove the recursive formula (9).

For the future regime probability in (8) we have

$$p_{t-1}(r_\tau) = \sum_{r_{t-1}=1,2} p_{t-1}(r_{t-1}) \cdot p_{t-1}(r_\tau \,|\, r_{t-1}), \tag{15}$$

where $p_{t-1}(r_{t-1})$ is discussed in (31). For the multi-period-ahead probability on the right-hand-side of (15), we form the time-constant Markov transition matrix M:

$$M = \begin{bmatrix} p_{11} & 1-p_{22} \\ 1-p_{11} & p_{22} \end{bmatrix}. \tag{16}$$

Using the $(\tau - (t-1))$-th power of M, the theory of Markov processes states that

$$p_{t-1}(r_\tau \,|\, r_{t-1}) = \left(M^{\tau-(t-1)}\right)_{r_\tau r_{t-1}}, \tag{17}$$

so that (15) can be computed.

In the remaining part of this appendix we prove (9). More precisely, we derive the special case $i = \tau - t$ of (9),

$$V_{t-1}\{\varepsilon_\tau \,|\, r_\tau\} = \omega_{r_\tau} + (\alpha_{r_\tau} + \beta_{r_\tau}) E_{t-1}\left\{ V_{t-1}\{\varepsilon_{\tau-1} \,|\, r_{\tau-1}\} | \, r_\tau \right\}, \tag{18}$$

as for the other i the derivation is analogous. The formula can be proved by repeatedly using the law of iterated expectations. Using definition (5), we get

$$\begin{aligned} V_{t-1}\{\varepsilon_\tau \,|\, r_\tau\} &= E_{t-1}\left[V_{\tau-1}\{\varepsilon_\tau \,|\, r_\tau\} \,|\, r_\tau \right] \\ &= E_{t-1}\left[\omega_{r_\tau} + \alpha_{r_\tau}\varepsilon_{\tau-1}^2 + \beta_{r_\tau} E_{\tau-1}\left\{ V_{\tau-2}\{\varepsilon_{\tau-1} \,|\, r_{\tau-1}\} | \, r_\tau \right\} \,|\, r_\tau \right]. \end{aligned} \tag{19}$$

For the ARCH part

$$\begin{aligned} E_{t-1}[\varepsilon_{\tau-1}^2 \,|\, r_\tau] &= E\{\varepsilon_{\tau-1}^2 \,|\, r_\tau, I_{t-1}\} \\ &= E[E\{\varepsilon_{\tau-1}^2 \,|\, r_{\tau-1}, r_\tau, I_{t-1}\} \,|\, r_\tau, I_{t-1}] \\ &= E_{t-1}[E_{t-1}\{\varepsilon_{\tau-1}^2 \,|\, r_{\tau-1}\} \,|\, r_\tau], \end{aligned} \tag{20}$$

where the last equality uses that the error distribution given the contemporaneous variance regime does not depend on the future variance regime.

For the GARCH part in (19) we use similar techniques to obtain

$$\begin{aligned}
&E_{t-1}[E_{\tau-1}(V_{\tau-2}\{\varepsilon_{\tau-1}\,|r_{\tau-1}\}\,|r_\tau)\,|r_\tau] \\
&\qquad = E[E(V\{\varepsilon_{\tau-1}\,|r_{\tau-1}, I_{\tau-2}\}\,|r_\tau, I_{\tau-1})\,|r_\tau, I_{t-1}] \\
&\qquad = E(V\{\varepsilon_{\tau-1}\,|r_{\tau-1}, I_{\tau-2}\}\,|r_\tau, I_{t-1}) \\
&\qquad = E[E(V\{\varepsilon_{\tau-1}\,|r_{\tau-1}, I_{\tau-2}\}\,|r_\tau, r_{\tau-1}, I_{t-1})\,|r_\tau, I_{t-1}] \\
&\qquad = E[V\{\varepsilon_{\tau-1}\,|r_{\tau-1}, I_{t-1}\}\,|r_\tau, I_{t-1}] \\
&\qquad = E_{t-1}[V_{t-1}\{\varepsilon_{\tau-1}\,|r_{\tau-1}\}\,|r_\tau]. \qquad (21)
\end{aligned}$$

The penultimate equality uses that $I_{\tau-2}$ given $r_\tau, r_{\tau-1}, I_{t-1}$ is independent of r_τ, since the Markov structure implies that the distribution of variance regimes $(r_{\tau-2}, r_{\tau-3}, ...)$ conditional on $r_{\tau-1}$ and r_τ is independent of r_τ; this makes the changes $(s_{\tau-2}, s_{\tau-3}, ...)$ also independent of r_τ once $r_{\tau-1}$ is given.

Substituting the results for the ARCH and GARCH parts in (19) gives formula (18). The required probability in (18) is

$$p_{t-1}(r_{\tau-1}\,|r_\tau) = \frac{p_{t-1}(r_\tau\,|r_{\tau-1})\cdot p_{t-1}(r_{\tau-1})}{p_{t-1}(r_\tau)}, \qquad (22)$$

where the switching probability follows from (2) and the regime probability $p_{t-1}(r_{\tau-1})$ follows in a similar way as $p_{t-1}(r_\tau)$ in (15); the denominator is given by (15).

B Unconditional Error Variance

Here we derive expression (10) for the "unconditional" error variance $V\{\varepsilon_t\,|r_t\}$ and the three necessary conditions for its existence.

