

Volatility Dynamics under Duration-Dependent Mixing*

John M. Maheu
University of Alberta
Dept. of Economics
john.maheu@ualberta.ca

Thomas H. McCurdy
University of Toronto
Rotman School of Management
tmccurdy@mgmt.utoronto.ca

first draft: September 1999; this draft: May 2000

Abstract

This paper proposes a discrete-state stochastic volatility model with duration-dependent mixing. The latter is directed by a high-order Markov chain with a sparse transition matrix. As in the standard first-order Markov switching model, this structure can capture turning points and shifts in volatility due, for example, to policy changes or news events. However, the duration-dependent Markov switching model can also exploit the persistence associated with volatility clustering. To evaluate the contribution of duration dependence, we compare with a benchmark Markov switching ARCH model. The empirical distribution generated by our proposed structure is assessed using interval forecasts and density forecasts. Implications for areas of the distribution relevant for risk management are also assessed.

*We are indebted to two anonymous referees and the Editor, Franz Palm, for very helpful suggestions. We are also grateful to: participants at the *Limberg Institute of Financial Economics & Journal of Empirical Finance Conference on Risk Management* held at Albufeira, Portugal, November 1999; participants at the *Canadian Econometrics Study Group* held at the University of Montreal, September, 1999; and Clifford Ball, Adolf Buse, Peter Christoffersen, Toby Daghish, Jin-Chuan Duan, Ronald Huisman, Chang-Jin Kim, Pok-sang Lam, Simon van Norden, and Ilias Tsiakas for helpful comments and discussions. We also thank Werner Antweiler for providing data. Both authors thank the Social Sciences and Humanities Research Council of Canada for financial support.

1 Introduction

To manage risk associated with uncertain outcomes, one relies on forecasts of the distribution of k -period-ahead returns. The popular value-at-risk (VaR) calculation of the loss that could occur with a specified confidence level over a given holding period, focuses on a single quantile of that forecasted distribution. There are several potential questions that would interest risk managers. Firstly, are the VaR numbers correct on average, or more generally, is the maintained unconditional distribution adequate? Secondly, given available information does one obtain accurate VaR numbers at each point of time, In other words is the maintained conditional distribution adequate? Finally, if the answer to either of these questions is no, how might the risk manager's model be improved? Implicit in these questions are difficult issues related to evaluation of the adequacy of alternative models, and the correct specification of the conditional and unconditional distribution of returns.

A first step to forecasting the distribution (or some region thereof) of future returns, is to investigate stylized facts concerning the statistical properties of returns for the asset or portfolio in question. Short-run price dynamics in financial markets generally exhibit heteroskedasticity and leptokurtosis. For the case of log price changes in foreign exchange markets, which is the application in this paper, there is also some evidence of convergence to a Gaussian distribution under time aggregation.¹

The next step might be to postulate a probability model which can replicate such stylized facts. Since short-run returns are not IID, attention to the conditional density is warranted. One needs to postulate and estimate a probability model for the law of motion of returns over time. For currencies at weekly frequencies, strong volatility dependence combined with largely serially uncorrelated returns suggests that, for modeling purposes, the main focus should be on the conditional volatility dynamics. Volatility clustering implies predictability in the conditional variances and covariances. This empirical fact should contribute to risk management strategies.

In this paper we concentrate on the choice of conditional volatility dynamics and ask what features a particular specification can deliver in terms of the conditional and unconditional distributions. We propose a model that emphasizes discrete changes in the level of volatility and introduces duration dependence to accommodate volatility clustering. The features of this model are compared with two benchmark models that do not include the duration structure. Of interest is whether a particular specification can simultaneously capture important properties in the conditional and the unconditional distributions.

