Journal of Financial Econometrics, 2004, Vol. 2, No. 4, 493–530

A New Approach to Markov-Switching GARCH Models

Markus Haas
University of Munich

Stefan Mittnik
University of Munich, Center for Financial Studies, and Ifo Institute for Economic Research

Marc S. Paolella
University of Zurich

ABSTRACT

The use of Markov-switching models to capture the volatility dynamics of financial time series has grown considerably during past years, in part because they give rise to a plausible interpretation of nonlinearities. Nevertheless, GARCH-type models remain ubiquitous in order to allow for nonlinearities associated with time-varying volatility. Existing methods of combining the two approaches are unsatisfactory, as they either suffer from severe estimation difficulties or else their dynamic properties are not well understood. In this article we present a new Markov-switching GARCH model that overcomes both of these problems. Dynamic properties are derived and their implications for the volatility process discussed. We argue that the disaggregation of the variance process offered by the new model is more plausible than in the existing variants. The approach is illustrated with several exchange rate return series. The results suggest that a promising volatility model is an independent switching GARCH process with a possibly skewed conditional mixture density.

KEYWORDS: conditional volatility, density forecasting, empirical finance, exchange rates, nonlinear time series, regime switching

We thank Cord Brannolte, René Garcia (the editor), and three anonymous referees for constructive comments. The research of M. Haas and S. Mittnik was supported by the *Deutsche Forschungsgemeinschaft*. Part of the research of M. S. Paolella has been carried out within the National Center of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. Address correspondence to Stefan Mittnik, Chair of Financial Econometrics, Institute of Statistics, University of Munich, D-80799 Munich, Germany, or e-mail: finmetrics@stat.uni-muenchen.de.

doi:10.1093/jjfinec/nbh020

A Markov-switching model is a nonlinear specification in which different states of the world affect the evolution of a time series. The dynamic properties depend on the present regime, with the regimes being realizations of a hidden Markov chain with a finite state space. Markov-switching models were introduced to the econometric mainstream by Hamilton (1988, 1989) and continue to gain popularity—especially in financial time-series analysis. For example, Engel and Hamilton (1990), Engel (1994), Vigfusson (1997), Bollen, Gray, and Whaley (2000), Dewachter (2001), Klaassen (2002), Brunetti et al. (2003), and Beine, Laurent, and Lecourt (2003), among others, investigate regime switching in foreign exchange rates; Bollen, Gray, and Whaley (2000) argue that different exchange rate policy regimes give rise to different exchange rate behavior; and Vigfusson (1997) constructs a two-state Markov-switching model for exchange rate dynamics, based on the Frankel and Froot (1988) model of chartists and fundamentalists in the market, and identifies their corresponding regimes, using daily Canada–U.S. exchange rates. Ahrens and Reitz (2004) apply Vigfusson's empirical model to German-U.S. data. Turner, Startz, and Nelson (1989), Pagan and Schwert (1990), Hamilton and Susmel (1994), Dueker (1997), Ryden, Teräsvirta, and Åsbrink (1998), Billio and Pelizzon (2000), Maheu and McCurdy (2000), Susmel (2000), Perez-Quiros and Timmermann (2001), and Bhar and Hamori (2004), among others, employ Markov-switching models for the modeling of stock returns. In particular, Hamilton and Susmel (1994) distinguish a low-, moderate-, and high-volatility regime in weekly stock return data, with the high-volatility regime being associated with economic recessions, while Maheu and McCurdy (2000) identify bull and bear markets and also find that volatility is much higher in the bear market.

Even the most basic Markov-switching model with constant regime parameters is capable of describing many of the typical characteristics of financial time series. To illustrate this, consider the time series $\{y_t\}$ generated by

$$y_t = \eta_t \sigma_{\Delta_t} + \mu_{\Delta_t}, \tag{1}$$

where $\eta_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$; and $\{\Delta_t\}$ is a Markov chain with k-dimensional state space. The unconditional distribution of y_t is a k-component mixture of normals with the vector of mixing weights being equal to the stationary distribution of the Markov chain. It is well known [see, e.g., McLachlan and Peel (2000, chap. 1)] that such mixture models can give rise to a skewed unconditional distribution if the regime means are different, and that the distribution is usually leptokurtic relative to the normal.[1] Due to the dependence in the Markov chain, the model also gives rise to conditional heteroskedasticity. However, as noted by Pagan and Schwert (1990)

[1] The use of normal mixtures to handle fat tails was first considered as early as 1886 by Newcomb [Newcomb (1980)] in his astronomical studies. He noted that "...the cases are quite exceptional in which the errors are found to really follow the [normal] law. ... [I]t is nearly always found that some of the outstanding errors seem abnormally large. The method of dealing with these abnormal errors has always been one of the most difficult questions in the treatment of observations."

and Timmermann (2000), the basic model given by Equation (1) can generate only limited dynamics in the variance. This is due to the fact that time-varying volatility is solely introduced by discrete shifts in the scale parameter σ^2, with no feedback from the return process and constant variance *within regimes*. However, economic intuition suggests that it is shocks that drive volatility, as these reflect the arrival of new information (or are erroneously taken to do so by noise traders). This is the idea behind the (G)ARCH [(generalized) autoregressive conditional heteroskedasticity] model class of Engle (1982) and Bollerslev (1986). Assuming normally distributed innovations, the popular GARCH(1,1) model is given by

$$\epsilon_t \mid \Psi_{t-1} \sim \mathrm{N}(0, \sigma_t^2), \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{2}$$

where Ψ_t is the information set at time t. A measure of the persistence of shocks to volatility in Equation (2) is given by the sum $\alpha_1 + \beta_1$ which, when estimated from daily data, tends to indicate a rather high degree of persistence. This finding led to the introduction of integrated GARCH processes by Engle and Bollerslev (1986), which take $\alpha_1 + \beta_1 = 1$, so that shocks on volatility do not die out over time. However, it has been argued that the strong persistence of shocks implied by estimated parameters may be due to shifts in the *unconditional variance* of the process, that is, shifts in the scale parameter α_0 in Equation (2), which are not captured by the single-regime model of Equation (2).

If these shifts are persistent, then there are two sources of volatility persistence, namely persistence due to shocks and persistence due to regime switching in the parameters of the variance process. Diebold (1986) appears to be the first to argue in this direction. He notes that, while economic variables "may appear to have integrated-variance disturbances, it may be due to a failure to include monetary-regime dummies for the conditional variance intercept α_0. This would correspond to stationary GARCH movements within regimes, with an unconditional 'jump' occuring between regimes." Recently Mikosch and Starica (2004) showed that deterministic shifts in the unconditional variance do indeed drive the estimate of $\alpha_1 + \beta_1$ toward unity.

Following Diebold's presumption, Lastrapes (1989), Lamoureux and Lastrapes (1990), and Kim and Kon (1999) allow for deterministic shifts in the scale parameter α_0 and find for exchange rates and stock returns that the estimates of the persistence parameter, $\alpha_1 + \beta_1$, decrease substantially. Independent of this research, various tests have been proposed for structural breaks in the conditional variance dynamics of asset returns, and these often indicate multiple breaks in returns over long periods [see Andreou and Ghysels (2002) for an overview].

The practical implication of these findings is clear, given that volatility is an ingredient for pricing financial assets: The omission of switching parameters causes a substantial upward bias in the estimates of the persistence parameters, which impairs volatility forecasts, particularly so in high-volatility periods [see Hamilton and Susmel (1994), Gray (1996), and Klaassen (2002)]. As a way of identifying the timing of the shifts in the unconditional variance, Lamoureux and Lastrapes (1990) suggest the use of Markov-switching models. Cai (1994) and Hamilton and Susmel (1994) follow this idea and combine the ARCH approach with the Markov-switching model of Hamilton (1989). Due to the problem of path dependence (to be discussed below), Cai (1994) and Hamilton and Susmel (1994) confined themselves to ARCH dependencies.

A generalization to Markov-switching GARCH models was developed by Gray (1996) and subsequently modified by Klaassen (2002). While the model of Gray is attractive in that it combines Markov-switching with GARCH effects, its analytical intractability is a serious drawback. As a consequence, conditions for covariance stationarity have yet to be established. Closely related is the lack of an analytic expression for the covariance structure of the squared process. In general, such results and expressions for higher-order moments of GARCH-type models are desirable, and are becoming more prominent in the literature; see, for example, Chen and An (1998), He and Teräsvirta (1999a), Karanasos (1999), Ling and McAleer (2002), Hafner (2003), Karanasos, Psaradakis, and Sola (2004), and Zaffaroni (2004). In the context of Markov-switching GARCH models, analytically tractable expressions for the covariance structure of the squared process can aid in understanding the interaction of the different sources of volatility persistence discussed above.

In this article, we present a new Markov-switching GARCH model. It is the "natural" generalization of the ARCH approach to a multi regime setting and has the advantage of being analytically tractable and allowing us to derive stationarity conditions and further dynamic properties of the process. Furthermore, we present closed-form volatility forecasts, which are not available for the existing models.

The article is organized as follows. The next section briefly reviews the model of Gray (1996) and the modification of Klaassen (2002). The new model, including stationarity conditions, is presented in Section 2, along with a discussion of its volatility dynamics. In Section 3, we apply the model to three exchange rate series, discuss the parameter estimates, and explore the in- and out-of-sample fit. Section 4 concludes. Various technical details are gathered in the appendix.

1 MARKOV-SWITCHING GARCH MODELS

Cai (1994) and Hamilton and Susmel (1994) propose Markov-switching ARCH models, though (standard) GARCH(1,1) models are known to provide better descriptions of market volatility than even high-order ARCH specifications. Their restriction to ARCH models was due to the path dependence in Markov-switching GARCH models that arises when "literally" translating the GARCH model of Bollerslev (1986) to a regime-switching setting. To illustrate this, let $\{\epsilon_t\}$ be generated by

$$\epsilon_t = \eta_t \sigma_t,$$

$$\sigma_t^2 = \alpha_0(\Delta_t) + \alpha_1(\Delta_t)\epsilon_{t-1}^2 + \beta_1(\Delta_t)\sigma_{t-1}^2, \tag{3}$$

with $\eta_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $\{\Delta_t\}$ being a Markov chain with k-dimensional state space. Given σ_0^2, recursive substitution in Equation (3) yields

$$\sigma_t^2 = \sum_{i=0}^{t-1}[\alpha_0(\Delta_{t-i}) + \alpha_1(\Delta_{t-i})\epsilon_{t-1-i}^2]\prod_{j=0}^{i-1}\beta_1(\Delta_{t-j}) + \sigma_0^2\prod_{i=0}^{t-1}\beta_1(\Delta_{t-i}),$$

showing that σ_t^2 depends on the entire history of regimes. The evaluation of the likelihood function for a sample of length T requires the integration over all k^T possible (unobserved) regime paths, rendering estimation of Equation (3) infeasible in practice. To circumvent the path dependence, Gray (1996) noted that the conditional distribution of ϵ_t in Equation (3) is a mixture of normals with time-varying mixing weights. The mixing weights can be computed from the state inferences $p_{t-1}(\Delta_t = j)$, $j = 1, \ldots, k$, expressing the probability that the chain is in state j at time t, given the information at time $t-1$. The conditional variance of ϵ_{t-1}, given information at time $t-2$, can be calculated by

$$h_{t-1} = \sum_{j=1}^{k} p_{t-2}(\Delta_{t-1} = j)\sigma_{jt-1}^2, \tag{4}$$

where σ_{jt-1}^2 is the variance of ϵ_{t-1}, given $\Delta_{t-1} = j$. Gray (1996) uses Equation (4) to replace Equation (3) with GARCH equations for each regime variance,

$$\sigma_{it}^2 = \alpha_{0i} + \alpha_{1i}\epsilon_{t-1}^2 + \beta_{1i}h_{t-1}, \quad i = 1, \ldots, k. \tag{5}$$

With Equation (5), the likelihood function can be computed using a first-order recursive scheme in the same way as in the basic switching model of Hamilton (1989).

Klaassen (2002) proposes the use of $p_{t-1}(\Delta_{t-1} = j)$, $j = 1, \ldots, k$, instead of $p_{t-2}(\Delta_{t-1} = j)$, that is, the use of information up to time $t-1$ instead of $t-2$. In addition, Klaassen (2002) makes use of the fact that the regime at time t belongs to the information set on which σ_t^2 is conditioned and employs the probabilities $p_{t-1}(\Delta_{t-1} = j \mid \Delta_t = i)$ to update the variance in the ith regime.

While the combination of a Markov-switching structure with regime-dependent GARCH effects is thus feasible, there are, as discussed in the introduction, several drawbacks of this approach. The Markov-switching GARCH model introduced below overcomes these drawbacks. In the new specification, the regime variances only depend on past shocks and their own lagged values. This gives rise to a Markov-switching GARCH model that is straightforwardly estimated by maximum likelihood, analytically tractable, and offers an appealing disaggregation of the conditional variance process.

2 A NEW APPROACH TO MARKOV-SWITCHING GARCH MODELS

2.1 Interpreting the Variance Process in Markov-Switching GARCH Models

As discussed at the end of the previous section, the interpretation of the variance process in Gray's approach is problematic. We address this in more detail here and propose an alternative formulation of Markov-switching GARCH models which appears more adequate for analyzing time-varying volatility dynamics.

Prima facie, the variance specification in Gray's model, given by Equations (4) and (5), seems reasonable because, as in the standard (single-regime) GARCH model, past shocks and past (process) variances affect today's (regime) variances. However, we find this a spurious analogy. Recall that the primary feature of a GARCH model (and one of the reasons for its success) is that shocks drive the volatility. Consequently the intuition behind the generalization from ARCH to GARCH is that the latter can parsimoniously represent a high-order ARCH process [Bollerslev (1986), Bera and Higgins (1993)]. To show the implications of this rationale for the multiregime case, consider initially the single-regime GARCH(1,1) model of Equation (2). If, in Equation (2), $\beta_1 < 1$, we can express σ_t^2 in terms of past ϵ_t^2's as an ARCH(∞), given by $\sigma_t^2 = \alpha_0(1-\beta_1)^{-1} + \alpha_1 \sum_1^{\infty} \beta_1^{i-1}\epsilon_{t-i}^2$. The ARCH($\infty$) representation reveals the role of parameters α_1 and β_1 in the GARCH(1,1): α_1 reflects the magnitude of a unit shock's immediate impact on the next period's variance, while β_1 is a parameter of inertia and indicates the memory in the variance. The total impact of a unit shock to future variances is $\alpha_1(1-\beta_1)^{-1}$.

One reason for specifying mixture models that allow for different GARCH behavior in each regime is to capture the difference in the variance dynamics in low- and high-volatility periods. For example, relatively large values of α_1 and relatively low values of β_1 in high-volatility regimes may indicate a tendency to overreact to news — possibly due to a prevailing panic-like mood — compared to "regular" periods, while there is less memory in these sub-processes. Such an interpretation requires a parameterization of Markov-switching GARCH models that implies a clear association between the GARCH parameters within regime j, that is, α_{0j}, α_{1j}, and β_j, and the corresponding σ_{jt}^2 process.