Suppose $V\{\varepsilon_t\,|\,r_t\}$ exists for both $r_t = 1, 2$ and for all ω_1, ω_2. Using the variance definition (5), repeated use of the law of iterated expectations yields

$$\begin{aligned}
&V\{\varepsilon_t\,|r_t\} \\
&= \omega_{r_t} + \alpha_{r_t} E\{\varepsilon_{t-1}^2\,|r_t\} + \beta_{r_t} E\{V_{t-2}\{\varepsilon_{t-1}\,|r_{t-1}\}\,|r_t\} \\
&= \omega_{r_t} + \alpha_{r_t} E\{E\{\varepsilon_{t-1}^2\,|r_{t-1}, r_t\}\,|r_t\} + \beta_{r_t} E\{E\{V_{t-2}\{\varepsilon_{t-1}\,|r_{t-1}\}\,|r_{t-1}, r_t\}\,|r_t\} \\
&= \omega_{r_t} + \alpha_{r_t} E\{V\{\varepsilon_{t-1}\,|r_{t-1}\}\,|r_t\} + \beta_{r_t} E\{V\{\varepsilon_{t-1}\,|r_{t-1}\}|\,r_t\} \\
&= \omega_{r_t} + (\alpha_{r_t} + \beta_{r_t})\cdot E\{V\{\varepsilon_{t-1}\,|r_{t-1}\}\,|r_t\}, \qquad (23)
\end{aligned}$$

where the penultimate equality uses that the distribution of the error given the contemporaneous variance regime does not depend on the future variance regime.

Next, assume that $V\{\varepsilon_t \,|\, r_t = 1\}$ and $V\{\varepsilon_t \,|\, r_t = 2\}$ do not depend on t and denote them by σ_1^2 and σ_2^2, respectively. Then

$$\begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \end{bmatrix}, \tag{24}$$

where $A_{ij} = P\{r_{t-1} = j \,|\, r_t = i\}(\alpha_i + \beta_i)$; expressions for the probabilities involved are at the end of this appendix.

Let A be the matrix with elements A_{ij}. Since we have assumed that both unconditional variances exist for all ω_1, ω_2, the matrix $I_2 - A$ is invertible, so that the unconditional variances are indeed given by (10).

Necessary conditions for the existence of both variances can be derived as follows. Because $\alpha_{r_t}, \beta_{r_t} \geqslant 0$ implies that the variances are strictly positive for all $\omega_1, \omega_2 (> 0)$ the four elements of $(I_2 - A)^{-1}$ in (10) must be nonnegative and $(I_2 - A)^{-1}$ may not have a zero row. The four elements follow from

$$(I_2 - A)^{-1} = \frac{1}{\det(I_2 - A)} \begin{bmatrix} 1 - A_{22} & \alpha_1 + \beta_1 - A_{11} \\ \alpha_2 + \beta_2 - A_{22} & 1 - A_{11} \end{bmatrix}, \tag{25}$$

where we have used $A_{12} = \alpha_1 + \beta_1 - A_{11}$ and $A_{21} = \alpha_2 + \beta_2 - A_{22}$. Since $\alpha_i + \beta_i - A_{ii} \geqslant 0$ for both regimes $i = 1, 2$, the nonnegativity of $(I_2 - A)^{-1}$ implies through (25) that $\det(I_2 - A) > 0$, so that $1 - A_{11} \geqslant 0$ and $1 - A_{22} \geqslant 0$. However, neither A_{11} nor A_{22} may be unity; otherwise $\det(I_2 - A) = -(\alpha_1 + \beta_1 - A_{11})(\alpha_2 + \beta_2 - A_{22})$, so that $\alpha_i + \beta_i - A_{ii} \geqslant 0$ for both regimes would imply that $\det(I_2 - A) \leqslant 0$, which is not the case. Hence, the three necessary conditions are $A_{11}, A_{22} < 1$ and $\det(I_2 - A) > 0$.

To compute the unconditional error variance in (24), we need the probability $p(r_{t-1} \,|\, r_t)$ that the previous regime was r_{t-1} given that the current regime is r_t. Using Bayes' rule, we have

$$p(r_{t-1} \,|\, r_t) = \frac{p(r_t \,|\, r_{t-1}) \cdot p(r_{t-1})}{\sum_{r_{t-1}=1,2} p(r_t \,|\, r_{t-1}) \cdot p(r_{t-1})}, \tag{26}$$

where $p(r_t \mid r_{t-1})$ is constant (see (2)) and the theory of Markov processes gives the unconditional probabilities (see Hamilton (1989)):

$$\begin{aligned} p(r_{t-1} = 1) &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \\ p(r_{t-1} = 2) &= \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \end{aligned} \tag{27}$$

C Estimation

In this appendix we derive the likelihood function of the regime-switching GARCH model and show that it has a first-order recursive structure, as claimed in subsection 2.2.4.