Alternative models have been proposed to capture and forecast time-varying volatility for financial returns. For example, the generalized ARCH

¹There are many papers which report stylized facts for exchange rates, including Boothe and Glassman (1987), Diebold (1988), Engle and Hamilton (1990), Hsieh (1989), Kaehler and Marnet (1993).

class (Engle (1982), Bollerslev (1986)) is based on an ARMA function of past innovations. These models have become the workhorse for parameterizing intertemporal dependence in the conditional variance of speculative returns.² Markov switching models for mixture distributions in which draws for component distributions follow a first-order Markov chain (Lindgren (1978), Hamilton (1988)) have also been applied to volatility dynamics.³ Although such discrete-mixtures models can generate most of the stylized facts of daily returns series (Rydén, Teräsvirta, and Åsbrink (1998)), there is some evidence (for example, Hamilton (1988), Pagan and Schwert (1990)) that the first-order Markov model with constant transition probabilities is inadequate with respect to capturing all of the volatility dependence.

This paper provides an alternative to the switching ARCH model introduced by Cai (1994) and Hamilton and Susmel (1994). Our alternative approach is a discrete-state stochastic volatility model which incorporates a parsimonious high-order Markov chain to allow for duration dependence. As in the standard first-order Markov switching (MS) model, this structure is useful for capturing shifts and turning points in volatility that are difficult to accommodate with the ARMA structure implicit in GARCH. However, unlike the standard model, a duration-dependent Markov switching (DDMS) model (Durland and McCurdy (1994), Maheu and McCurdy (2000), Lam (1997)) is particularly suited to exploiting the persistence associated with volatility clustering. This is achieved by several important features of our specification. Firstly, the duration variable provides a parsimonious parameterization of potential high-order dependence. Secondly, unlike the GARCH case, persistence is permitted to be time varying by allowing the duration of a state to affect the transition probabilities. Thirdly, including duration as a conditioning variable in the conditional variance specification allows the model to capture a broad range of volatility levels.

Markov switching (MS) models generate mixtures of distributions. These models can be motivated from information flows into the market. The application in this paper is to log-differences in foreign exchange rates. Exchange rates are good candidates for mixtures models because an exchange rate is a relative price which will be influenced by policy changes and news arrivals. News theories of price changes attribute an important role to the *current* innovation. In this case, a stochastic volatility component may be an important addition to time-varying but deterministic representations of volatility as in GARCH. The DDMS model is a discrete-time, discrete-state, stochastic volatility model. Therefore, our approach provides a straightforward method of obtaining maximum likelihood estimates for stochastic volatility with the added features discussed above.

Like standard first-order MS models, and other time dependent discrete-

²See, for example, a recent survey by Palm (1996) and references therein.

³For example, see Pagan and Schwert (1990), Kaehler and Marnet (1993), Klaassen (1998), Kim, Nelson, and Startz (1998), Taylor (1999), and Timmermann (2000).

mixture models, we model the stochastic process governing a switch from one volatility state to another. Incorporating the possibility of a discrete change in the level of volatility can have a substantial effect on the implied conditional and unconditional distributions. This may contribute to an improved fit, particularly with respect to the tails of the distribution. Further, unlike standard MS models, allowing the transition probabilities to be duration dependent may improve the model's ability to capture intertemporal dependence. In addition, our approach models the evolution of volatility within each state. The DDMS model considered in this study is only a two-state model, however it acts like a large N state model in that it can capture a broad range of volatility levels through conditioning on duration in the conditional variance.

In section 2, we discuss the data and associated descriptive statistics. Section 3 summarizes the candidate models for volatility dynamics compared in this paper. Particular emphasis is applied to comparing the stochastic properties inherent in the forecasts implied by each. This is followed by model estimates in Section 4. Adequacy of the conditional distributions implied by those estimates is reported in Section 5. For these assessments, we apply both interval forecasts (Christoffersen (1998)) and density forecasts (Diebold, Gunther, and Tay (1998), Berkowitz (1999)). In section 6 we use simulation methods to investigate the properties of the unconditional distributions implied by the estimates of the alternative volatility models. Finally, section 7 provides a brief summary of the features of the data that are matched by our alternative parameterizations.

2 Data and Descriptive Statistics

Let e_t denote the spot price in units of foreign currency for 1 US dollar. In this paper, we report results for the Deutsche mark (DEM-USD), and British pound (GBP-USD). Define y_t as the scaled log-difference of e_t , or the continuously compounded percentage return from holding a \$1-equivalent of foreign currency for a week,

$$y_t = 100 \log \frac{e_t}{e_{t-1}} \quad (2.1)$$

Information available to the econometrician at time t is $\Omega_t = \{y_t, y_{t-1}, \dots\}$. Sample size is $T = 1304$ covering the time period from 1974/01/02 to 1998/12/23.