In Equations (4) and (5), the relation between the pattern with which σ_{jt}^2 responds to shocks and the parameters α_{0j}, α_{1j}, and β_{1j} is far from obvious. For example, β_{1j} in Equation (5) cannot be viewed as the inertia in σ_{jt}^2, as h_{t-1} is composed of *all* component variances as well as the time-varying conditional regime probabilities. Thus the variance of regime j will be affected by shocks, even if α_{1j} is zero, as long as h_t changes and $\beta_j \neq 0$. In fact, due to time-varying regime probabilities inferred from past data, the regime variances will not be constant, even if all α_{1j} are zero. That is, the regime variances (and not just the overall variance as in the basic Markov-switching framework) will vary just because the conditional regime probabilities are not constant, contrary to the notion of shocks driving the regime processes. The economic significance of such variance dynamics is unclear, and the disaggregation of the overall variance process provided by this model is at best difficult to interpret.

In the following subsection we present a multiregime GARCH model that preserves the direct correspondence between regime j's GARCH parameters and its variance process. It allows a clear-cut interpretation of the variance dynamics in each regime, can be estimated in a straightforward fashion, and admits formal analysis of the properties of the overall variance process, such as covariance stationarity and the dependence structure of ϵ_t^2.

2.2 Definition of the Model

Let a time-series $\{\epsilon_t\}$ satisfy

$$\epsilon_t = \eta_t \sigma_{\Delta_t, t}, \tag{6}$$

where $\eta_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$, and $\{\Delta_t\}$ is a Markov chain with finite state space $S = \{1, 2, \ldots, k\}$ and an irreducible and primitive $k \times k$ transition matrix, P, with typical element $p_{ij} = P\,(\Delta_t = j \,|\, \Delta_{t-1} = i)$, that is,

$$P = [p_{ij}] = [P(\Delta_t = j \,|\, \Delta_{t-1} = i)], \quad i, j = 1, \ldots, k. \tag{7}$$

The $k \times 1$ vector $\sigma_t^{(2)} = [\sigma_{1t}^2, \sigma_{2t}^2, \ldots, \sigma_{kt}^2]'$ of regime variances follows the GARCH(1,1) equation

$$\sigma_t^{(2)} = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^{(2)}, \tag{8}$$

where $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ik}]'$, $i = 0, 1$; $\beta = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_k)$; and inequalities $\alpha_0 > 0$, α_1, $\beta \geq 0$ are assumed to hold element-wise to guarantee positivity of the variance process. The stationary distribution of the Markov chain will be denoted by $\pi_\infty = [\pi_\infty^1, \pi_\infty^2, \ldots, \pi_\infty^k]'$. We will denote the k-regime MS-GARCH model defined in Equations (6)–(8) by MSG(k), in short.

If, in Equations (6)–(8), $\max\{\beta_1, \ldots, \beta_k\} < 1$ holds, then Equation (8) can be inverted such that $\sigma_t^{(2)} = (I - \beta)^{-1}\alpha_0 + \sum_1^\infty \beta^{i-1}\alpha_1 \epsilon_{t-i}^2$. Diagonality of β implies $\sigma_{jt}^2 = \alpha_{0j}(1 - \beta_j)^{-1} + \alpha_{1j}\sum_1^\infty \beta_j^{i-1}\epsilon_{t-i}^2$, for $j = 1, \ldots, k$. Thus the evolution of σ_{jt}^2 depends only on the within-regime GARCH parameters. In addition, the ARCH(∞) representation of σ_{jt}^2, $j = 1, \ldots, k$, resembles the single-regime model, as $\alpha_{1j}\,(1 - \beta_j)^{-1}$ is the total impact of a unit shock to component j's future variances, α_{1j} measures the magnitude of a shock's immediate impact on the next period's σ_{jt}^2, and β_j reflects the memory in component j's variance in response to such a shock.

A comparison between the process of Equations (6)–(8) and the Markov-switching GARCH model of Equation (3) shows that, for period t, the variance dynamics are the same if $\Delta_t = \Delta_{t-1}$. If a regime shift takes place, then the variance specifications will differ. Suppose that there are two regimes, one with low and one with high volatility, and that there is a shift from the low-volatility to the high-volatility regime at time t. In Equation (3), the low-volatility regime variance in $t - 1$ determines the volatility dynamics irrespective of the shift, whereas the proposed model is governed right away by the high-volatility regime. This results in an instantaneous shift in variance, which is in agreement with the notion of a regime *shift* as put forth by Hamilton (1990), who refers to "occasional discrete shifts" in the underlying stochastic process. It is also in agreement with the observation reported in Dueker (1997) that volatility often increases substantially in a short amount of time at the onset of a turbulent period. Therefore the property of the standard GARCH model that low (high) volatility today results, on average, in low (high) next-period volatility is reasonable within a given regime, but not necessarily so for periods following a regime shift.

2.3 Dynamic Properties

2.3.1 Moment conditions. Next, we provide stationarity conditions for the process defined in Equations (6)–(8) and derive an expression for the unconditional variance. The proofs of these results are given in the appendix. Although we state the results for normally distributed innovations, the results are trivially extended to more flexible distributions of the η_t's. Consider the matrices

$$M_{ji} = p_{ji}(\beta + \alpha_1 e_i'), \quad i, j = 1, 2, \ldots, k, \tag{9}$$

where e_i is the ith $k \times 1$ unit vector.

To understand the role of matrices M_{ji}, suppose $\sigma_t^{(2)}$ and $\Delta_{t-1} = j$ are known with certainty. Then, with probability p_{ji}, the process will be in state i at time t. In this case, $\mathrm{E}(\epsilon_t^2 \mid \Delta_t = i, \sigma_t^{(2)}) = \sigma_{it}^2$, $\mathrm{E}(\sigma_{t+1}^{(2)} \mid \Delta_t = i, \sigma_t^{(2)}) = \alpha_0 + (\beta + \alpha_1 e_i')\sigma_t^{(2)}$, and

$$\mathrm{E}(\sigma_{t+1}^{(2)} \mid \Delta_{t-1} = j,\, \sigma_t^{(2)}) = \alpha_0 + \sum_{i=1}^{k} M_{ji}\sigma_t^{(2)}.$$

In order to derive the stationarity condition for the process defined by Equations (6) and (8), collect all M_{ji} matrices in the $k^2 \times k^2$ matrix

$$M = \begin{bmatrix} M_{11} & M_{21} & \cdots & M_{k1} \\ M_{12} & M_{22} & \cdots & M_{k2} \\ \vdots & \vdots & & \vdots \\ M_{1k} & M_{2k} & \cdots & M_{kk} \end{bmatrix}. \tag{10}$$

For example, for $k = 2$ we have

$$M = \begin{bmatrix} p_{11}(\alpha_{11} + \beta_1) & 0 & p_{21}(\alpha_{11} + \beta_1) & 0 \\ p_{11}\alpha_{12} & p_{11}\beta_2 & p_{21}\alpha_{12} & p_{21}\beta_2 \\ p_{12}\beta_1 & p_{12}\alpha_{11} & p_{22}\beta_1 & p_{22}\alpha_{11} \\ 0 & p_{12}(\alpha_{12} + \beta_2) & 0 & p_{22}(\alpha_{12} + \beta_2) \end{bmatrix}. \tag{11}$$

As shown in the appendix, the process defined by Equations (6)–(8) is stationary if and only if $\rho(M) < 1$, where $\rho(\cdot)$ denotes the largest eigenvalue in modulus of a matrix. It follows from Equation (8) that $\beta_j < 1$ for all j is a necessary condition for the process to be stationary. The nonnegativity of the parameters implies $M > P \otimes \beta$, and because the largest eigenvalue of a nonnegative matrix is increasing in its elements, we have $\rho(M) > \rho(P \otimes \beta) = \rho(P)\rho(\beta) = \rho(\beta) = \max\{\beta_1, \ldots, \beta_k\}$. Clearly this is a simple consequence of the fact that, in our model, β_j measures the memory in σ_{jt}^2, with $\beta_j > 1$ implying that the impact of a shock increases in time.

If condition $\rho(M) < 1$ holds, then the unconditional expectation of $\sigma_t^{(2)}$ is given by

$$\mathrm{E}(\sigma_t^{(2)}) = [I_k, \ldots, I_k](I_{k^2} - M)^{-1}(\pi_\infty \otimes \alpha_0), \tag{12}$$

and the unconditional variance of the process is

$$\mathrm{E}(\epsilon_t^2) = (\mathrm{vec}\ P)'(I_{k^2} - M)^{-1}(\pi_\infty \otimes \alpha_0). \tag{13}$$

To derive the condition for the existence of the unconditional fourth moment (and hence kurtosis), we construct the $k^3 \times k^3$-matrix Q from the $k^2 \times k^2$-submatrices Q_{ji}, given by

$$Q_{ji} = p_{ji}[3(\alpha_1 \otimes \alpha_1)\mathrm{vec}[\mathrm{diag}(e_i)]' + (\alpha_1 e_i') \otimes \beta + \beta \otimes (\alpha_1 e_i') + \beta \otimes \beta], \quad i, j = 1, \ldots, k, \tag{14}$$

analogous to matrix M defined in Equation (10) and composed of matrices M_{ji} as given in Equation (9). As we show in the appendix, the unconditional fourth moment of the process defined by Equations (6)–(8) exists if and only if the process is covariance stationary and the maximal eigenvalue of Q is smaller than unity, that is, $\rho(Q) < 1$. An expression for the fourth moment is derived in the appendix and given in Equation (42).

2.3.2 Two special cases. We now examine two special cases of the MSG(k) model introduced above. Consider first the k-regime Markov-switching ARCH(1) model with $\beta = 0$. This variant of the model introduced above coincides with the ARCH(1) version of the model given by Equation (3), which is studied in detail by Francq, Roussignol, and Zakoïan (2001). It follows from their results that this special process is covariance stationary, if the eigenvalues of the matrix $A = \mathrm{diag}\,(\alpha_1)P$ are less than one in magnitude. It is straightforward to see that the eigenvalues of A are the nonzero eigenvalues of the matrix in Equation (10): Let λ be any eigenvalue of A and $x = [x_1, x_2, \ldots, x_k]'$ a corresponding right eigenvector. Then, λ is an eigenvalue of Equation (10) with corresponding eigenvector $\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{k^2}]'$, where $\tilde{x}_{1+(j-1)k} = x_j$ for $j = 1, 2, \ldots, k$, and $\tilde{x}_{i+(j-1)k} = (\alpha_i/\alpha_1)x_j$ for $i = 2, \ldots, k$ and $j = 1, 2, \ldots, k$.

The second special case is the mixed normal (MN) GARCH model [Vlaar and Palm (1993), Alexander and Lazar (2004), Haas, Mittnik, and Paolella (2004)], which assumes that the regimes are drawn independently from a multinomial distribution. This results in a mixture GARCH model with a constant vector of mixing weights, say $\lambda = [\lambda_1, \ldots, \lambda_k]'$, for which

$$M_{1j} = M_{2j} = \cdots = M_{kj} = \lambda_j(\beta + \alpha_1 e_j'), \quad j = 1, \ldots, k. \tag{15}$$

It follows from the results of Haas, Mittnik, and Paolella (2004) that, given $\max\{\beta_1, \ldots, \beta_k\} < 1$ holds, a necessary and sufficient condition for the MN-GARCH process to be covariance stationary is

$$1 - \sum_{j=1}^{k} \lambda_j \frac{\alpha_{1j}}{1 - \beta_j} = \sum_{j=1}^{k} \frac{\lambda_j}{1 - \beta_j}(1 - \alpha_{1j} - \beta_j) > 0. \tag{16}$$

This is a straightforward generalization of the condition in the standard GARCH(1,1) model [Bollerslev (1986)], and is easily shown to be equivalent to $\rho(M) < 1$.

Strictly speaking, the MN-GARCH model is not just a special case of the MS-GARCH, as it allows for the introduction of different component means, and thus skewness, in the conditional mixture *without* altering the first-order dynamics of the process. In particular, adding a regime-specific mean μ_{Δ_t} in Equation (6) with nonzero Markov persistence results in autocorrelated ϵ_t's and, thus, loss of the GARCH property, namely lack of correlation for the raw paired with presence of correlation for the squared series [Timmermann (2000)]. However, if the process $\{\Delta_t\}$ is taken to be iid multinomial, and enforcing the restriction $\mu_k = -\lambda_k^{-1}\sum_{j=1}^{k-1}\lambda_j\mu_j$ so that $\mathrm{E}(\epsilon_t \mid \Psi_{t-1}) = \mathrm{E}(\epsilon_t) = 0$ and hence $\mathrm{E}(\epsilon_{t-\tau}\epsilon_t) = \mathrm{E}[\epsilon_{t-\tau}\mathrm{E}(\epsilon_t \mid \Psi_{t-1})] = 0$ for $\tau \geq 1$, as in Haas, Mittnik, and Paolella (2004), the raw series will still be uncorrelated. Hence the MN-GARCH model may be particularly suitable for modeling and predicting return distributions that appear to be skewed. Analogous to our abbreviation MSG(k), our short notation for the k-component MN-GARCH process will be MNG(k).

Apart from the more conceptual caveats indicated in the last paragraph, the MSG structure put forward here can, in principle, be extended to support different regime means, $\mu_{j,t}$, $j = 1, \ldots, k$, without invalidating the discussion of Section 2.1. It is worthwhile to briefly discuss the notion of a shock, ϵ_t, in regime-switching models with switching means, which has not been unambiguously defined in the literature. Denote the variable of interest, that is, the return at time t, by r_t. Then, in the spirit of Francq and Zakoïan (2001), who study Markov-switching autoregressive moving average (ARMA) processes, we could define $\epsilon_t = r_t - \mu_{\Delta_t,t}$. As this definition of a shock implies that ϵ_t depends on the hidden state variable and thus is not observable, its use considerably complicates parameter estimation, as discussed in Section 1. In contrast, Gray's (1996) recombining approach, outlined in Section 1, sets $\epsilon_t = r_t - \mathrm{E}(r_t \mid \Psi_{t-1}) = r_t - \sum_j p_{t-1}(\Delta_t = j)\,\mu_{j,t}$, which allows computation of ϵ_t by Hamilton's (1989) standard recursive algorithm. Although use of the recombining approach with the MSG process defined in Equations (6)–(8) is feasible, a definition that may fit better into its structure is that employed by Wong and Li (2001) in the context of a mixture autoregressive ARCH model, which was also adopted by Lanne and Saikkonen (2003). Here, shocks are defined with respect to a given regime, and the shock within the jth regime is $\epsilon_{jt} = r_{t-1} - \mu_{j,t-1}$, which gives rise to a modification of Equation (8), namely, $\sigma_{jt}^2 = \alpha_{0j} + \alpha_{1j}(r_t - \mu_{j,t})^2 + \beta_j\sigma_{j,t-1}^2$, $j = 1, \ldots, k$. Also, with this specification, as long as only the intercept in the mean equation is due to regime switching, all the theoretical results discussed in the present section can be extended in a conceptually straightforward manner, although the algebra becomes rather lengthy and tedious.[2]

[2] No theoretical results exist for MS models with regime-specific dynamics in both the means and variances. Lanne and Saikkonen (2003) use simulation methods to evaluate the stationarity properties of a general dynamic mixture model.