To obtain the likelihood function, we first need the density of the exchange rate change at time t conditional on only observable information. Let $p_{t-1}(s_t)$ denote this density evaluated at an exchange rate change equal to s_t. It can be split up as

$$p_{t-1}(s_t) = \sum_{r_t=1,2} p_{t-1}(s_t \,|\, r_t) \cdot p_{t-1}(r_t). \tag{28}$$

The first term on the right-hand-side, $p_{t-1}(s_t \,|\, r_t)$, denotes the density of the exchange rate change at time t evaluated at the value s_t conditional on I_{t-1} and on the regime having value r_t. This t-density follows from formulas (1), (5) and (6). It is, however, not straightforward how to compute the conditional variance in (5), as this requires integrating out the regime path $\widetilde{r}_{t-1}$ in $E_{t-1}[V_{t-2}\{\varepsilon_{t-1} \mid \widetilde{r}_{t-1}\}|r_t]$. Because $V_{t-2}\{\varepsilon_{t-1} \,|\, \widetilde{r}_{t-1}\}$ depends only on r_{t-1}, we just need $p_{t-1}(r_{t-1}|r_t)$, the probability that the previous regime was r_{t-1} given that the current regime is r_t and given the information I_{t-1}:

$$p_{t-1}(r_{t-1} \,|\, r_t) = \frac{p_{t-1}(r_{t-1}) \cdot p_{t-1}(r_t \,|\, r_{t-1})}{p_{t-1}(r_t)}, \tag{29}$$

where

$$p_{t-1}(r_t) = \sum_{r_{t-1}=1,2} p_{t-1}(r_{t-1}) \cdot p_{t-1}(r_t \,|\, r_{t-1}). \tag{30}$$

The constant switching probability $p_{t-1}(r_t \,|\, r_{t-1})$ follows from (2).

The remaining term in (29) and (30) is $p_{t-1}(r_{t-1})$. This probability is crucial, since all regime probabilities in the paper can be derived from it. Using similar techniques as in Gray (1996a), the following formula shows that this probability has a first-order recursive structure, which simplifies its computation substantially:

$$\begin{aligned} p_{t-1}(r_{t-1}) &= p_{t-2}(r_{t-1} \,|\, s_{t-1}) \\ &= \frac{p_{t-2}(s_{t-1} \,|\, r_{t-1}) \cdot p_{t-2}(r_{t-1})}{p_{t-2}(s_{t-1})} \\ &= \frac{p_{t-2}(s_{t-1} \,|\, r_{t-1}) \cdot \sum_{r_{t-2}=1,2} p_{t-2}(r_{t-2}) \cdot p_{t-2}(r_{t-1} \,|\, r_{t-2})}{p_{t-2}(s_{t-1})}. \end{aligned} \tag{31}$$

Hence, the variables to compute $p_{t-1}(r_{t-1})$ are its previous values $p_{t-2}(r_{t-2})$ and the constant $p_{t-2}(r_{t-1}|\, r_{t-2})$ for $r_{t-2} = 1, 2$ and the previous densities $p_{t-2}(s_{t-1} \,|\, r_{t-1})$ and $p_{t-2}(s_{t-1})$. This makes the computation of $p_{t-1}(r_{t-1})$ a first-order recursive process.

The second term on the right-hand-side of (28), $p_{t-1}(r_t)$, is the conditional probability that the current regime is r_t. It is given by (30).

Having discussed both terms on the right-hand-side of (28), we can now compute the density of interest, $p_{t-1}(s_t)$, being a mixture of two t-densities. This density can then be used to build the sample log-likelihood $\sum_{t=1}^{T} \log(p_{t-1}(s_t))$ with which all parameters in the regime-switching GARCH model can be estimated.

From a practical point of view, it is important to realize that the log-likelihood has a first-order recursive structure, similar to that of a standard, one-regime GARCH(1,1) model. After all, for (29) and (30) one needs the constant $p_{t-1}(r_t \mid r_{t-1})$ and the first-order recursive probability $p_{t-1}(r_{t-1})$ in (31) for all four combinations of (r_t, r_{t-1}); density (28) can then be computed from (30), the previous change s_{t-1}, (29) and the previous variances $V_{t-2}\{\varepsilon_{t-1}|r_{t-1}\}$ for $r_{t-1}=1,2$. This first-order recursiveness of $p_{t-1}(s_t)$ speeds up the calculation of the sample log-likelihood substantially. To start up the recursive process, we set the required variables equal to their expectation without conditioning on the information set, that is, the "unconditional" variance $\sigma^2_{r_t}$ in (10).

D Regime Inference

As stated in subsection 2.2.5, the smoothed probability that the regime was r_t at time t, $p_T(r_t)$, can be computed recursively. More generally, any ex post regime probability $p_\tau(r_t)$, for a given future time $\tau \in \{t, t+1, \ldots, T\}$, can be computed in a recursive manner, starting from the ex ante probability $p_{t-1}(r_t)$. In this appendix, we verify that claim.

One can write $p_\tau(r_t)$ for both regimes $r_t=1,2$ as

$$\begin{aligned} p_\tau(r_t) &= p_{\tau-1}(r_t|s_\tau) \\ &= \frac{p_{\tau-1}(s_\tau|r_t)\cdot p_{\tau-1}(r_t)}{\sum_{r_t=1,2} p_{\tau-1}(s_\tau|r_t)\cdot p_{\tau-1}(r_t)}. \end{aligned} \tag{32}$$

Suppose first that $\tau=t$. Then $p_\tau(r_t)$ follows directly, as $p_{\tau-1}(r_t)$ and $p_{\tau-1}(s_\tau|r_t)$ in (32) are known from the estimation process (see appendix C).

Hence, let us suppose from now on that $\tau > t$. The computation of (32) requires two inputs. The first is the previous ex post probability $p_{\tau-1}(r_t)$, which is known from the previous recursion for both r_t. The second ingredient of (32) is the density $p_{\tau-1}(s_\tau|r_t)$ for both regime outcomes. Its computation requires a number of steps. We first write it as

$$p_{\tau-1}(s_\tau|r_t) = \sum_{r_\tau=1,2} p_{\tau-1}(s_\tau|r_\tau)\cdot p_{\tau-1}(r_\tau|r_t), \tag{33}$$

where we use that the conditional distribution of s_τ given r_τ does not depend on the earlier regime r_t. This formula itself has two ingredients. The first one is the density $p_{\tau-1}(s_\tau|r_\tau)$ for both regime combinations, which is known from the estimation process.

The second term needed in (33) is the $(\tau-t)$-period-ahead regime-switching probability $p_{\tau-1}(r_\tau|r_t)$ for all regime outcomes. Once it has been computed, it should be saved, since it will be needed in the next recursive step. Making use of the Markov structure of the regime process, it can be written in terms of $(\tau-1-t)$-period-ahead switching probabilities

$$p_{\tau-1}(r_\tau|r_t) = \sum_{r_{\tau-1}=1,2} p_{\tau-1}(r_\tau|r_{\tau-1}) \cdot p_{\tau-1}(r_{\tau-1}|r_t). \tag{34}$$

Again there are two ingredients. First, we need $p_{\tau-1}(r_\tau|r_{\tau-1})$ for all regime combinations. These are constant and follow from (2).

The second ingredient of (34) is $p_{\tau-1}(r_{\tau-1}|r_t)$ for all regime combinations. We have

$$\begin{aligned} p_{\tau-1}(r_{\tau-1}|r_t) &= p_{\tau-2}(r_{\tau-1}|r_t, s_{\tau-1}) \\ &= \frac{p_{\tau-2}(s_{\tau-1}|r_{\tau-1}) \cdot p_{\tau-2}(r_{\tau-1}|r_t)}{\sum_{r_{\tau-1}=1,2} p_{\tau-2}(s_{\tau-1}|r_{\tau-1}) \cdot p_{\tau-2}(r_{\tau-1}|r_t)}, \end{aligned} \tag{35}$$

where we use that the conditional density of $s_{\tau-1}$ is independent of the earlier regime r_t once $r_{\tau-1}$ is given. We have two ingredients. First, the conditional density $p_{\tau-2}(s_{\tau-1}|r_{\tau-1})$ for both regime outcomes. It is known from the estimation process. Secondly, we need the $(\tau-1-t)$-period-ahead switching probability $p_{\tau-2}(r_{\tau-1}|r_t)$ for all regime combinations. This one was saved during the previous recursion, if $\tau > t+1$. If $\tau = t+1$, it equals one.

This completes the algorithm to compute (33), which is the second ingredient of (32). For each recursion one has to compute (35), use the result to compute (34) and use this to compute (33). Using this in (32) yields the ex post probability $p_\tau(r_t)$ from $p_{\tau-1}(r_t)$. Therefore, starting from the ex ante probability $p_{t-1}(r_t)$ one can recursively compute the ex post probability $p_\tau(r_t)$.

References

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

Ang, A. and G. Bekaert (1998), "Regime Switches in Interest Rates," NBER Working Paper No. 6508.

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

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5-59.

Cai, J. (1994), "A Markov Model of Unconditional Variance in ARCH," *Journal of Business and Economic Statistics*, 12, 309-316.

Chaudhuri, K. and F. Klaassen (2000), "Have East Asian Stock Markets Calmed Down? Evidence from a Regime Switching Model," mimeo, University of Amsterdam.

Dewachter, H. (1997), "Sign Predictions of Exchange Rate Changes: Charts as Proxies for Bayesian Inferences," *Weltwirtschaftliches Archiv*, 133, 39-55.

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

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

Fong, W.M. (1998), "The Dynamics of DM/£ Exchange Rate Volatility: A SWARCH Analysis," *International Journal of Finance and Economics*, 3, 59-71.

Gray, S.F. (1996a), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27-62.

Gray, S.F. (1996b), "An Analysis of Conditional Regime-Switching Models," Working Paper, Fuqua School of Business, Duke University.

Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Hamilton, J.D. and R. Susmel (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307-333.

Hansen, B.E. (1992), "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61-S82.

Kaufmann, S. and M. Scheicher (1996), "Markov-Regime Switching in Economic Variables: Part I. Modelling, Estimating and Testing. – Part II. A Selective Survey," Institute for Advanced Studies, University of Vienna, Economics Series No. 38.

Klaassen, F.J.G.M (1999), "Long Swings in Exchange Rates: Are They Really in the Data?" CentER for Economic Research, Tilburg University, Discussion Paper No. 9908.

Lamoureux, C.G. and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225-234.

MacDonald, R. and M.P. Taylor (1992), "Exchange Rate Economics: A Survey," *IMF Staff Papers*, 39, 1-57.

Meese, R.A. and K. Rogoff (1983), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14, 3-24.

Merton, R.C. (1980), "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323-361.

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Newey, W.K. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.

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

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.

Scheicher, M. (1999), "Nonlinear Dynamics: Evidence for a Small Stock Exchange," *Empirical Economics*, 24, 45-59.

Schwert, G.W. (1989), "Why does Stock Market Volatility Change over Time?" *Journal of Finance*, 44, 1115-1153.

Timmermann, A. (2000), "Moments of Markov Switching Models," *Journal of Econometrics*, 96, 75-111.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

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

West, K.D. and M.W. McCracken (1998), "Regression-Based Tests of Predictive Ability," *International Economic Review*, 39, 817-840.