Note, however, that whichever notion of a shock is adopted, the disentanglement of the distributional shape from the return autocorrelation structure, which is a feature of the MNG process, is lost [see also the discussion in Haas, Mittnik, and Paolella (2004)]. Recently Geweke and Amisano (2003) proposed a method to get skewness into Markov-switching models without implying serial correlation in returns: They assume that the regime densities are skewed normal mixtures with zero means. Combining this approach with the structure proposed herein may be more appealing than those discussed in the preceding paragraph.

2.3.3 Persistence in variance and the role of the mixing process. In this section we investigate the relation between the switching of parameters in the GARCH equation and the degree of volatility persistence. Assume that the unconditional fourth moment of the process exists. In the appendix we derive the autocovariance function of the squared process $\{\epsilon_t^2\}$ defined by Equations (6)–(8) and show that, for $k = 2$,

$$\begin{aligned}\operatorname{cov}(\epsilon_t^2, \epsilon_{t-\tau}^2) = (\operatorname{vec} P)'(M^{\tau-1}\Theta P^{\star}\Gamma + (I_4 - M)^{-1}M^{\tau}[(C_1\,\tilde{\sigma}^{(2)}) \otimes \alpha_0] \\ + (\delta I_4 - M)^{-1}(\delta^{\tau} I_4 - M^{\tau})[(C_2\,\tilde{\sigma}^{(2)}) \otimes \alpha_0]),\end{aligned} \tag{17}$$

where M is defined in Equation (10), $\delta = p_{11} + p_{22} - 1$ is the degree of persistence due to the Markov effects, and Θ, $P^{\star}$, Γ, C_1, C_2, and $\tilde{\sigma}^{(2)}$ are constant matrices defined in the appendix. From Equation (17) it can be seen that $\operatorname{cov}(\epsilon_t^2, \epsilon_{t-\tau}^2)$ depends on the powers of both M and δ, which implies that the autocovariance function is dominated by exponential decay at a rate of $\max\{\rho(M), \delta\}$. If we assume that the regime shift occurs in the scale α_0, while the genuine persistence parameters α_1 and β_1 in Equation (2) are constant across regimes, then $\rho(M) = \alpha_1 + \beta_1$, and the autocovariances decay at a rate of $\max\{\alpha_1 + \beta_1, \delta\}$. Here $\alpha_1 + \beta_1$ measures the persistence of shocks to the variance, while δ measures the degree of dependence in the squared process that is caused by persistent shifts in the level parameter α_0.

By way of comparison, in the single-regime GARCH(1,1) model of Equation (2), $\operatorname{cov}(\epsilon_{t-\tau}^2, \epsilon_t^2) \propto (\alpha_1 + \beta_1)^{\tau-1}$ [Bollerslev (1988)]. Thus if shifts in α_0 are ignored, and given that the true process is well approximated by an MSG process, we expect that the persistence of shocks will be overestimated if $\delta > \rho(M)$. For example, as noted by Timmermann (2000) and Morana (2002), those values of the transition probabilities p_{11} and p_{22} representing infrequent mixing of regimes, that is, $\delta = p_{11} + p_{22} - 1 \approx 1$, may be interpreted as closely resembling structural break models. In this case, δ will dominate the persistence of shocks, and estimates of $\alpha_1 + \beta_1$ from models ignoring the switching may be plagued by substantial upward bias.

To get an impression of these effects, we carry out a simulation experiment and consider a GARCH(1,1) process with regime switching in α_0, as given by Equation (6)–(8), with

$$\alpha_0 = \begin{bmatrix} 0.05 \\ \alpha_{02} \end{bmatrix}, \quad \alpha_1 = \begin{bmatrix} 0.05 \\ 0.05 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, \quad P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \tag{18}$$

where we let p vary from 0.5 to 0.995, corresponding to $\delta = 2p - 1$ varying from 0 to 0.99. For α_{02}, we use values 0.1 and 0.25, corresponding to an unconditional variance that is five and two times larger in the high-volatility than in the low-volatility regime, respectively. In Equation (18), $\alpha_1 + \beta_1 = 0.75$ is small relative to those values commonly estimated from financial time series. For each value of δ and α_{02}, we fit a standard (single-regime) GARCH(1,1) model to 2,500 datasets of length 10,000 generated from Equation (18) and compute the average estimate of α_1, β_1, and $\alpha_1 + \beta_1$.

Figure 1 plots the relation between δ and the average $\hat{\alpha}_1'$s, $\hat{\beta}_1'$s, and $\hat{\alpha}_1 + \hat{\beta}_1'$s. The small downward bias of $\hat{\alpha}_1 + \hat{\beta}_1$ at $\delta = 0$ just reflects the well-known fact that $\hat{\alpha}_1 + \hat{\beta}_1$ in the GARCH(1,1) model is downward biased, even for fairly large sample sizes, due to a downward biased $\hat{\beta}_1$ [cf. Bollerslev, Engle, and Nelson (1994)].[3] However, as long as δ is small, relative to $\alpha_1 + \beta_1$, at the outset, as δ increases, we observe a *decrease* in the average $\hat{\alpha}_1 + \hat{\beta}_1$. This decrease is caused by a decreasing $\hat{\beta}_1$, which is only partially compensated by an increasing $\hat{\alpha}_1$. The size of the effects depends on the magnitude of the switch in the unconditional variance, as they are larger for $\alpha_{02} = 0.25$ than for $\alpha_{02} = 0.1$.

The findings in Figure 1 have an intuitive explanation. The presence of Markov-switching effects in the intercept of the variance equation gives rise to another source of inertia in σ_t^2, since, if $\delta \geq 0$, a large (small) α_0 today tends to be followed by a large (small) α_0 tomorrow. As this source of persistence is not related to shocks, it mainly affects $\hat{\beta}_1$, while there is a certain amount of compensating movement in $\hat{\alpha}_1$. Apparently the estimated $\alpha_1 + \beta_1$ is some kind of average of both sources of persistence, where the relative weight of δ is determined by its magnitude, and so, as δ becomes larger but is still small in relation to the GARCH persistence, the trade-off between the growing relative weight of (still relatively small) δ and the increase in δ initially induces a decreasing estimated persistence. As δ becomes larger and exceeds $\alpha_1 + \beta_1$, it is clear from Figure 1 that the persistence due to the shifts in α_0 is "thrown into the persistence of an individual shock" [Gray (1996)]. The results have some similarity with those of Granger and Teräsvirta (1999) and Diebold and Inoue (2001), who find by simulation that Markovian regime switching in the mean may lead to the spurious presence of long memory, given that the "staying probabilities" p_{11} and p_{22} are near unity.

In general, we expect that the switching takes place not only in α_0, but in all parameters of the GARCH equation; for example, bull and bear markets may be characterized by *different variance dynamics*. Then, $\rho(M)$ is a complicated function of both the regime-specific GARCH parameters and the persistence properties of the Markov chain. As noted by Yang (2000), an interesting feature

---

[3] The average estimated $\hat{\alpha}_1 + \hat{\beta}_1$, when simulating from the single-regime GARCH(1,1) process $\sigma_t^2 = 0.05 + 0.05\epsilon_{t-1}^2 + 0.7\sigma_{t-1}^2$, with $T = 10{,}000$, is 0.74, as in Figure 1. Thus the downward bias shown at $\delta = 0$ is not caused by the switching intercept. (With $T = 25{,}000$, the bias practically disappears.)

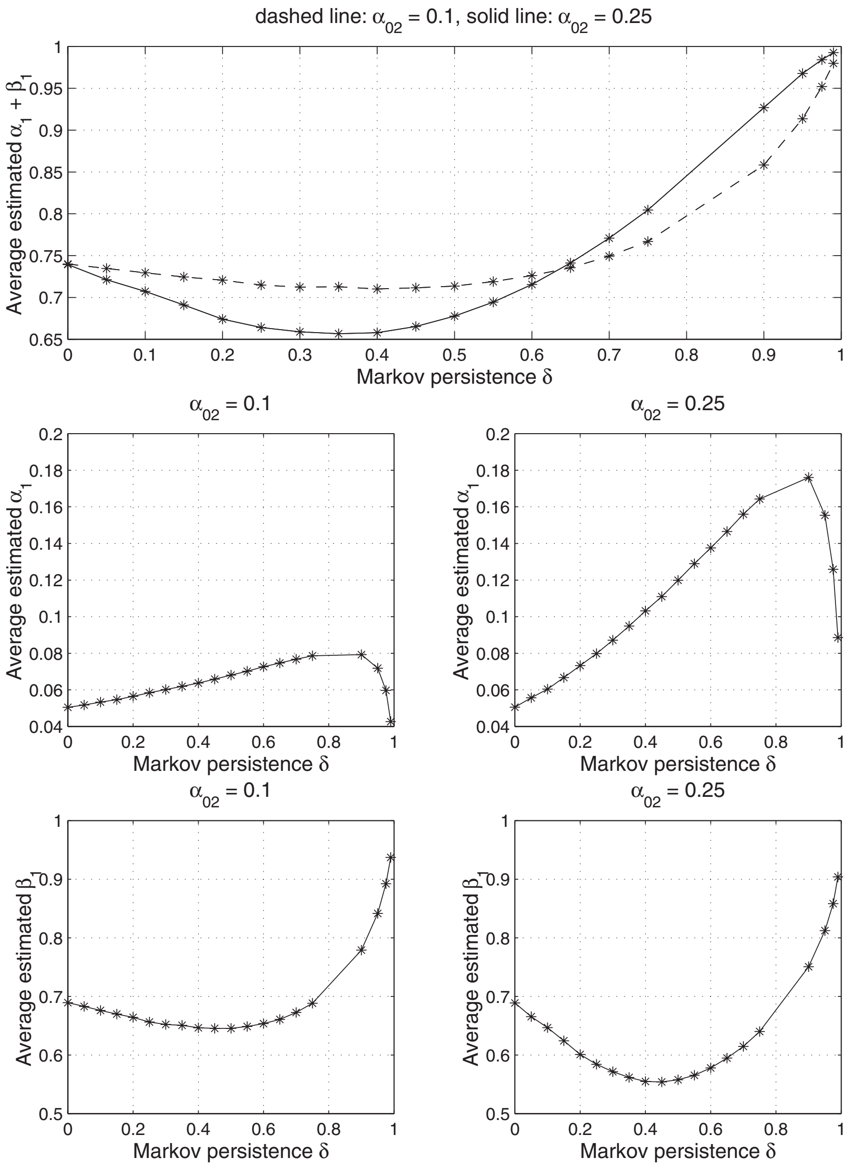

Figure 1 For each value of $\delta = 2p - 1$ and α_{02} in Equation (18), we show the average estimates of $\alpha_1 + \beta_1$ (top panel), α_1 (middle panel), and β_1 (bottom panel), assuming a GARCH process of Equation (2); that is, a GARCH(1,1) without a switch in the scale parameter α_0.

of Markov-switching autoregressions is that the process can occasionally follow a nonstationary process while maintaining stationarity in the long run. As will be demonstrated in the empirical applications below, the MS-GARCH model studied here has the same property.

3 MODELING EXCHANGE RATE DYNAMICS

3.1 Data and Model Specification

Daily returns of the Japanese yen and the British pound exchange rates against the U.S. dollar from January 1978 to June 2003 ($T = 6{,}336$ and 6,313 observations, respectively) and the Singapore dollar against the U.S. dollar from January 1981 to June 2003 ($T = 5{,}313$ observations) are used to empirically investigate the usefulness of the proposed models.[4] Specifically, continuously compounded percentage returns, $r_t = 100(\log S_t - \log S_{t-1})$, are considered, where S_t denotes the exchange rate at time t.

We will estimate the MSG model with both two and three regimes. Recall that the MSG process has been restricted to zero regime means, that is, a symmetric conditional and unconditional distribution. This contrasts the MNG model, where the consideration of skewness does not entail any theoretical complication. Thus we will also consider the MNG process with two and three components and different component means (see the end of Section 2.3.2 for details). This differs from Klaassen (2002) and Alexander and Lazar (2003), who assume a priori that the distribution of exchange rate returns is symmetric, and solely consider symmetric (Markov) mixture models. For the purpose of comparison, we also estimate the single-regime GARCH(1,1) model with normal as well as Student's t innovations, which we subsequently denote simply as normal-GARCH and t-GARCH, respectively. Given the widespread evidence for excess kurtosis in the conditional distribution of asset returns, the t-GARCH process is a more worthy competitor for the mixture models proposed herein. In this specification, the density of the innovation η_t depends on the degrees of freedom parameter, ν, and is given by

$$f(\eta_t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{\eta_t^2}{\nu}\right)^{-(\nu+1)/2}.$$

If $\nu > 4$, the conditional kurtosis of ϵ_t exists and is given by $\mathrm{E}(\eta_t^4)/\mathrm{E}^2(\eta_t^2) = 3(\nu-2)/(\nu-4)$. Fat-tailed distributions such as Student's t could, of course, be assumed for the innovations driving the mixture GARCH processes as well, as in Hamilton and Susmel (1994), Dueker (1997), and Klaassen (2002).[5] Nevertheless, we prefer to retain the normality assumption for the mixture components. From the conceptual viewpoint, it is a particularly appealing feature of the mixture approach that normality within regimes can be preserved. In a mixture GARCH model, for example, there are already two sources of excess kurtosis, and *conditional* fat-tailedness *within* each regime is likely to be superfluous. In fact, empirical results reported in the literature [Alexander and Lazar (2004), Haas, Mittnik, and Paolella (2004)] indicate that MN-GARCH models can

---

[4] The data were obtained from the Internet site http://www.federalreserve.gov/releases/H10/hist/default1999.htm of the New York Federal Reserve Bank.

[5] Perez-Quiros and Timmermann (2001) propose a mixture model for stock returns that combines a normal and a Student density.

capture the excess kurtosis in return processes, and the same conclusions can be drawn from the findings in the present article, as reported, for example, in Table 1.

We use the first 3,000 observations for the specification of the mean equation and the remaining to evaluate the forecasting performance. Using the Bayesian

Table 1 Properties of estimated GARCH models.