Table 1: Estimation results

			Low var. regime 1				High variance regime 2							logLik
	μ	ν^{-1}	σ_1^2	α_{11}	α_{21}	p_{11}	σ_2^2	α_{12}	α_{22}	α_{32}	α_{42}	β_2	p_{22}	-G(1,1)
BRITISH POUND														
Const. variance	.01	.25					.46							–244.34
	(.01)	(.02)					(.02)							
GARCH(1,1)	.01	.19					.61	.06				.93		-4437.26
	(.01)	(.01)					(.18)	(.01)				(.01)		
ARCH(0;0)	.01	.17	.18			.984	.65						.987	–46.53
	(.01)	(.02)	(.01)			(.004)	(.04)						(.004)	
ARCH(2;4)	.01	.19	.16	.11	.11	.991	.58	.07	.04	.07	.07		.996	–19.56
	(.01)	(.02)	(.01)	(.04)	(.04)	(.003)	(.03)	(.02)	(.02)	(.03)	(.02)		(.002)	
GrayG(0,2;1,1)	.01	.19	.18	.21	.18	.999	.74	.06				.93	.999	+9.24
	(.01)	(.01)	(.03)	(.06)	(.06)	(.001)	(.23)	(.01)				(.01)	(.000)	
GARCH(0,2;1,1)	.01	.18	.15	.17	.10	.989	.58	.05				.93	.997	+19.34
	(.01)	(.02)	(.02)	(.07)	(.05)	(.007)	(.09)	(.01)				(.01)	(.002)	
GERMAN MARK														
Const. variance	-.00	.25					.50							–183.92
	(.01)	(.02)					(.02)							
GARCH(1,1)	-.01	.20					.70	.08				.91		-4778.34
	(.01)	(.01)					(.20)	(.01)				(.01)		
ARCH(0;0)	-.00	.15	.23			.984	.77						.982	–24.85
	(.01)	(.02)	(.02)			(.004)	(.04)						(.005)	
ARCH(0;4)	-.00	.16	.21			.983	.70	.06	.01	.07	.05		.987	–14.97
	(.01)	(.02)	(.02)			(.004)	(.05)	(.03)	(.02)	(.03)	(.02)		(.004)	
GrayG(0,0;1,1)	-.00	.20	.11			.998	.70	.08				.89	.999	–3.29
	(.01)	(.01)	(.01)			(.001)	(.11)	(.01)				(.01)	(.000)	
GARCH(0,0;1,1)	-.01	.19	.12			.981	.57	.07				.90	.998	+8.44
	(.01)	(.02)	(.01)			(.011)	(.07)	(.01)				(.02)	(.002)	
JAPANESE YEN														
Const. variance	-.01	.28					.50							–111.62
	(.01)	(.02)					(.03)							
GARCH(1,1)	-.01	.25					.61	.09				.87		-4682.58
	(.01)	(.02)					(.11)	(.02)				(.02)		
ARCH(0;0)	-.01	.20	.21			.975	.73						.975	+9.07
	(.01)	(.02)	(.02)			(.007)	(.04)						(.007)	
ARCH(1;4)	-.01	.20	.23	.08		.981	.71	.03	.04	.05	.00		.981	+16.28
	(.01)	(.02)	(.02)	(.04)		(.006)	(.05)	(.03)	(.03)	(.03)	(.02)		(.006)	
GrayG(0,1;1,1)	-.01	.21	.23	.10		.980	.80	.04				.48	.980	+12.59
	(.01)	(.02)	(.02)	(.04)		(.006)	(.07)	(.03)				(.17)	(.005)	
GARCH(0,1;1,1)	-.01	.21	.22	.09		.978	.68	.06				.78	.984	+16.52
	(.01)	(.02)	(.02)	(.04)		(.007)	(.05)	(.02)				(.10)	(.005)	

Standard errors in parentheses. "logLik-G(1,1)" denotes the log-likelihood of a model minus that of GARCH(1,1); for GARCH(1,1) it is the log-likelihood itself.
The estimated models belong to the class of models described by equations (1), (2), (5) and (6), except for GrayG(0,Q_1;1,1) (Gray's (1996a) variant of GARCH(0,Q_1;1,1)), which uses (4) instead of (5). The parameter μ denotes the conditional mean, ν^{-1} is the inverted degrees of freedom of the t-distribution for the innovation, σ_r^2 denotes the unconditional variance in regime r, α_{qr} and β_r are the regime specific ARCH and GARCH parameters, respectively, and the p_{rr} are the regime-staying probabilities. The estimated first-order autoregressive coefficient used for the pound only is 0.03 (0.01) for all models.

Table 2: Diagnostics for residual conditional heteroskedasticity

	BRITISH POUND		GERMAN MARK		JAPANESE YEN	
	ρ_1	Q_{10}	ρ_1	Q_{10}	ρ_1	Q_{10}
Const. variance	0.11∗	533.26∗	0.12∗	371.09∗	0.09∗	164.25∗
	(0.01)	[0.00]	(0.01)	[0.00]	(0.01)	[0.00]
GARCH(1,1)	0.01	5.29	0.00	8.02	0.02	13.00
	(0.01)	[0.87]	(0.01)	[0.63]	(0.01)	[0.22]
ARCH(0;0)	0.03∗	44.69∗	0.03∗	25.29∗	0.03∗	16.15
	(0.01)	[0.00]	(0.01)	[0.00]	(0.01)	[0.10]
ARCH(Q_1;4)	-0.01	26.15∗	0.00	25.59∗	0.01	8.29
	(0.01)	[0.00]	(0.01)	[0.00]	(0.01)	[0.60]
GrayG(0,Q_1;1,1)	-0.00	6.30	0.01	27.08∗	0.00	16.67
	(0.01)	[0.79]	(0.01)	[0.00]	(0.01)	[0.08]
GARCH(0,Q_1;1,1)	-0.00	6.47	0.00	9.33	0.01	8.87
	(0.01)	[0.77]	(0.01)	[0.50]	(0.01)	[0.54]

Standard errors in parentheses and p-values in square brackets; * is significant at the 5% level.
The first-order autocorrelation, ρ_1, and the Box-Pierce statistic of order ten, Q_{10}, are computed from the squared normalized residuals. Note that the normalization of residuals under a regime-switching model entails integrating out the unobserved regime in the variance, as in (8).
The specifications of the models are given in the notes below table 1.

Table 3: Out-of-sample volatility forecasting statistics

	One-day forecast horizon								Ten-day forecast horizon							
	MSE-G(1,1)		γ_0		γ_1		$\widetilde{R}^2$		MSE-G(1,1)		γ_0		γ_1		$\widetilde{R}^2$	
BRITISH POUND																
Const. variance	.051∗	.058∗			.88	.98	0	0	4.00∗	4.66			.89	.98	0	0
	(.017)	(.028)			(.10)	(.12)			(1.24)	(3.30)			(.11)	(.13)		
GARCH(1,1)	.697	.998	.05	.08	.84	.78	.062	.058	11.20	16.39	.55	1.03	.82	.72	.256	.226
	(.112)	(.262)	(.04)	(.06)	(.11)	(.16)			(2.55)	(4.44)	(.39)	(.59)	(.10)	(.16)		
ARCH(0;0)	.011	.024	-.04	-.02	1.06	1.04	.048	.030	0.84	2.10	-.77	-.02	1.14	1.00	.198	.121
	(.009)	(.024)	(.06)	(.07)	(.19)	(.24)			(0.70)	(2.54)	(.80)	(.90)	(.23)	(.29)		
ARCH(2;4)	.009∗	.015	.00	-.02	.95	1.01	.050	.043	0.61	1.71	-1.00	-.52	1.18	1.09	.214	.139
	(.004)	(.015)	(.06)	(.08)	(.17)	(.22)			(0.62)	(2.14)	(.77)	(1.19)	(.23)	(.34)		
GrayG(0,2;1,1)	.002	-.002	.03	.05	.89	.85	.059	.059	–	–	–	–	–	–	–	–
	(.002)	(.007)	(.05)	(.06)	(.13)	(.15)										
GARCH(0,2;1,1)	.001	-.004	.02	.07	.90	.80	.060	.061	-0.03	-0.56	.22	.79	.90	.76	.259	.253
	(.002)	(.003)	(.05)	(.06)	(.14)	(.14)			(0.27)	(0.34)	(.43)	(.58)	(.12)	(.15)		
GERMAN MARK																
Const. variance	.036∗	.033∗			.89	.98	0	0	1.87	0.42			.89	.98	0	0
	(.017)	(.016)			(.09)	(.11)			(1.47)	(1.91)			(.10)	(.11)		
GARCH(1,1)	.817	1.208	.12∗	.12∗	.68∗	.69∗	.040	.029	14.51	22.18	1.39∗	1.93∗	.61∗	.54∗	.122	.031
	(.121)	(.323)	(.05)	(.05)	(.11)	(.12)			(2.60)	(6.08)	(.56)	(.90)	(.11)	(.15)		
ARCH(0;0)	.000	-.013	.03	-.12	.87	1.31	.039	.038	-1.23	-2.77	-.10	-.76	.96	1.20	.178	.143
	(.009)	(.011)	(.07)	(.08)	(.18)	(.24)			(0.92)	(1.58)	(.87)	(.89)	(.22)	(.23)		
ARCH(0;4)	-.002	-.016	.04	-.01	.86	1.05	.042	.040	-1.31	-2.80	-.43	-.80	1.02	1.20	.184	.143
	(.007)	(.012)	(.08)	(.06)	(.18)	(.15)			(0.90)	(1.57)	(.85)	(.86)	(.21)	(.21)		
GrayG(0,0;1,1)	-.004∗	-.008	.09	.11	.73∗	.76	.046	.034	–	–	–	–	–	–	–	–
	(.002)	(.005)	(.05)	(.06)	(.11)	(.15)										
GARCH(0,0;1,1)	-.009∗	-.011∗	.06	.08	.82	.81	.050	.036	-1.48∗	-1.69∗	.43	1.44	.83	.68	.199	.095
	(.003)	(.004)	(.06)	(.06)	(.13)	(.15)			(0.44)	(0.60)	(.67)	(.94)	(.14)	(.18)		
JAPANESE YEN																
Const. variance	.005	.002			.87	.96	0	0	-0.14	-2.27			.87	.96	0	0
	(.014)	(.011)			(.09)	(.09)			(0.88)	(1.69)			(.10)	(.09)		
GARCH(1,1)	.996	1.414	.16∗	.19∗	.55∗	.56∗	.009	.005	17.35	23.59	1.19∗	3.17∗	.57∗	.30∗	.052	-.096
	(.157)	(.481)	(.05)	(.08)	(.10)	(.14)			(2.45)	(7.61)	(.61)	(1.08)	(.10)	(.17)		
ARCH(0;0)	-.022∗	-.021	-.07	.03	1.09	.90	.025	.018	-2.16∗	-3.08	-1.87	1.41	1.34	.68	.102	.038
	(.011)	(.014)	(.08)	(.10)	(.23)	(.23)			(0.71)	(1.65)	(1.25)	(.98)	(.32)	(.20)		
ARCH(1;4)	-.021∗	-.027	.06	.02	.82	.93	.024	.023	-2.29∗	-3.12	-.40	1.30	1.00	.70	.113	.041
	(.009)	(.016)	(.06)	(.10)	(.16)	(.23)			(0.64)	(1.67)	(.91)	(.98)	(.23)	(.20)		
GrayG(1,0;1,1)	-.021∗	-.022	.05	.05	.82	.86	.025	.019	–	–	–	–	–	–	–	–
	(.008)	(.017)	(.07)	(.12)	(.17)	(.27)										
GARCH(1,0;1,1)	-.024∗	-.010	.04	.15∗	.84	.65∗	.028	.011	-2.32∗	-1.44	-.46	2.72∗	1.01	.40∗	.117	.012
	(.008)	(.007)	(.07)	(.06)	(.17)	(.12)			(0.62)	(0.83)	(.90)	(.98)	(.22)	(.16)		