	$\rho(M)$	Variance	$\rho(Q)$	Skewness	Kurtosis	K	L	AIC	BIC
Japanese yen									
$\{\hat{\epsilon}_t\}$	—	0.486	—	−0.492	6.847	—	—	—	—
Normal	0.966	0.490	0.941	0	3.433	4	−6406.7	12821	12849
t	0.982	0.555	0.996	0	92.64	5	−6150.0	12310	**12344**
MSG(2)	0.973	0.483	0.951	0	6.467	9	−6153.5	12325	12386
MSG(3)	0.979	0.493	0.963	0	7.936	16	−6120.2	12272	12380
MNG(2)	0.974	0.476	0.953	−0.259	6.273	9	−6142.6	12303	12364
MNG(3)	0.988	0.465	0.978	−0.355	6.698	14	−**6118.6**	−**12265**	12360
British pound									
$\{\hat{\epsilon}_t\}$	—	0.396	—	−0.077	5.958	—	—	—	—
Normal	0.989	0.397	0.983	0	3.848	7	−5590.0	11194	11241
t	0.992	0.444	0.997	0	32.75	8	−5443.4	10903	**10957**
MSG(2)	0.989	0.394	0.985	0	6.008	12	−5437.9	10900	10981
MSG(3)	0.993	0.380	0.990	0	6.284	19	−5422.8	10884	11012
MNG(2)	0.989	0.393	0.984	−0.088	5.932	12	−5435.2	10894	10975
MNG(3)	0.993	0.379	0.989	−0.124	6.168	17	−**5422.1**	**10878**	10993
Singapore dollar									
$\{\hat{\epsilon}_t\}$	—	0.120	—	−0.855	17.88	—	—	—	—
Normal	0.986	0.133	0.988	0	6.827	5	−944.4	1898.7	1931.6
t	0.993	0.293	1.577	—	—	6	−577.6	1167.3	1206.8
MSG(2)	0.991	0.128	0.996	0	26.00	10	−571.4	1162.8	1228.6
MSG(3)	0.989	0.105	0.989	0	13.35	17	−**526.1**	**1086.1**	**1197.9**
MNG(2)	0.990	0.138	1.001	—	—	10	−584.8	1189.6	1255.4
MNG(3)	0.994	0.144	1.001	—	—	15	−553.3	1136.7	1235.3

Shown are unconditional properties and likelihood-based goodness-of-fit criteria of fitted GARCH models over the whole sample period for the Japanese yen, the British pound, and the Singapore dollar, respectively. The models denoted by normal and t are the single-regime GARCH(1,1) processes with normal and Student's t innovations, respectively. MSG(k), $k=2, 3$, is the Markov-switching GARCH with k components, and MNG(k), $k=2, 3$, is the mixed normal GARCH with k components, where the mixture process is multinomial. The row labeled $\{\hat{\epsilon}_t\}$ reports the respective properties for the residuals of an AR(0), AR(3), and AR(1) model for the yen, the pound, and the Singapore dollar, respectively, corresponding to the AR orders that accompany each of the GARCH models. For the models with normal innovations, $\rho(M)$, $\rho(Q)$, and the unconditional variance and kurtosis are derived in the appendix, while for Student's t GARCH(1,1), $\rho(M) = \nu(\nu-2)^{-1}\alpha_{11} + \beta_1, \rho(Q) = 3\nu^2(\nu-2)^{-1}(\nu-4)^{-1}\alpha_{11}^2 + 2\nu(\nu-2)^{-1}\alpha_{11}\beta_1 + \beta_1^2$, where ν is the degrees of freedom parameter, and the unconditional moments are modified accordingly [see, e.g., He and Teräsvirta (1999b, Corollary 1.1)]. The column labeled K reports the number of parameters of a model, including the mean equation; L is the log-likelihood; AIC $-2L+2K$ and BIC $= -2L + K \log T$, where T is the sample size. For each of the criteria, boldface entries indicate the best model for the particular criterion.

information criterion (BIC) [Schwarz (1978)] to choose among autoregressive (AR) mean specifications, we model the conditional mean by a constant for the Japanese yen, an AR(3) for the British pound, and an AR(1) for the Singapore dollar.[6] That is, we model the returns as

$$r_t = c_t + \epsilon_t, \tag{19}$$

with $c_t = a_0$ for the yen, $c_t = a_0 + \sum_{i=1}^{3} a_i r_{t-i}$ for the pound, and $c_t = a_0 + a_1 r_{t-1}$ for the Singapore dollar.

We now turn to the estimation results and discuss certain features that arise with regime-switching models. Subsequently we compare the models' predictive performance by evaluating one-step-ahead density forecasts.

3.2 Estimation Results

The estimation results over the whole sample period are reported in Tables 2–4.[7] The parameter estimates of the mean equations are not shown, these being only of secondary interest. The three upper sections of the tables display the parameters characterizing the dynamics within each of the regimes, with the regimes being ordered with respect to a declining stationary regime probability, that is, $\pi_\infty^1 > \pi_\infty^2 > \pi_\infty^3$. V_j, $j = 1, 2, 3$, denotes the regime-specific volatility persistence, that is, $V_j = \kappa\alpha_{1j} + \beta_j$, where $\kappa = 1$ if the innovations η_t are normal, and $\kappa = \nu/(\nu - 2)$ if they are Student's t with ν degrees of freedom. The fourth section concerns the degrees of freedom parameter for the t-GARCH(1,1) model, the transition matrix P, and its implied degree of Markov dependence, δ, that is, its largest eigenvalue in modulus different from the Frobenius root 1.

Table 1 shows quantities characterizing the estimated processes as a whole, namely the maximum eigenvalues, $\rho(M)$ and $\rho(Q)$, of matrices M and Q defined in Equations (10) and (14), indicating the existence of unconditional second and fourth moments, respectively. As these quantities are smaller than unity for all currencies and all processes (with the exception of $\rho(Q)$ implied by the MNG models for the Singapore dollar), we can compute the unconditional variances and skewness/kurtosis coefficients implied by the estimated parameters for the models, also shown in Table 1. The row labeled $\{\hat{\epsilon}_t\}$ reports the corresponding values for the residuals of an AR(0), AR(3), and AR(1) model for the Japanese yen, the British pound, and the Singapore dollar, respectively, corresponding to the autoregressive lag orders that accompany each of the employed GARCH models. K refers to the number of parameters (including the mean equation) and L to the log-likelihood of the estimated models. It should be noted that standard likelihood

[6] The optimal AR specifications were derived by estimating normal-GARCH(1,1) models for the error term.

[7] Standard errors were obtained by numerically computing the Hessian matrix at the ML estimates. The delta method [see, e.g., Lehmann (1999, p. 315)] was used to approximate the standard errors of functions of estimated quantities, namely, V_i, $i = 1, 2, 3$, the Markov dependence δ, and, in case of the MNG models, $\mu_k = -\lambda_k^{-1} \sum_{j=1}^{k-1} \lambda_j \mu_j$, $k = 2, 3$ (see Section 2.3.2).

ratio tests cannot be employed for testing the number of regimes [see, e.g., Hansen (1992) and McLachlan and Peel (2000)]. We also report the AIC [Akaike (1973)] and BIC, which may serve as an additional (informal) assessment of (relative) model fit.

Table 2 Parameter estimates for the Japanese yen.

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
α_{01}	0.017	0.006	0.003	0.004	0.002	0.002
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
α_{11}	0.065	0.032	0.023	0.028	0.023	0.027
	(0.007)	(0.005)	(0.004)	(0.006)	(0.004)	(0.007)
β_1	0.900	0.926	0.945	0.958	0.947	0.969
	(0.012)	(0.011)	(0.008)	(0.008)	(0.009)	(0.009)
V_1	0.966	0.982	0.968	0.986	0.970	0.995
	(0.006)	(0.005)	(0.005)	(0.004)	(0.006)	(0.004)
π_∞^1	1	1	0.736	0.573	0.753	0.528
					(0.033)	(0.063)
$\mathrm{E}\sigma_{1t}^2$	0.490	0.321	0.249	0.421	0.254	0.473
μ_1	0	0	0	0	0.043	0.049
					(0.009)	(0.023)
α_{02}	—	—	0.097	0.001	0.101	0.002
			(0.045)	(0.001)	(0.047)	(0.002)
α_{12}	—	—	0.227	0.033	0.231	0.044
			(0.072)	(0.007)	(0.069)	(0.015)
β_2	—	—	0.818	0.881	0.813	0.858
			(0.063)	(0.028)	(0.063)	(0.040)
V_2	—	—	1.045	0.914	1.044	0.902
			(0.031)	(0.023)	(0.033)	(0.028)
π_∞^2	0	0	0.264	0.287	0.247	0.369
					(0.033)	(0.063)
$\mathrm{E}\sigma_{2t}^2$	—	—	1.132	0.144	1.129	0.160
μ_2	—	—	0	0	−0.130	0.028
					(0.034)	(0.017)
α_{03}	—	—	—	0.252	—	0.318
				(0.136)		(0.149)
α_{13}	—	—	—	0.387	—	0.593
				(0.148)		(0.199)
β_3	—	—	—	0.690	—	0.570
				(0.132)		(0.141)
V_3	—	—	—	1.077	—	1.163
				(0.084)		(0.140)
π_∞^3	0	0	0	0.140	0	0.103
						(0.027)
$\mathrm{E}\sigma_{3t}^2$	—	—	—	1.428	—	1.380
μ_3	—	—	—	0	—	−0.351
						(0.096)

continued

Table 2 (*continued*)

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
ν	—	4.747 (0.286)	—	—	—	—
P	1	1	$\begin{bmatrix} 0.744 & 0.715 \\ (0.037) & (0.056) \\ 0.256 & 0.285 \\ (0.037) & (0.056) \end{bmatrix}$	$\begin{bmatrix} 0.948 & 0.104 & 0.000 \\ (0.017) & (0.039) & (0.013) \\ 0.000 & 0.652 & 0.713 \\ (0.023) & (0.044) & (0.062) \\ 0.052 & 0.244 & 0.287 \\ (0.020) & (0.036) & (0.061) \end{bmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_3'$
δ	—	—	0.029 (0.072)	0.868 (0.045)	0	0

Standard errors are given in parentheses.
The top row of the table labels the estimated model. The models denoted by normal and t are the GARCH(1,1) processes with normal and Student's t innovations, respectively. MSG(k), $k=2$, 3, is the Markov-switching GARCH with k components, and MNG(k), $k=2$, 3, is the mixed normal GARCH with k components, where the mixture process is multinomial. V_i, $i=1$, 2, 3, is the component-specific degree of volatility persistence, that is, $V_i=\alpha_{1i}+\beta_i$ for the models with normally distributed innovations, and $V_1=\nu(\nu-2)^{-1}\,\alpha_{11}+\beta_1$ for the Student's t GARCH(1,1) process, where ν is the degrees of freedom parameter of the conditional Student's distribution. δ is the degree of persistence in the (Markov) mixture process, and $\mathbf{1}_k$, $k=2$, 3, is a k-dimensional column of ones. The mean Equation (19) is specified as a constant, that is, $c_t=a_0$.

Comparing the empirical kurtosis coefficients computed from the series $\{\hat{\epsilon}_t\}$ and those implied by the estimated models, we note the (widely observed) fact that the normal GARCH model cannot account for the excess kurtosis in financial return series. In contrast, the MSG and MNG models fit the empirical kurtosis rather well, even with two components. The Singapore dollar may be seen as an exception from this good fit, but it is unclear whether the fourth moment of this process exists. This indicates that, in the context of conditional heteroskedastic mixture models, the normal assumption may be adequate. In these models, the excess kurtosis stems from both the GARCH and the mixture effects, so that the use of fat-tailed regime densities will be superfluous. It is difficult to give any meaning to the unconditional kurtosis values computed from the Student's t distribution, given that the respective values of $\rho(Q)$, although just below unity for the yen and the pound, indicate a non existent fourth moment. However, given recent evidence for the existence of the fourth moment of major exchange rates against the U.S. dollar, including the yen and the pound [Huisman et al. (2002)], this may be viewed as a drawback of the conditional Student's distribution.

Table 1 also shows the respective skewness coefficients, but it is difficult to assess the significance of the asymmetries in the data, given their evident fat-tailedness, and we defer formal testing to the forecasting section. Nevertheless, the AIC and BIC, as indicators of relative model fit, suggest the importance of asymmetries, given that for two of the three series, the asymmetric MNG models are preferred relative to the MSG models. For the Singapore dollar, however, the

Markov model with three regimes clearly outperforms the multinomial mixture process.

Figure 2 provides the empirical autocorrelations of the squared AR residuals along with their theoretical counterparts, implied by the estimated normal and

Table 3 Parameter estimates for the British pound.

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
α_{01}	0.004	0.002	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
α_{11}	0.049	0.035	0.037	0.030	0.038	0.030
	(0.005)	(0.005)	(0.007)	(0.008)	(0.007)	(0.008)
β_1	0.941	0.939	0.927	0.969	0.926	0.970
	(0.006)	(0.008)	(0.016)	(0.008)	(0.015)	(0.008)
V_1	0.989	0.992	0.964	0.999	0.963	0.999
	(0.003)	(0.003)	(0.010)	(0.004)	(0.009)	(0.004)
π_∞^1	1	1	0.626	0.546	0.634	0.521
					(0.048)	(0.081)
$\mathrm{E}\sigma_{1t}^2$	0.397	0.299	0.205	0.412	0.207	0.418
μ_1	0	0	0	0	0.022	0.016
					(0.009)	(0.020)
α_{02}	—	—	0.010	0.000	0.010	0.000
			(0.004)	(0.001)	(0.004)	(0.001)
α_{12}	—	—	0.071	0.049	0.073	0.051
			(0.016)	(0.025)	(0.017)	(0.010)
β_2	—	—	0.947	0.872	0.946	0.871
			(0.012)	(0.052)	(0.012)	(0.029)
V_2	—	—	1.018	0.921	1.018	0.923
			(0.007)	(0.030)	(0.007)	(0.023)
π_∞^2	0	0	0.374	0.357	0.366	0.380
					(0.048)	(0.083)
$\mathrm{E}\sigma_{2t}^2$	—	—	0.710	0.146	0.712	0.153
μ_2	—	—	0	0	−0.039	0.014
					(0.018)	(0.015)
α_{03}	—	—	—	0.057	—	0.056
				(0.049)		(0.037)
α_{13}	—	—	—	0.242	—	0.246
				(0.139)		(0.117)
β_3	—	—	—	0.860	—	0.855
				(0.084)		(0.063)
V_3	—	—	—	1.102	—	1.101
				(0.073)		(0.075)
π_∞^3	0	0	0	0.097	0	0.099
						(0.050)
$\mathrm{E}\sigma_{3t}^2$	—	—	—	1.063	—	1.024
μ_3	—	—	—	0	—	−0.137
						(0.078)

continued

Table 3 (*continued*)

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
ν	—	6.108 (0.448)	—	—	—	—
P	1	1	$\begin{bmatrix} 0.642 & 0.598 \\ (0.058) & (0.070) \\ 0.358 & 0.402 \\ (0.058) & (0.070) \end{bmatrix}$	$\begin{bmatrix} 0.378 & 0.681 & 1.000 \\ (0.138) & (0.201) & (1.140) \\ 0.467 & 0.285 & 0.000 \\ (0.194) & (0.119) & (1.140) \\ 0.155 & 0.034 & 0.000 \\ (0.138) & (0.134) & (0.000) \end{bmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_3'$
δ	—	—	0.044 (0.096)	−0.406 (0.193)	0	0

See Table 2 for explanations, but note the following modification: The mean Equation (19) is specified as an AR(3) process, that is, $c_t = a_0 + \sum_{i=1}^{3} a_i r_{t-i}$.

MS-GARCH models. The differences between the functions are minor for the pound, but more striking for the yen and the Singapore dollar. The empirical sequences for these series are characterized by high values at the first few lags and a sudden subsequent decline. The normal GARCH(1,1) model cannot mimic this pattern. Its autocorrelation function either does not capture the large value at the first lag (in case of the yen) or decreases too slow after the first lags (Singapore dollar). The theoretical autocorrelation functions of the MSG models are, of course, more flexible, and in particular the sequence implied by MSG(3) reproduces the properties described above.

Turning to the parameter estimates reported in Tables 2–4, we draw attention to two features, which are similar across the three currencies. The first concerns the presence of regimes with different volatility dynamics. All mixture models identify at least one nonstationary regime (i.e., for some j, we have $\alpha_{1j} + \beta_j > 1$). Also, for all currencies, in the nonstationary regimes there is more weight on the respective α_{1j}'s and less weight on the β_j's than in the stationary regimes. This can be interpreted along the lines indicated in Section 2.1.