Standard errors in parentheses (details in section 3.3); * is significant at 5%.
"MSE-G(1,1)" is the mean squared forecast error of a model minus that of GARCH(1,1); for GARCH(1,1) it is the MSE itself; γ_0 and γ_1 are the intercept and slope in the forecast efficiency regression (13); $\widetilde{R}^2$ is the restricted R^2 defined by (14).
There are two estimates for each statistic. This reflects the two different out-of-sample periods: the left (right) estimate is based on forecasts for the second (first) half of the sample, that is, 2,491 days, using the first (second) half for estimation.
The specifications of the models are given in the notes below table 1. For Gray's (1996a) variant of GARCH(0,Q_1;1,1) we are unable to compute multi-day-ahead forecasts.

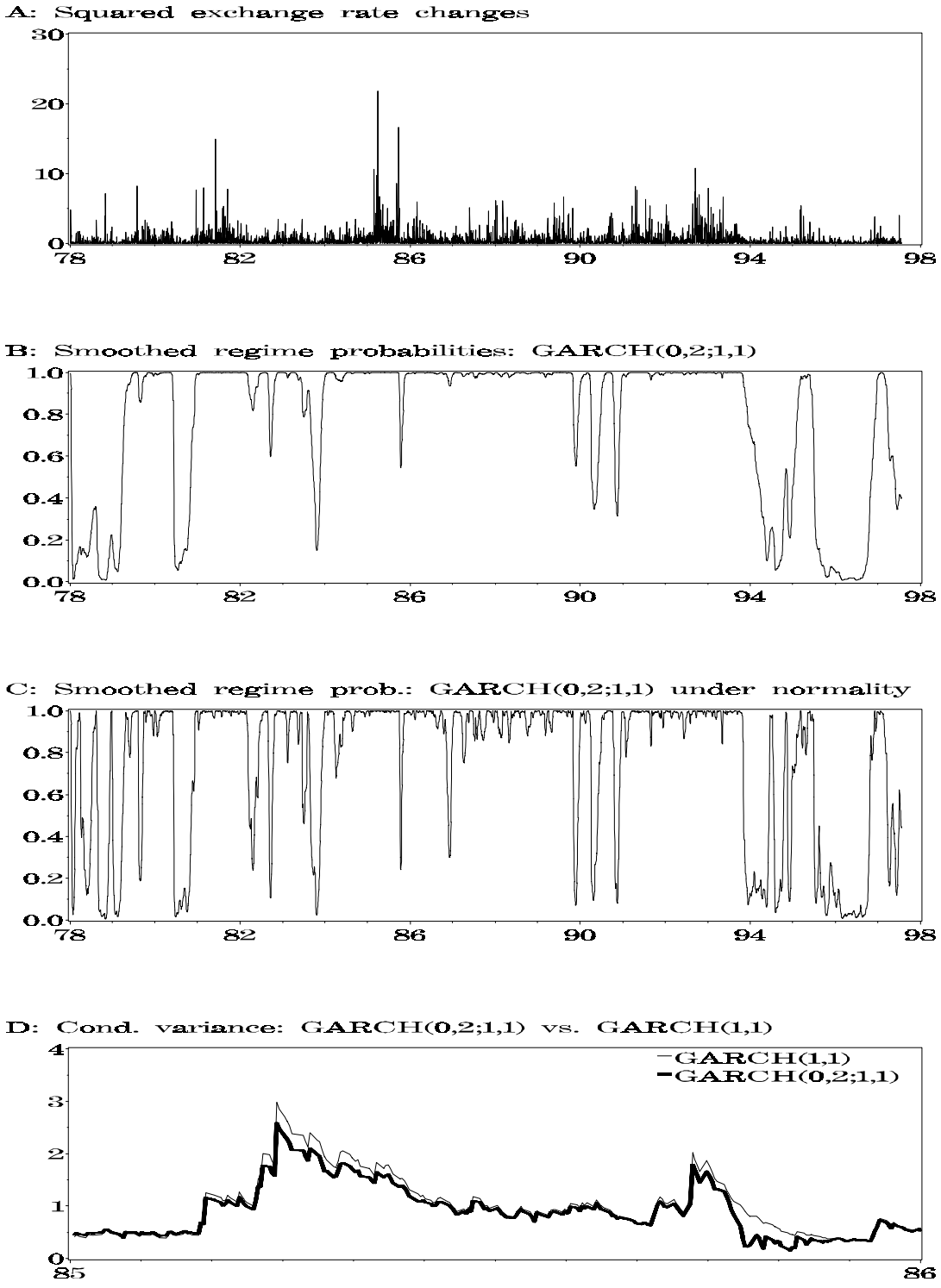

Figure 1: British pound

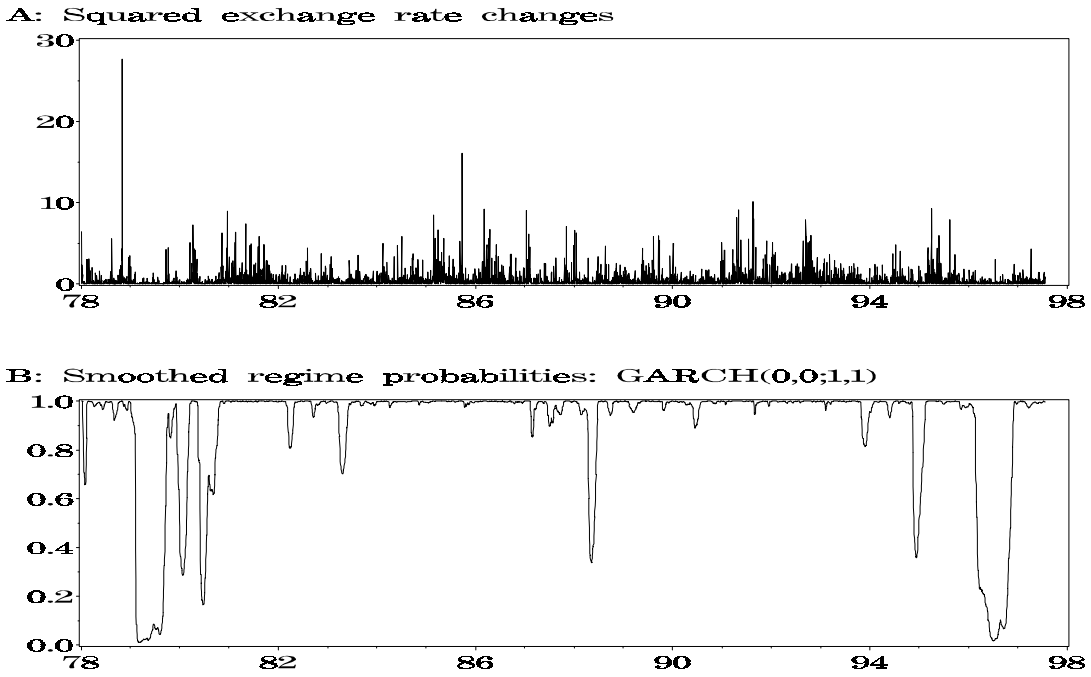

Figure 2: German mark

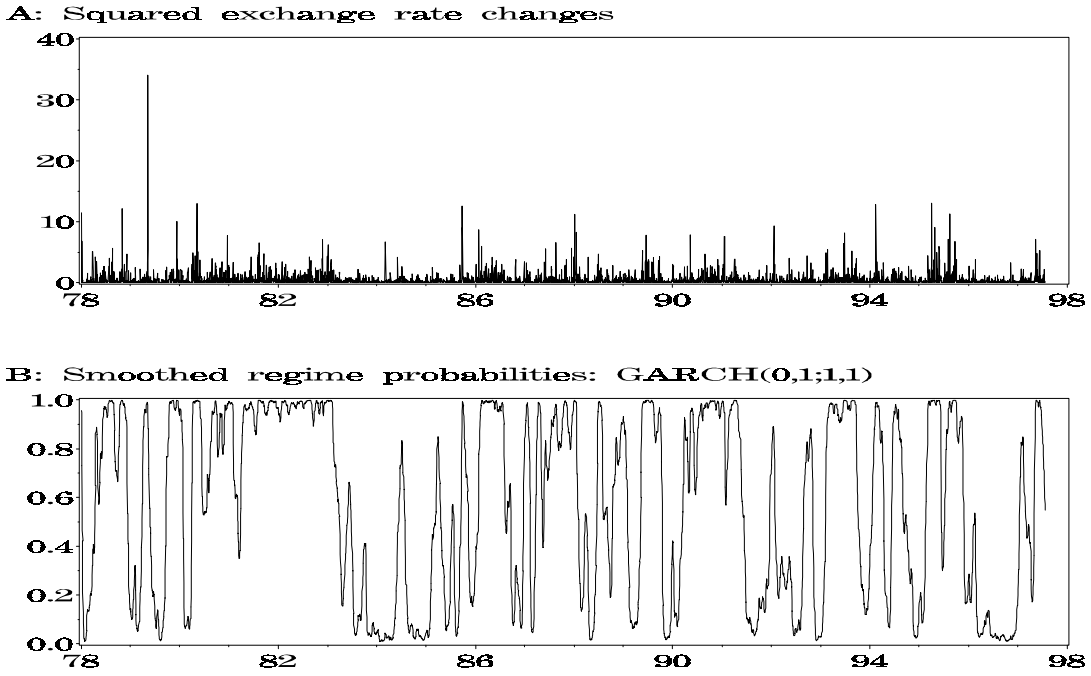

Figure 3: Japanese yen