The second feature relates to the degree of Markov persistence and the persistence of shocks to volatility. We note that, in the case of the yen and the pound, although the three-regime MSG and MNG models both have a relatively moderate sum $\alpha_{12} + \beta_2$ in the second regime, we do not observe a "switch" of volatility persistence from the GARCH structure to the Markov chain, compared to the single-regime model. In fact, the Markov-persistence is even insignificant in the MSG(2) models for the yen and the pound, and at least close to insignificant in the MSG(3) model for the pound.[8] This explains why the MSG models are,

[8] Note, however, that the standard errors of the elements of P in this case have to be interpreted with care. If the "true value" of p_{31} was on the boundary (i.e., $p_{31} = 1$), as suggested by the results reported in Table 3, then the usual asymptotic behavior does not hold.

according to AIC and BIC, dominated by the MNG class for the yen and the pound.

In addition, the parameter estimates obtained for the Singapore dollar, where MSG outperforms MNG according to the in-sample criteria reported in Table 1, do

Table 4 Parameter estimates for the Singapore dollar.

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
α_{01}	0.002	0.001	0.001	0.002	0.001	0.000
	(0.0003)	(0.0002)	(0.0002)	(0.001)	(0.0002)	(0.0002)
α_{11}	0.090	0.066	0.049	0.044	0.076	0.023
	(0.012)	(0.008)	(0.008)	(0.014)	(0.011)	(0.007)
β_1	0.896	0.867	0.902	0.940	0.861	0.947
	(0.013)	(0.016)	(0.013)	(0.018)	(0.017)	(0.013)
V_1	0.986	0.993	0.951	0.983	0.937	0.970
	(0.004)	(0.004)	(0.007)	(0.008)	(0.008)	(0.008)
π^1_∞	1	1	0.798	0.462	0.837	0.481
					(0.021)	(0.075)
$\mathrm{E}\sigma^2_{1t}$	0.133	0.153	0.069	0.010	0.082	0.063
μ_1	0	0	0	0	0.008	0.015
					(0.003)	(0.010)
α_{02}	—	—	0.004	0.000	0.004	0.003
			(0.001)	(0.0001)	(0.001)	(0.012)
α_{12}	—	—	0.106	0.025	0.147	0.325
			(0.025)	(0.005)	(0.040)	(0.076)
β_2	—	—	0.952	0.926	0.943	0.681
			(0.010)	(0.012)	(0.013)	(0.067)
V_2	—	—	1.058	0.951	1.090	1.007
			(0.018)	(0.008)	(0.029)	(0.079)
π^2_∞	0	0	0.202	0.450	0.163	0.410
					(0.021)	(0.081)
$\mathrm{E}\sigma^2_{2t}$	—	—	0.353	0.038	0.428	0.157
μ_2	—	—	0	0	−0.040	−0.007
					(0.018)	(0.017)
α_{03}	—	—	—	0.005	—	0.003
				(0.002)		(0.001)
α_{13}	—	—	—	0.113	—	0.077
				(0.044)		(0.034)
β_3	—	—	—	0.963	—	0.969
				(0.013)		(0.011)
V_3	—	—	—	1.077	—	1.045
				(0.032)		(0.025)
π^3_∞	0	0	0	0.088	0	0.109
						(0.023)
$\mathrm{E}\sigma^2_{3t}$	—	—	—	0.447	—	0.449
μ_3	—	—	—	0	—	−0.040
						(0.030)

continued

Table 4 (*continued*)

	Normal	t	MSG (2)	MSG (3)	MNG (2)	MNG (3)
ν	—	4.171 (0.209)	—	—	—	—
P	1	1	$\begin{bmatrix} 0.873 & 0.499 \\ (0.016) & (0.061) \\ 0.127 & 0.501 \\ (0.016) & (0.061) \end{bmatrix}$	$\begin{bmatrix} 0.944 & 0.056 & 0.010 \\ (0.015) & (0.017) & (0.078) \\ 0.027 & 0.880 & 0.470 \\ (0.018) & (0.016) & (0.088) \\ 0.029 & 0.064 & 0.520 \\ (0.014) & (0.016) & (0.064) \end{bmatrix}$	$\pi_\infty \mathbf{1}_2'$	$\pi_\infty \mathbf{1}_3'$
δ	—	—	0.375 (0.061)	0.891 (0.028)	0	0

See Table 2 for explanations, but note the following modification: The mean Equation (19) is specified by an AR(1) process, that is, $c_t = a_0 + a_1\, r_{t-1}$.

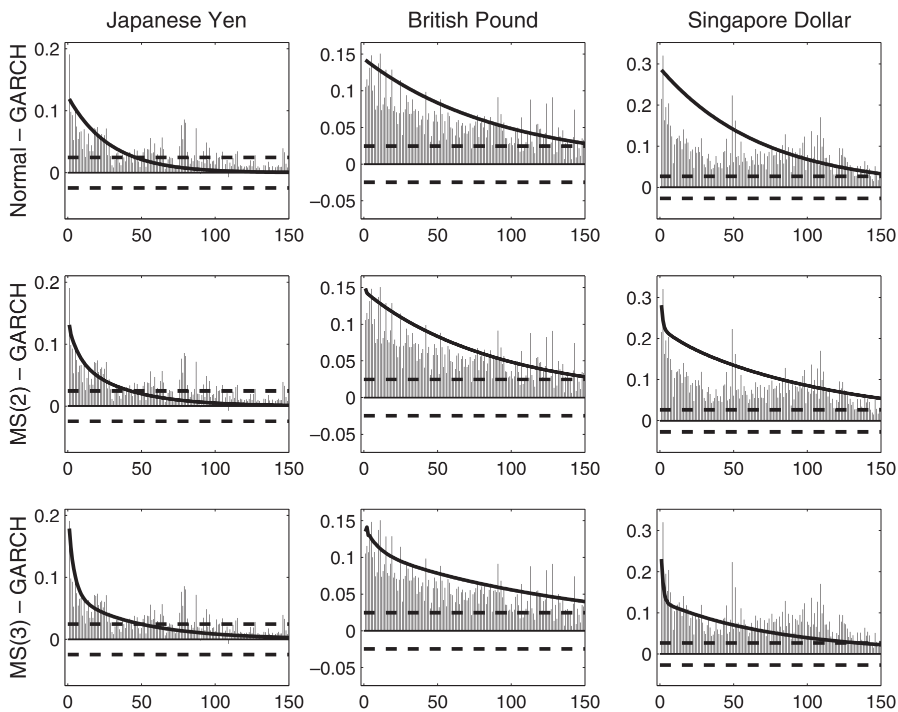

Figure 2 Shown are the sample autocorrelations of squared AR(0), AR(3), and AR(1) residuals for the returns of the Japanese yen, British pound, and Singapore dollar against the U.S. dollar, respectively, along with their theoretical counterparts implied by estimated GARCH models, as derived in Equation (48). Dashed lines represent the usual 95% one-at-a-time asymptotic confidence intervals associated with a white-noise process with finite second moments [cf. Anderson (1992)].

not imply a substitution of Markov for GARCH effects. Here the degree of Markov dependence, δ, is still lower than the persistence of shocks, $\rho(M)$, and the latter is as large as in the single-regime model. This, of course, does not imply that the arguments put forward in the introduction concerning the presence of spurious high persistence due to parameter instabilities do not apply. It should be kept in mind that the exponential decay at rate $\rho(M)$ dominates the autocorrelation function and volatility forecasts only asymptotically. As an illustration, the autocorrelation functions of the squared errors implied by the estimated MSG(3) models for the yen and the Singapore dollar, shown in Figure 2, are extremely different from those implied by the single-regime GARCH model, in that their decay is much faster at the beginning. Also, we observe that, in the high-volatility regimes, the effect of an *individual shock* dies out faster than in the tranquil regimes, as reflected in the relatively low β_j's in the turbulent components.

Part of the results reported in Tables 2–4 may also be explained by the inappropriateness of the Markov process to capture the parameter instabilities, or by the fact that the periods under study are rather long — too long perhaps even for a three-regime model to reflect the time variability of the process parameters.

3.3 Forecasting Performance

3.3.1 Evaluating the density forecasts. We assess the forecasting performance of the models by using transformed one-step-ahead forecast errors. Even if a multi-regime model were correctly specified, then standardized residuals would not be identically distributed, which prevents us from directly evaluating the distributional properties of the estimated residuals, $\hat{\epsilon}_t$. To circumvent this, the residuals are transformed by computing the conditional cdf values,

$$\hat{u}_t = \hat{F}(\hat{\epsilon}_t \mid \Psi_{t-1}), \quad t = 1, \ldots, T. \tag{20}$$

Under a correct specification, the transformed residuals are iid uniform [Rosenblatt (1952); see also Diebold, Gunther, and Tay (1998)]. A Pearson goodness-of-fit test may be employed to test this distributional property. However, as is well known, the results of such a test, when applied to a continuous distribution, can be extremely sensitive to the number of bins specified. Therefore we perform a second transformation, namely,

$$z_t = \Phi^{-1}(\hat{u}_t), \quad t = 1, \ldots, T, \tag{21}$$

where $\Phi(\cdot)$ is the standard normal cdf. The z_t are iid N(0,1) distributed if the underlying model is correct. Berkowitz (2001) shows that inaccuracies in the density forecast will be preserved in the transformed data. To test for the correct specification of skewness and kurtosis, we evaluate the quantities $T\gamma_1^2/6$ and $T(\gamma_2 - 3)^2/24$. Here, $\gamma_1 = T^{-1}\sum_t (z_t - \bar{z})^3 / \hat{\sigma}^3$ is the sample skewness, $\gamma_2 = T^{-1}\sum_t (z_t - \bar{z})^4 / \hat{\sigma}^4$ the sample kurtosis, $\hat{\sigma}$ the sample standard deviation, and $\bar{z}$ the sample mean of the transformed residuals of Equation (21). If the z_t are normal, then these statistics each have an asymptotic $\chi^2(1)$ distribution [see, e.g., Lehmann (1999, p. 344)]. In addition, we report the results of a joint test for zero mean, unit variance, zero skewness, and the absence of excess kurtosis, employing the likelihood ratio framework proposed by Berkowitz (2001). Under the alternative, we let z_t be distributed according to the skewed exponential power (SEP) distribution of Fernandez, Osiewalski, and Steel (1995), with density

$$f(z_t; \mu, \sigma, d, \theta) = K \begin{cases} \exp\left\{-\frac{1}{2}\left(\frac{|z_t - \mu|\theta}{\sigma}\right)^d\right\}, & \text{if } z_t < \mu \\ \exp\left\{-\frac{1}{2}\left(\frac{z_t - \mu}{\sigma\theta}\right)^d\right\}, & \text{if } z_t \geq \mu, \end{cases} \tag{22}$$

where $K = [\sigma(\theta + \theta^{-1})2^{1/d}d^{-1}\Gamma(d^{-1})]^{-1}$. This distribution nests the normal for $\theta = 1$ and $d = 2$. For $\theta < 1$ ($\theta > 1$), the density is skewed to the left (right), and is fat-tailed for $d < 2$. If $\hat{\mu}$, $\hat{\sigma}$, $\hat{d}$, and $\hat{\theta}$ are the values which maximize the log-likelihood

$$L(\mu, \sigma, d, \theta) = \sum_{t=1}^{T} \log f(z_t; \mu, \sigma, d, \theta), \tag{23}$$

then the likelihood ratio statistic,

$$\mathrm{LR_B} = 2[L(\hat{\mu}, \hat{\sigma}, \hat{d}, \hat{\theta}) - L(0, 1, 2, 1)], \tag{24}$$

is asymptotically distributed as $\chi^2(4)$ if the z_t are N(0, 1) distributed.

We also evaluate the models' forecasting performance with respect to the value-at-risk (VaR), a widely employed tool in risk management [see, e.g., Dowd (2002)]. The VaR for period t with shortfall probability ξ, denoted by $\mathrm{VaR}_t(\xi)$, implied by model $\mathcal{M}$ is defined by

$$\hat{F}_t^{\mathcal{M}}(\mathrm{VaR}_t(\xi)) = \xi, \tag{25}$$

where $\hat{F}_t^{\mathcal{M}}$ is the return-distribution function at time t predicted from model $\mathcal{M}$ using information up to time $t-1$. For a correctly specified model, we expect $100 \times \xi\%$ of the observed return values not to exceed the respective VaR forecast.

To formally test whether a model correctly estimates the risk inherent in a given financial position, that is, whether the empirical shortfall probability coincides with the specified shortfall probability, ξ, we use the likelihood-ratio test statistic [see, e.g., Kupiec (1995)]

$$\mathrm{LR_{VaR}} = -2(\log[\xi^x(1-\xi)^{T-x}] - \log[\hat{\xi}^x(1-\hat{\xi})^{T-x}]) \overset{asy}{\sim} \chi^2(1), \tag{26}$$

where T denotes the number of forecasts evaluated, x is the observed shortfall count, and $\hat{\xi} = x/T$ is the empirical shortfall probability.

Finally, in view of the conditional heteroskedasticity, it makes sense to examine whether the variance process is adequately described by the specified GARCH structures. Berkowitz (2001) makes use of the filtered and transformed residuals of Equation (21) for testing the dynamic properties of the conditional distribution. To this end we employ the Lagrange multiplier (LM) test of Engle (1982) to test for remaining ARCH dependencies in z_t. The relevant test statistic is computed by regressing z_t^2 on q lags, $z_{t-1}^2, \ldots, z_{t-q}^2$, and a constant and computing

$$\mathrm{LM}_{\mathrm{ARCH}}(q) = TR^2, \tag{27}$$

which is approximately $\chi^2(q)$ distributed, with R^2 denoting the coefficient of determination obtained for the regression.

3.3.2 Empirical results. To keep the effects of parameter uncertainty small, we estimate initial models from the first 3,000 observations and update the parameter vectors after five trading days, using an expanding window of data. Given the sample sizes for the three currencies, we can generate 3,335, 3,310 and 2,310 one-step-ahead forecasts, respectively, for the yen, the pound, and the Singapore dollar. The predictive performance of the models under investigation is shown in Tables 5–7. The upper section in each table reports the test results on the coverage of the VaR forecasts; the lower section provides results concerning the overall density forecasts.

As was to be expected, the single-regime GARCH models with normal innovations perform poorly for all three series. The conditional variance dynamics,

Table 5 Value-at-risk and density forecasts of competing GARCH models for the returns of the Japanese yen.

$100 \times \xi\%$	Normal	Student's t	MSG(2)	MSG(3)	MNG(2)	MNG(3)
0.5	1.439***	0.720*	0.780**	0.870***	0.750*	0.660
1.0	2.009***	1.439**	1.169	1.409**	0.990	0.930
2.5	3.448***	3.268***	2.969*	3.148**	2.369	2.429
5.0	5.247	5.937**	5.637*	5.397	5.037	4.918
10.0	8.576***	11.09**	11.12**	10.65	10.38	10.20
90.0	91.81***	90.13	90.13	90.40	89.27	89.87
95.0	95.41	94.93	95.08	95.29	94.33*	94.72
97.5	97.60	97.90	98.26***	97.99*	97.54	97.60
99.0	98.86	99.49***	99.52***	99.37**	99.13	99.04
99.5	99.22**	99.82***	99.76**	99.76**	99.73**	99.70*
Skewness	−0.454***	−0.135***	−0.173***	−0.157***	−0.055	−0.042
Kurtosis	2.379***	−0.202**	0.035	0.014	0.031	−0.045
$\mathrm{LR_B}$	225.0***	21.01***	19.03***	16.26***	3.790	4.731
$\mathrm{LM_{ARCH}}(1)$	0.089	1.023	2.887*	3.598*	1.669	3.729*
$\mathrm{LM_{ARCH}}(5)$	3.125	5.952	7.902	9.385*	7.352	9.549*
$\mathrm{LM_{ARCH}}(10)$	8.104	12.80	13.03	13.68	13.97	15.29

The upper part of the table reports the observed one-step-ahead percentage shortfall frequencies for given target probabilities ξ, as defined in Equation (25). The lower part reports test results on the distributional properties of the filtered one-step ahead forecast residuals, as defined by Equations (21) and (20). "Skewness" denotes the coefficient of skewness $\gamma_1 = \hat{m}_3/\hat{\sigma}^3$ and "kurtosis" the coefficient of excess kurtosis $\gamma_2 - 3 = \hat{m}_4/\hat{\sigma}^4 - 3$, where $\hat{m}_p = T^{-1}\sum_t (z_t - \bar{z})^p$; $\bar{z}$ is the sample mean; and $\hat{\sigma}$ is the sample standard deviation. Under normality, $T\gamma_1^2/6 \overset{\text{asy}}{\sim} \chi^2(1)$ and $T(\gamma_2 - 3)^2/24 \overset{\text{asy}}{\sim} \chi^2(1)$. $\mathrm{LR_B}$ is the likelihood ratio statistic of Equation (24); and $\mathrm{LM_{ARCH}}(q)$ is the Lagrange multiplier test for ARCH(q), as given by Equation (27). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6 Value-at-risk and density forecasts of competing GARCH models for the returns of the British pound.

$100 \times \xi\%$	Normal	Student's t	MSG(2)	MSG(3)	MNG(2)	MNG(3)
0.5	0.785**	0.423	0.363	0.332	0.272**	0.211***
1.0	1.420**	0.906	0.755	0.906	0.574***	0.604**
2.5	2.840	2.810	2.296	2.296	1.964**	2.205
5.0	4.592	5.227	4.592	4.773	4.350*	4.502
10.0	8.187***	9.789	9.728	9.698	9.245	9.517
90.0	91.93***	90.63	90.66	90.79	90.09	90.30
95.0	95.77**	95.35	95.83**	95.65*	95.32	95.20
97.5	97.43	97.46	97.95*	97.79	97.73	97.37
99.0	98.73	99.12	99.27*	99.21	99.18	99.00
99.5	99.15**	99.67	99.73**	99.67	99.70*	99.58
Skewness	−0.071*	−0.036	−0.047	−0.036	0.041	0.074*
Kurtosis	1.533***	−0.017	0.008	0.069	0.001	−0.014
LR_B	142.9***	3.599	6.153	5.185	5.846	4.713
$LM_{ARCH}(1)$	0.096	0.656	0.439	0.169	0.134	0.094
$LM_{ARCH}(5)$	2.100	2.289	3.782	2.360	3.903	4.006
$LM_{ARCH}(10)$	2.736	4.464	5.436	4.486	5.325	6.267

See the legend of Table 5 for explanations.

Table 7 Value-at-risk and density forecasts of competing GARCH models for the returns of the Singapore dollar.

$100 \times \xi\%$	Normal	Student's t	MSG(2)	MSG(3)	MNG(2)	MNG(3)
0.5	1.169***	0.606	0.693	0.649	0.476	0.476
1.0	1.602***	0.996	0.823	1.039	0.736	0.952
2.5	2.814	2.727	2.511	2.684	2.208	2.468
5.0	3.939**	4.675	4.848	4.805	4.416	4.589
10.0	7.186***	10.13	10.52	9.524	10.00	9.740
90.0	91.82***	89.05	89.22	89.96	88.74**	89.35
95.0	95.50	94.37	94.03**	94.55	93.94**	94.24
97.5	97.27	97.58	98.05*	97.75	97.23	97.45
99.0	98.57*	99.26	99.31	99.13	99.39**	99.22
99.5	99.00***	99.70	99.61	99.61	99.61	99.57
Skewness	−0.304***	−0.108**	−0.129**	−0.114**	−0.082	−0.081
Kurtosis	4.548***	−0.135	−0.057	−0.065	0.145	0.121
LR_B	246.6***	13.84***	11.76**	10.44**	8.673*	8.407*
$LM_{ARCH}(1)$	1.523	9.222***	0.045	0.001	4.567**	0.052
$LM_{ARCH}(5)$	5.082	15.77***	3.500	2.584	8.071	4.766
$LM_{ARCH}(10)$	7.592	18.24*	7.290	6.589	10.05	6.133

See the legend of Table 5 for explanations.

however, seem adequately captured, as the Lagrange multiplier test for ARCH does not reject the homoskedasticity of the transformed residuals of Equation (21).

Turning to the mixture models MSG and MNG, asymmetries appear to be important for the yen and the Singapore dollar, as the residuals of Equation (21) of the MSG models exhibit significant skewness, while those of the MNG do not. This coincides with the results of Mittnik and Paolella (2000), who find that, for modeling Asian exchange rates, employing a skewed extension of the *t* distribution improves density forecasts relative to the (symmetric) Student's *t*-GARCH of Bollerslev (1987). In particular, the ignorance of asymmetries also worsens the VaR forecasts of the MSG models for the yen in both the left and right tail. In contrast, the density forecasts for the yen returns of both MNG(2) and MNG(3) are satisfactory with respect to skewness, kurtosis, and VaR. Note, however, that both the residuals of MSG(3) and MNG(3) exhibit significant ARCH, indicating overparameterization of the three-regime models. The superior performance of model MNG(2) is worth mentioning because the estimation results over the whole period for the MSG(2) model indicate insignificant Markov dependence. This suggests the appropriateness of the multinomial mixture process for this currency.

All mixture models provide more or less accurate density forecasts for the pound, with a weak dominance of MSG(3). The best model for the Singapore dollar is MNG(3), as this is the only model that provides accurate VaR forecasts throughout and passes the tests for conditional skewness, kurtosis, and ARCH. This contrasts with the penalty-based criteria in Table 1, which clearly favor the MSG(3) model. It may still be true, however, that the Markov models capture important properties of the volatility process of this currency, given that they exhibit the lowest values of the LM test statistics of Equation (27) for ARCH, reported in the bottom part of Table 7.

For the currencies considered, the standardized residuals of all mixture models employed do not exhibit any significant excess kurtosis, indicating the appropriateness of the normal distribution in the multiregime framework.

Comparing the mixture models with the *t*-GARCH, the former are preferred for the yen and the Singapore dollar. Clearly, in contrast to the MNG process, the *t*-GARCH process cannot capture the asymmetries in the data, but more interestingly, also predicts too much kurtosis for the yen, which corresponds to the in-sample results in Table 1, and fails to pass the heteroskedasticity tests for the Singapore dollar. However, the *t*-GARCH process is a satisfactory model for the pound, at least if the final aim of modeling is the production of accurate density forecasts. In particular, the results suggest that the returns on the pound have a symmetric distribution.

4 CONCLUSION

A new Markov-switching GARCH model for modeling the volatility process of financial time series has been proposed. We provided a detailed discussion of the variance processes in multiregime GARCH models and argued that the disaggregation of the overall variance process offered by the proposed model is more appealing than in the existing variants. Maximum-likelihood estimation is straightforward using standard techniques for dealing with Markov switching. In addition, and in stark contrast to the common approach of Gray (1996), the analytical tractability of the new model allows derivation of stationarity conditions and dynamic properties. In particular, the autocorrelation function of the squared process was derived and employed to investigate the relation between two possible sources of volatility persistence, namely persistence due to parameter instability and persistence of shocks to the variance.

The proposed model has been applied to three exchange rate return series to demonstrate its practical relevance. With respect to penalty-based model selection criteria and out-of-sample density forecasting, it turned out that the independent multinomial-switching GARCH process with a skewed mixed normal conditional distribution is particularly well suited for modeling and predicting the exchange rates under study. For two of the three series, it also outperforms the popular t-GARCH model in forecasting the exchange rate density. Allowing for skewness was found to be crucial for obtaining accurate density and VaR forecasts for the Japanese yen and the Singapore dollar. The MN-GARCH model adequately captures skewness and other conditional and unconditional distributional properties of the return series. In addition, when estimating the models over the whole sample period, the Markov dependence was insignificant in the two- and three-regime models for the pound, and the two-regime model for the yen. On the other hand, the Markov process appears to capture the volatility properties of the Singapore dollar series somewhat better than the MN-GARCH process.

Finally, it is worthwhile to recall that all mixture models employed in this article are *mixtures of normal distributions*. They proved adequate for capturing the widely observed nonnormality of financial return series. For example, all mixture models successfully reproduced the excess kurtosis in the series under investigation. Thus it may often be the case that the specification of heavy-tailed distributions in the context of multiregime models is avoidable, and that normality is adequate when conditioning on both the regime and the return process. Based on the empirical findings, future work will consider how to effectively incorporate skewness into the Markov-switching GARCH model introduced herein.

APPENDIX

A.1 Moment Conditions

We derive the stationarity conditions of the model specified by Equations (6), (7), and (8), using the notation defined in Section 2. To do so we shall simultaneously consider $\sigma_t^{(2)}$ and the matrix of cross-products $\sigma_t^{(2)}\sigma_t^{(2)'}$. Straightforward calculation gives

$$\begin{aligned}\sigma_t^{(2)}\sigma_t^{(2)'} &= (\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^{(2)})(\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^{(2)})' \\ &= \alpha_0\alpha_0' + \alpha_0\alpha_1'\epsilon_{t-1}^2 + \alpha_1\alpha_0'\epsilon_{t-1}^2 + \alpha_0\sigma_{t-1}^{(2)'}\beta + \beta\sigma_{t-1}^{(2)}\alpha_0' \\ &\quad + \alpha_1\alpha_1'\epsilon_{t-1}^4 + \alpha_1\sigma_{t-1}^{(2)'}\epsilon_{t-1}^2\beta + \beta\sigma_{t-1}^{(2)}\epsilon_{t-1}^2\alpha_1' + \beta\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}\beta. \qquad (28)\end{aligned}$$

Define the $k(k+1)\times 1$ vector

$$\Sigma_t = \begin{bmatrix} \sigma_t^{(2)} \\ \mathrm{vec}(\sigma_t^{(2)}\sigma_t^{(2)'}) \end{bmatrix}. \tag{29}$$

Consider the conditional expectation of vector $\sigma_t^{(2)}$, given the information at time $t-\tau-1$. In general, the information at time t consists of the values of the process up to time t, $\Psi_t = \{\epsilon_t, \epsilon_{t-1}, \ldots\}$, and hence the vector $\sigma_{t+1}^{(2)}$, and a probability distribution $\pi_t = [\pi_t^1, \pi_t^2, \ldots, \pi_t^k]'$ over the state space S of $\{\Delta_t\}$. Define

$$\Pi = \begin{bmatrix} \Pi_1 \\ \Pi_2 \end{bmatrix} = \begin{bmatrix} \pi_{t-\tau-1} \otimes \sigma_{t-\tau}^{(2)} \\ \pi_{t-\tau-1} \otimes \mathrm{vec}(\sigma_{t-\tau}^{(2)}\sigma_{t-\tau}^{(2)'}) \end{bmatrix}, \quad A_t = \begin{bmatrix} A_t^1 \\ A_t^2 \end{bmatrix} = \begin{bmatrix} \pi_t \otimes \alpha_0 \\ \pi_t \otimes (\alpha_0 \otimes \alpha_0) \end{bmatrix}, \tag{30}$$

where $\pi_t = P^\tau\, \pi_{t-\tau}$, and

$$\mathbb{I} = \begin{bmatrix} I_k & \cdots & I_k & 0_{k\times k^2} & \cdots & 0_{k\times k^2} \\ 0_{k^2\times k} & \cdots & 0_{k^2\times k} & I_{k^2} & \cdots & I_{k^2} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_k' \otimes I_k & 0_{k\times k^3} \\ 0_{k^2\times k^2} & \mathbf{1}_k' \otimes I_{k^2} \end{bmatrix}, \tag{31}$$

where $\mathbf{1}_k$ is a k-dimensional column of ones. Conditional on Ψ_{t-2} and $\Delta_{t-1} = j$, we have

$$\mathrm{E}[\sigma_t^{(2)} \mid \Psi_{t-2}, \Delta_{t-1} = j] = \alpha_0 + (\alpha_1 e_j' + \beta)\sigma_{t-1}^{(2)}. \tag{32}$$

For the elements of Equation (28), we get, using $\mathrm{vec}(xy') = y \otimes x$ and $\mathrm{vec}(ABC) = (C' \otimes A)\,\mathrm{vec}(B)$,

$$\begin{aligned}
\mathrm{vec}(\alpha_0\alpha_0') &= \alpha_0 \otimes \alpha_0, \\
\mathrm{vec}(\alpha_0\sigma_{t-1}^{(2)'}\beta) &= (\beta \otimes \alpha_0)\sigma_{t-1}^{(2)}, \\
\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\alpha_0') &= (\alpha_0 \otimes \beta)\sigma_{t-1}^{(2)}, \\
\mathrm{E}[\mathrm{vec}(\alpha_0\alpha_1'\epsilon_{t-1}^2) \mid \Psi_{t-2}, \Delta_{t-1} = j] &= (\alpha_1 \otimes \alpha_0)e_j'\sigma_{t-1}^{(2)} = ((\alpha_1 e_j') \otimes \alpha_0)\sigma_{t-1}^{(2)}, \\
\mathrm{E}[\mathrm{vec}(\alpha_1\alpha_0'\epsilon_{t-1}^2) \mid \Psi_{t-2}, \Delta_{t-1} = j] &= (\alpha_0 \otimes \alpha_1)e_j'\sigma_{t-1}^{(2)} = (\alpha_0 \otimes (\alpha_1 e_j'))\sigma_{t-1}^{(2)}, \\
\mathrm{E}[\mathrm{vec}(\alpha_1\alpha_1'\epsilon_{t-1}^4) \mid \Psi_{t-2}, \Delta_{t-1} = j] &= 3\mathrm{vec}(\alpha_1\alpha_1')e_j'\sigma_{t-1}^{(4)} \\
&= \{3(\alpha_1 \otimes \alpha_1)\mathrm{vec}[\mathrm{diag}(e_j)]'\}\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}), \\
\mathrm{E}[\mathrm{vec}(\alpha_1\sigma_{t-1}^{(2)'}\epsilon_{t-1}^2\beta) \mid \Psi_{t-2}, \Delta_{t-1} = j] &= \mathrm{vec}(\alpha_1\sigma_{t-1}^{(2)'}e_j'\sigma_{t-1}^{(2)}\beta) \\
&= (\beta \otimes (\alpha_1 e_j'))\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}), \\
\mathrm{E}[\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\epsilon_{t-1}^2\alpha_1') \mid \Psi_{t-2}, \Delta_{t-1} = j] &= \mathrm{vec}(\beta\sigma_{t-1}^{(2)}e_j'\sigma_{t-1}^{(2)}\alpha_1') \\
&= ((\alpha_1 e_j') \otimes \beta)\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}), \\
\mathrm{vec}(\beta\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}\beta) &= (\beta \otimes \beta)\mathrm{vec}(\sigma_{t-1}^{(2)}\sigma_{t-1}^{(2)'}).
\end{aligned}$$

Letting

$$\Lambda = \begin{bmatrix} M & 0_{k^2 \times k^3} \\ R & Q \end{bmatrix}, \tag{33}$$

where M and Q are defined by Equations (10) and (14), respectively, and R is the $k^3 \times k^2$ matrix similar to M and Q, but with $k^2 \times k$-matrix elements

$$R_{ji} = p_{ji}[(\alpha_1 e_i') \otimes \alpha_0 + \alpha_0 \otimes (\alpha_1 e_i') + \alpha_0 \otimes \beta + \beta \otimes \alpha_0], \quad i, j = 1, \ldots, k,$$

then, as $\{\eta_t\}$ is an iid sequence of random variables,

$$\mathrm{E}[\Sigma_{t-\tau+1} \mid \Psi_{t-\tau-1}, \pi_{t-\tau-1}] = \mathbb{I}\Lambda\Pi + \mathbb{I}A_{t-\tau},$$
$$\mathrm{E}[\Sigma_{t-\tau+2} \mid \Psi_{t-\tau-1}, \pi_{t-\tau-1}] = \mathbb{I}\Lambda^2\Pi + \mathbb{I}(A_{t-\tau+1} + \Lambda A_{t-\tau}),$$
$$\vdots$$
$$\mathrm{E}[\Sigma_t \mid \Psi_{t-\tau-1}, \pi_{t-\tau-1}] = \mathbb{I}\Lambda^\tau\Pi + \mathbb{I}\sum_{i=0}^{\tau-1}\Lambda^i A_{t-1-i}. \tag{34}$$

As Λ is lower block-triangular, we have, from Equation (34),

$$\mathrm{E}[\sigma_t^{(2)} \mid \Psi_{t-\tau-1},\ \pi_{t-\tau-1}] = (\mathbf{1}_k' \otimes I_k)M^\tau\Pi_1 + (\mathbf{1}_k' \otimes I_k)\sum_{i=0}^{\tau-1} M^i A_{t-1-i}^1. \tag{35}$$

The first term on the right-hand side of Equation (35) tends to zero as $\tau \to \infty$, provided that $\rho(M) < 1$.

To see that the sum in the second term, $S = \sum_{i=0}^{\tau-1} M^i A_{t-1-i}^1$, converges to a limit that is independent of initial conditions, consider first the case $k = 2$ (we will see that the result readily generalizes). Then, $\pi_{t-1-i} = \pi_\infty + \delta^{\tau-1-i}\ (\pi_{t-\tau} - \pi_\infty)$, where $\delta = p_{11} + p_{22} - 1$ and $|\delta| < 1$ if P is irreducible and aperiodic [cf. Karlin and Taylor (1975, p. 78)]. Hence

$$S = \sum_{i=0}^{\tau-1} M^i A_\infty^1 + \sum_{i=0}^{\tau-1} M^i \delta^{\tau-1-i}(A_{t-\tau}^1 - A_\infty^1). \tag{36}$$

Because $A_{t-\tau}^1 - A_\infty^1$ is bounded, it suffices to demonstrate that $\sum_{i=0}^{\tau-1} M^i \delta^{\tau-1-i}$ approaches a zero matrix as τ goes to infinity. If $\rho(M) < 1$, then there is a matrix norm, denoted by $\|\cdot\|$, such that $\|M\| < 1$ [cf. Horn and Johnson (1985, p. 297)]. Using this norm and the norm equivalence on a finite-dimensional vector space, we can choose ξ with $\max\{|\delta|, \|M\|\} < \xi < 1$ such that

$$\left\| \sum_{i=0}^{\tau-1} M^i \delta^{\tau-1-i} \right\| \le \sum_{i=0}^{\tau-1} |\delta|^{\tau-1-i} \|M^i\| \le \sum_{i=0}^{\tau-1} |\delta|^{\tau-1-i} \|M\|^i < \sum_{i=0}^{\tau-1} \xi^{\tau-1} = \tau\xi^{\tau-1} \overset{\tau \to \infty}{\longrightarrow} 0. \tag{37}$$

For $k > 2$, the arguments are easily generalized by Perron's formula for matrix powers [cf. Gantmacher (1959, p. 116)]. Assuming that the process started in the infinite past with finite variance, we have

$$\lim_{\tau \to \infty} \mathrm{E}[\sigma_t^{(2)} \mid \Psi_{t-\tau-1}, \pi_{t-\tau-1}] = (\mathbf{1}_k' \otimes I_k)\left(\sum_{i=0}^{\infty} M^i\right) A_\infty^1 = (\mathbf{1}_k' \otimes I_k)(I_{k^2} - M)^{-1} A_\infty^1. \tag{38}$$

As this limit does not depend upon the initial conditions, Equation (38) represents the unconditional expectation $\mathrm{E}(\sigma_t^{(2)})$, as stated in Equation (12).

To derive Equation (13), consider all paths of the Markov chain starting from $\pi_{t-\tau}$ at time $t-\tau$ that arrive in state 1 at the $(\tau-1)^{\text{th}}$ iteration. The first element of the vector $M^\tau\Pi + \sum_{i=0}^{\tau-1} M^i A^1_{t-1-i}$ is the sum of the expected values of σ^2_{1t} given such a path, each summand multiplied with its respective path probability; the second element is the sum of expected values of σ^2_{2t} given a path that arrives in state 1 at the $(\tau-1)^{\text{th}}$ iteration, each value multiplied with its respective path probability. Analogous interpretations hold for the other elements of the vector. Thus

$$\mathrm{E}[\epsilon_t^2 \mid \Psi_{t-\tau-1}, \pi_{t-\tau-1}] = (\mathrm{vec}\ P)'\left(M^\tau\Pi_1 + \sum_{i=0}^{\tau-1} M^i A^1_{t-1-i}\right), \tag{39}$$

and Equation (13) follows. We derive closed-form volatility forecasts from Equation (39) below.

From Equation (34), the block-triangular structure of Equation (33) and the same line of arguments used to derive the convergence of $\mathrm{E}(\sigma_t^{(2)} \mid \Psi_{t-\tau-1})$, $\mathrm{E}(\sigma_t^{(2)}\sigma_t^{(2)'})$, and hence $\mathrm{E}(\epsilon_t^4)$ exists if and only if $\mathrm{E}(\sigma_t^{(2)})$ exists and the largest eigenvalue of the nonnegative matrix Q is smaller than one. An expression for $\mathrm{E}(\sigma_t^{(2)}\sigma_t^{(2)'})$, if it exists, can be obtained from $\mathrm{E}\,(\Sigma_t) = \mathbb{I}(I-\Lambda)^{-1} A_\infty$ and partitioned inversion of $I-\Lambda$. Thus define

$$\Gamma = (I-Q)^{-1}[A^2_\infty + R(I-M)^{-1}A^1_\infty]. \tag{40}$$

Let p_i, $i = 1, \ldots, k$, be the ith column of P, and

$$\omega = [\mathrm{vec}(\mathrm{diag}(p_1))', \ldots, \mathrm{vec}(\mathrm{diag}(p_k))']'. \tag{41}$$

The same argument that gave us Equation (39) shows that the unconditional fourth moment of ϵ_t, given it exists, can be calculated via

$$\mathrm{E}(\epsilon_t^4) = 3\omega'\Gamma. \tag{42}$$

A.2 Autocorrelation Function of the Squared Process

To derive the autocorrelation function of $\{\epsilon_t^2\}$, we introduce the following matrices:

$$\tilde{\sigma}^{(2)} = [\mathrm{diag}(p_1), \ldots, \mathrm{diag}(p_k)](I_{k^2} - M)^{-1}A^1_\infty, \tag{43}$$

$$\Theta = \begin{bmatrix} 3\alpha_1 e_1' + \beta & 0_{k\times k} & \cdots & 0_{k\times k} \\ 0_{k\times k} & 3\alpha_1 e_2' + \beta & \cdots & 0_{k\times k} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{k\times k} & 0_{k\times k} & \cdots & 3\alpha_1 e_k' + \beta \end{bmatrix}, \tag{44}$$

and

$$P^\star = \begin{bmatrix} p_{11}I_k & 0_{k\times k} & \cdots & 0_{k\times k} & \cdots & p_{k1}I_k & 0_{k\times k} & \cdots & 0_{k\times k} \\ 0_{k\times k} & p_{12}I_k & \cdots & 0_{k\times k} & \cdots & 0_{k\times k} & p_{k2}I_k & \cdots & 0_{k\times k} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0_{k\times k} & 0_{k\times k} & \cdots & p_{1k}I_k & \cdots & 0_{k\times k} & 0_{k\times k} & \cdots & p_{kk}I \end{bmatrix}. \tag{45}$$

From the discussion preceding Equation (39), the ith element of $\tilde{\sigma}^{(2)}$ is given by

$$\begin{aligned} &\sum_{j=1}^{k} p(\Delta_t = i \mid \Delta_{t-1} = j) p(\Delta_{t-1} = j) \mathrm{E}(\sigma_{it}^2 \mid \Delta_{t-1} = j) \\ &\quad = \sum_{j=1}^{k} \frac{p(\Delta_t = i \cap \Delta_{t-1} = j)}{p(\Delta_{t-1} = j)} p(\Delta_{t-1} = j) \mathrm{E}(\sigma_{it}^2 \mid \Delta_{t-1} = j \cap \Delta_t = i) \\ &\quad = p(\Delta_t = i) \sum_{j=1}^{k} p(\Delta_{t-1} = j \mid \Delta_t = i) \mathrm{E}(\sigma_{it}^2 \mid \Delta_{t-1} = j \cap \Delta_t = i) \\ &\quad = \pi_\infty^i \mathrm{E}(\sigma_{it}^2 \mid \Delta_t = i), \end{aligned} \tag{46}$$

where the last equation follows from the "tower property" of the conditional expectation, that is, $\mathrm{E}[\mathrm{E}(Y \mid X,Z) \mid X] = \mathrm{E}(Y \mid X)$ [cf. Grimmett and Stirzaker (2001, p. 69)]. Similarly, the ith $(k \times 1)$ subvector of the $(k^2 \times 1)$ vector $P^\star\Gamma$ has the interpretation $\pi_\infty^i \mathrm{E}(\sigma_{it}^2 \sigma_t^{(2)} \mid \Delta_t = i)$.

We have

$$\mathrm{E}(\epsilon_{t-\tau}^2 \epsilon_t^2) = \mathrm{E}(\eta_{t-\tau}^2 \sigma_{\Delta_{t-\tau},t-\tau}^2 \epsilon_t^2) = \mathrm{E}_\Delta \Big\{ \mathrm{E}[\eta_{t-\tau}^2 \sigma_{\Delta_{t-\tau},t-\tau}^2 \mathrm{E}(\epsilon_t^2 \mid \Psi_{t-\tau}, \Delta_{t-\tau}) \mid \Delta_{t-\tau}] \Big\}. \tag{47}$$

From Equations (35) and (39), we get, noting that knowledge of $\Delta_{t-\tau}$ implies $\pi_{t-\tau} = e_{\Delta_{t-\tau}}$,

$$\begin{aligned} \mathrm{E}(\epsilon_t^2 \mid \Psi_{t-\tau}, \Delta_{t-\tau}) &= (\mathrm{vec}\, P)' \left[M^{\tau-1}(e_{\Delta_{t-\tau}} \otimes \sigma_{t-\tau+1}^{(2)}) + \sum_{i=0}^{\tau-2} M^i \{(P^{\tau-i-1} e_{\Delta_{t-\tau}}) \otimes \alpha_0\} \right] \\ &= (\mathrm{vec}\, P)' \left[M^{\tau-1} \{ e_{\Delta_{t-\tau}} \otimes [\alpha_0 + (\eta_{t-\tau}^2 \alpha_1 e'_{\Delta_{t-\tau}} + \beta) \sigma_{t-\tau}^{(2)}] \} \right. \\ &\qquad \left. + \sum_{i=0}^{\tau-2} M^i \{(P^{\tau-i-1} e_{\Delta_{t-\tau}}) \otimes \alpha_0\} \right] \\ &= (\mathrm{vec}\, P)' \left[M^{\tau-1} \{ e_{\Delta_{t-\tau}} \otimes [(\eta_{t-\tau}^2 \alpha_1 e'_{\Delta_{t-\tau}} + \beta) \sigma_{t-\tau}^{(2)}] \} \right. \\ &\qquad \left. + \sum_{i=0}^{\tau-1} M^i \{(P^{\tau-i-1} e_{\Delta_{t-\tau}}) \otimes \alpha_0\} \right]. \end{aligned}$$

Substituting in Equation (47), we get

$$\begin{aligned}
\mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2) &= (\mathrm{vec}\ P)'M^{\tau-1}\mathrm{E}_\Delta\Big[\mathrm{E}(\{e_{\Delta_{t-\tau}}\otimes[(\eta_{t-\tau}^2\alpha_1 e'_{\Delta_{t-\tau}}+\beta)\sigma_{t-\tau}^{(2)}]\}\eta_{t-\tau}^2\sigma_{\Delta_{t-\tau},t-\tau}^2)\Big] \\
&\quad + (\mathrm{vec}\ P)'\mathrm{E}_\Delta\left[\mathrm{E}\left(\sum_{i=0}^{\tau-1}M^i\{(P^{\tau-i-1}e_{\Delta_{t-\tau}}\sigma_{\Delta_{t-\tau},t-\tau}^2)\otimes\alpha_0\}\right)\right] \\
&= (\mathrm{vec}\ P)'\left(M^{\tau-1}\Theta P^\star\Gamma+\sum_{i=0}^{\tau-1}M^i\left[(P^{\tau-i-1}\,\tilde{\sigma}^{(2)})\otimes\alpha_0\right]\right). \qquad (48)
\end{aligned}$$

Equation (48) can be used to compute $\mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2)$ for all values of τ, $\tau\geq 1$, and hence the autocorrelation function

$$r(\tau)=\frac{\mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2)-\mathrm{E}^2(\epsilon_t^2)}{\mathrm{E}(\epsilon_t^4)-\mathrm{E}^2(\epsilon_t^2)}. \qquad (49)$$

However, in case of a small number k of regimes, the computation of Equation (48) can be considerably simplified by using Perron's formula for the powers of P [cf. Gantmacher (1959, p. 116)]. We shall consider the case where all eigenvalues of P, $\psi_1,\ldots,\psi_k$, are distinct, and where M and P have no eigenvalues in common. Then,

$$P^t=\sum_{j=1}^k C_j\psi_j^t, \qquad (50)$$

where C_j are constant matrices, which can be computed from P and its eigenvalues [Gantmacher (1959)]. Using Equation (50) in Equation (48), we get

$$\begin{aligned}
\sum_{i=0}^{\tau-1}M^i[(P^{\tau-i-1}\,\tilde{\sigma}^{(2)})\otimes\alpha_0] &= \sum_{j=1}^k\psi_j^{\tau-1}\sum_{i=0}^{\tau-1}(M/\psi_j)^i[(C_j\,\tilde{\sigma}^{(2)})\otimes\alpha_0] \\
&= \sum_{j=1}^k\psi_j^{\tau-1}(I_{k^2}-M/\psi_j)^{-1}(I_{k^2}-(M/\psi_j)^\tau)[(C_j\,\tilde{\sigma}^{(2)})\otimes\alpha_0] \\
&= \sum_{j=1}^k(\psi_j I_{k^2}-M)^{-1}(\psi_j^\tau I_{k^2}-M^\tau)[(C_j\,\tilde{\sigma}^{(2)})\otimes\alpha_0]. \qquad (51)
\end{aligned}$$

In particular, in the practically relevant case of two regimes, we have $\psi_1=1$, $\psi_2=\delta=p_{11}+p_{22}-1$,

$$C_1=\begin{bmatrix}\pi_\infty^1 & \pi_\infty^1\\ \pi_\infty^2 & \pi_\infty^2\end{bmatrix},\quad C_2=\begin{bmatrix}\pi_\infty^2 & -\pi_\infty^1\\ -\pi_\infty^2 & \pi_\infty^1\end{bmatrix}, \qquad (52)$$

and $\pi_\infty^1=(1-p_{22})/(2-p_{11}-p_{22})$, $\pi_\infty^2=(1-p_{11})/(2-p_{11}-p_{22})$. By inserting these quantities into

$$\begin{aligned}
\mathrm{E}(\epsilon_{t-\tau}^2\epsilon_t^2) &= (\mathrm{vec}\ P)'(M^{\tau-1}\Theta P^\star\Gamma+(I_4-M)^{-1}(I_4-M^\tau)[(C_1\,\tilde{\sigma}^{(2)})\otimes\alpha_0] \\
&\quad +(\delta I_4-M)^{-1}(\delta^\tau I_4-M^\tau)[(C_2\,\tilde{\sigma}^{(2)})\otimes\alpha_0]), \qquad (53)
\end{aligned}$$

the autocorrelation can easily be calculated for each $\tau\geq 1$.

In general, if P has distinct eigenvalues, C_j, $j = 1, \ldots, k$, in Equation (50) is given by

$$C_j = \left(\prod_{i \neq j} (\psi_j - \psi_i) \right)^{-1} \mathrm{Adj}(\psi_j I_k - P),$$

where Adj(A) denotes the adjoint matrix of A [Gantmacher (1959, p. 116)], and if the eigenvalues are ordered such that $\psi_1 = 1$, $C_1 = \boldsymbol{\pi}_\infty \mathbf{1}_k'$, where $\mathbf{1}_k$ is a k-dimensional column of ones. However, note that $\mathrm{Adj}(\psi_j \ I_k - P)$ cannot be computed with the formula $(\psi_\mathrm{j} \ I_k - P)^{-1} \det(\psi_\mathrm{j} \ I_k - P)$, so that the genuine definition of the adjoint matrix must be used. Also note that to derive Equation (51), the assumptions of P having distinct eigenvalues, and P and M no common eigenvalues, are crucial. Hence it does not hold for the MNG process, for which a compact formula for $\mathrm{cov}(\epsilon_t^2, \epsilon_{t-\tau}^2)$ is provided in Haas, Mittnik, and Paolella (2004), while Alexander and Lazar (2004) show how it can be computed recursively.

A.3 Closed-Form Volatility Forecasts

Starting from Equation (39), and employing essentially the same methods used to derive the autocorrelation function, it can be shown that for the Markov-switching GARCH model with two regimes, the volatility forecast in closed form is

$$\begin{aligned} \mathrm{E}(\epsilon_t^2 \mid \Psi_{t-\tau-1}, \boldsymbol{\pi}_{t-\tau-1}) = \mathrm{vec}(P)' \{ & M^\tau \Pi_1 + (I - M)^{-1}(I - M^\tau)[(C_1 \boldsymbol{\pi}_{t-\tau}) \otimes \alpha_0] \\ & + (\delta I - M)^{-1}(\delta^\tau I - M^\tau)[(C_2 \boldsymbol{\pi}_{t-\tau}) \otimes \alpha_0] \}, \end{aligned} \tag{54}$$

with $\pi_{t-\tau} = P\pi_{t-\tau-1}$, and the generalization to k regimes being straightforward. Clearly, as the product $C_1 \ \pi_{\mathrm{t}-\tau} = \pi_\infty$ does not depend on $\pi_{\mathrm{t}-\tau}$, we have

$$\lim_{\tau \to \infty} \mathrm{E}(\epsilon_t^2 \mid \Psi_{t-\tau-1}, \boldsymbol{\pi}_{t-\tau-1}) = \mathrm{vec}(P)'(I - M)^{-1}(\boldsymbol{\pi}_\infty \otimes \alpha_0) = \mathrm{E}(\epsilon_t^2),$$

where the speed of convergence is governed by $\max\{\delta, \rho(M)\}$.

Received March 18, 2004; revised July 2, 2004; accepted July 5, 2004.

REFERENCES

Ahrens, R., and S. Reitz. (2004). "Heterogeneous Expectations in the Foreign Exchange Market: Evidence from the Daily Dollar/DM Exchange Rate." Forthcoming in *Journal of Evolutionary Economics*.

Akaike, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle." In B. N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado.

Alexander, C., and E. Lazar. (2003). "Symmetric Normal Mixture GARCH." ISMA Centre Discussion Papers in Finance 2003-09, Business School for Financial Markets, University of Reading.

Alexander, C., and E. Lazar. (2004). "Normal Mixture GARCH(1,1). Applications to Exchange Rate Modelling." ISMA Centre Discussion Papers in Finance 2004-06, Business School for Financial Markets, University of Reading.

Anderson, T. W. (1992). "The Asymptotic Distribution of Autocorrelation Coefficients." In K. V. Mardia (ed.), *The Art of Statistical Science, A Tribute to G. S. Watson*. Chichester: John Wiley & Sons.

Andreou, E., and E. Ghysels. (2002). "Detecting Multiple Breaks in Financial Market Volatility Dynamics." *Journal of Applied Econometrics 17*, 579–600.

Beine, M., Laurent, S., and C. Lecourt. (2003). "Official Central Bank Interventions and Exchange Rate Volatility: Evidence from a Regime-Swichting Analysis." *European Economic Review 47*, 891–911.

Bera, A. K., and M. L. Higgins. (1993). "ARCH Models: Properties, Estimation and Testing." *Journal of Economic Surveys 7*, 305–366.

Berkowitz, J. (2001). "Testing Density Forecasts, with Applications to Risk Management." *Journal of Business and Economic Statistics 19*, 465–474.

Bhar, R., and S. Hamori. (2004). "Empirical Characteristics of the Permanent and Transitory Components of Stock Return: Analysis in a Markov Switching Heteroskedasticity Framework." *Economics Letters 82*, 157–165.

Billio, M., and L. Pelizzon. (2000). "Value at Risk: A Multivariate Switching Regime Approach." *Journal of Empirical Finance 7*, 531–554.

Bollen, N. P. B., S. F. Gray, and R. E. Whaley. (2000). "Regime Switching in Foreign Exchange Rates: Evidence from Currency Option Prices." *Journal of Econometrics 94*, 239–276.

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

Bollerslev, T. (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics 69*, 542–547.

Bollerslev, T. (1988). "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process." *Journal of Time Series Analysis 9*, 121–131.

Bollerslev, T., R. F. Engle, and D. B. Nelson. (1994). "ARCH Models." In R. F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: Elsevier Science B.V.

Brunetti, C., R. S. Mariano, C. Scotti, and A. H. H. Tan. (2003). "Markov Switching GARCH Models of Currency Crises in Southeast Asia." Working Paper 03-008, Penn Institute for Economic Research.

Cai, J. (1994). "A Markov Model of Switching-Regime ARCH." *Journal of Business and Economic Statistics 12*, 309–316.

Chen, M., and H. Z. An. (1998). "A Note on the Stationarity and the Existence of Moments of the GARCH Model." *Statistica Sinica 8*, 505–510.

Dewachter, H. (2001). "Can Markov Switching Models Replicate Chartist Profits in the Foreign Exchange Market?" *Journal of International Money and Finance 20*, 25–41.

Diebold, F. X. (1986). "Modeling the Persistence of Conditional Variances: A Comment." *Econometric Reviews 5*, 51–56.

Diebold, F. X., T. A. Gunther, and A. S. Tay. (1998). "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review 39*, 863–883.

Diebold, F. X., and A. Inoue. (2001). "Long Memory and Regime Switching." *Journal of Econometrics 105*, 131–159.

Dowd, K. (2002). *Measuring Market Risk*. Chichester: John Wiley & Sons.

Dueker, M. J. (1997). "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility." *Journal of Business and Economic Statistics 15*, 26–34.

Engel, C. (1994). "Can the Markov Switching Model Forecast Exchange Rates?" *Journal of International Economics 36*, 151–165.

Engel, C., and J. D. Hamilton. (1990). "Long Swings in the Dollar: Are they in the Data and do Markets know it?" *American Economic Review 80*, 689–713.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica 50*, 987–1008.

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

Fernandez, C., J. Osiewalski, and M. F. J. Steel. (1995). "Modelling and Inference With V-Spherical Distributions." *Journal of the American Statistical Association 90*, 1331–1340.

Francq, C., M. Roussignol, and J.-M. Zakoïan. (2001). "Conditional Heteroskedasticity Driven by Hidden Markov Chains." *Journal of Time Series Analysis 22*, 197–220.

Francq, C., and J.-M. Zakoïan. (2001). "Stationarity of Multivariate Markov-Switching ARMA Models." *Journal of Econometrics 102*, 339–364.

Frankel, J. A., and K. A. Froot. (1988). "Chartists, Fundamentalists and the Demand for Dollars." *Greek Economic Review 10*, 49–102.

Gantmacher, F. (1959). *Matrix Theory*, vol. 1. New York: Chelsea Publishing.

Geweke, J., and G. Amisano. (2003). "Compound Markov Mixture Models with Applications in Finance," working paper, University of Iowa.

Granger, C. W. J., and T. Teräsvirta. (1999). "A Simple Nonlinear Time Series Model with Misleading Linear Properties." *Economics Letters 62*, 161–165.

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

Grimmett, G., and D. Stirzaker. (2001). *Probability and Random Processes*. New York: Oxford University Press.

Haas, M., S. Mittnik, and M. S. Paolella. (2004). "Mixed Normal Conditional Heteroskedasticity." *Journal of Financial Econometrics 2*, 211–250.

Hafner, C. M. (2003). "Fourth Moment Structure of Multivariate GARCH Models." *Journal of Financial Econometrics 1*, 26–54.

Hamilton, J. D. (1988). "Rational-Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control 12*, 385–423.

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. (1990). "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics 45*, 39–70.

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.

He, C., and T. Teräsvirta. (1999a). "Fourth Moment Structure of the GARCH(p,q) Process." *Econometric Theory 15*, 824–846.

He, C., and T. Teräsvirta. (1999b). "Properties of Moments of a Family of GARCH Processes." *Journal of Econometrics 92*, 173–192.

Horn, R. A., and C. R. Johnson. (1985). *Matrix Analysis*. Cambridge: Cambridge University Press.

Huisman, R., K. Koedijk, C. Kool, and F. Palm. (2002). "The Tail-Fatness of FX Returns Reconsidered." *De Economist 150*, 299–312.

Karanasos, M. (1999). "The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH Model." *Journal of Econometrics 90*, 63–76.

Karanasos, M., Z. Psaradakis, and M. Sola. (2004). "On the Autocorrelation Properties of Long-Memory GARCH Processes." *Journal of Time Series Analysis 25*, 265–281.

Karlin, S., and H. M. Taylor. (1975). *A First Course in Stochastic Processes, 2nd ed.* New York: Academic Press.

Kim, D., and S. J. Kon. (1999). "Structural Change and Time Dependence in Models of Stock Returns." *Journal of Empirical Finance 6*, 283–308.

Klaassen, F. (2002). "Improving GARCH Volatility Forecasts with Regime-Switching GARCH." *Empirical Economics 27*, 363–394.

Kupiec, P. H. (1995). "Techniques for Verifying the Accuracy of Risk Management Models." *Journal of Derivatives 3*, 73–84.

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.

Lanne, M., and P. Saikkonen. (2003). "Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes." *Journal of Financial Econometrics 1*, 96–125.

Lastrapes, W. D. (1989). "Exchange Rate Volatility and U.S. Monetary Policy: An ARCH Application." *Journal of Money, Credit and Banking 21*, 66–77.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag.

Ling, S., and M. McAleer. (2002). "Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models." *Econometric Theory 18*, 722–729.

Maheu, J. M., and T. H. McCurdy. (2000). "Identifying Bull and Bear Markets in Stock Returns." *Journal of Business and Economic Statistics 18*, 100–112.

McLachlan, G. J., and D. Peel. (2000). *Finite Mixture Models*. New York: John Wiley & Sons.

Mikosch, T., and C. Starica. (2004). "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects." *Review of Economics and Statistics 86*, 378–390.

Mittnik, S., and M. S. Paolella. (2000). "Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates." *Journal of Forecasting 19*, 313–333.

Morana, C. (2002). "IGARCH Effects: An Interpretation." *Applied Economics Letters 9*, 745–748.

Newcomb, S. (1980). "A Generalized Theory of the Combination of Observations so as to Obtain the Best Result." In S. M. Stigler (ed.), *American Contributions to Mathematical Statistics in the Nineteenth Century*, vol. 2. New York: Arno Press.

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

Perez-Quiros, G., and A. Timmermann. (2001). "Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities." *Journal of Econometrics 103*, 259–306.

Rosenblatt, M. (1952). "Remarks on a Multivariate Transformation." *Annals of Mathematical Statistics 23*, 470–472.

Ryden, T., T. Teräsvirta, and S. Åsbrink. (1998). "Stylized Facts of Daily Return Series and the Hidden Markov Model." *Journal of Applied Econometrics 13*, 217–244.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics 6*, 461–464.

Susmel, R. (2000). "Switching Volatility in International Equity Markets." *International Journal of Finance and Economics 5*, 265–283.

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

Turner, C. M., R. Startz, and C. R. Nelson. (1989). "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market." *Journal of Financial Economics 25*, 3–22.

Vigfusson, R. (1997). "Switching between Chartists and Fundamentalists: A Markov Regime-Switching Approach." *International Journal of Finance and Economics 2*, 291–305.

Vlaar, P. J. G., and F. C. Palm. (1993). "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps." *Journal of Business and Economic Statistics 11*, 351–360.

Wong, C. S., and W. K. Li. (2001). "On a Mixture Autoregressive Conditional Heteroscedastic Model." *Journal of the American Statistical Association 96*, 982–995.

Yang, M. (2000). "Some Properties of Vector Autoregressive Processes with Markov-Switching Coefficients." *Econometric Theory 16*, 23–43.

Zaffaroni, P. (2004). "Stationarity and Memory of ARCH(∞) Models." *Econometric Theory 20*, 147–160.