Table 1 reports descriptive statistics for y_t associated with each of the currencies. Unconditional mean returns are insignificantly different from zero, there is no significant skewness, and excess kurtosis is significant for both currencies but larger for the GBP-USD case. Using a modified Ljung-Box portmanteau statistic, there is no significant serial correlation in returns but very significant serial dependence associated with squared returns. Figure 1 plots the levels e_t and log price changes y_t for the two currencies.

3 Alternative Parameterizations of Volatility

In this section, alternative parameterizations of conditional volatility are discussed. As a reference point, we begin by briefly summarizing the popular ARCH and a Markov switching ARCH model (MS-ARCH). Then duration dependence is introduced in a discrete-state parameterization of volatility (the DDMS model) by allowing the conditional transition probabilities, and also the state-specific levels of volatility, to be functions of duration. The final subsection discusses some salient differences between the alternative parameterizations. In particular we emphasize the potential impact of the duration-dependent components on volatility dynamics and forecasts.

3.1 GARCH

To introduce our discussion of alternative parameterizations of volatility, consider the popular generalized ARCH model, referred to as GARCH(1,1), which is defined as⁴

$$\sigma_t^2 = \omega_t + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (3.1)$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1) \quad (3.2)$$

where ϵ_t is the innovation to the process y_t parameterized as,

$$y_t = \mu + \phi y_{t-1} + \epsilon_t. \quad (3.3)$$

For the purpose of residual-based diagnostic tests, the standardized residuals are formed using estimates of ϵ_t/σ_t .

Note that except for the non-negativity constraint on σ_t^2 , volatility is a continuous variable. Also, in this formulation, conditional volatility σ_t^2 is time-varying but deterministic given the information set Ω_{t-1} . As the acronym ARCH implies, this model parameterizes volatility as autoregressive conditional heteroskedasticity. In other words, it allows volatility clustering.

A re-arrangement of (3.1) implies squared innovations follow an ARMA model, as in,

$$(1 - \alpha L - \beta L)\epsilon_t^2 = \omega + (1 - \beta L)(\epsilon_t^2 - \sigma_t^2)$$

where L is the lag operator. Therefore, many of the properties of the stationary ARMA model, such as exponential decay rates, are imposed by this GARCH model on the squared innovations for returns. Furthermore, ARMA models are not well suited to capturing discrete jumps (up or down) in volatility. This potential shortcoming of the plain vanilla GARCH parameterization in (3.1) is one of the motivations for exploring alternative models that can capture discrete regime changes in volatility.

⁴Asymmetries and seasonal effects are not expected to be important components of foreign exchange returns at weekly frequencies.

3.2 Markov switching ARCH model (MS-ARCH)

One possibility is to combine an ARCH specification with a discrete-state MS model in which the directing process is a first-order Markov chain. Such models were introduced by Cai (1994) and Hamilton and Susmel (1994).⁵ Volatility in a MS model is assumed to be stochastic and driven by an unobserved or hidden variable. Unlike the conventional stochastic volatility model, the assumption that the unobserved state variable is governed by a finite-state Markov chain makes estimation straightforward using maximum likelihood methods.⁶

As in Hamilton (1988), Markov switching models assume the existence of an unobserved discrete-valued variable S_t that determines the dynamics of y_t . Usually S_t is directed by a first-order Markov chain. Formally, assume the existence of a discrete-valued variable S_t that indexes the unobserved states. Our parameterization of an AR(1), MS-ARCH(p) model follows. In this case, log-differences of exchange rates are assumed to follow

$$y_t = \mu + \phi y_{t-1} + \epsilon_t \quad (3.4)$$

$$\epsilon_t = \sigma(S_t) z_t, \quad z_t \sim N(0, 1), \quad S_t = 1, 2 \quad (3.5)$$

$$\sigma_t^2(S_t) = \omega(S_t) + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 \quad (3.6)$$

$$P(S_t = 1 | S_{t-1} = 1) = \frac{\exp(\gamma_1(1))}{1 + \exp(\gamma_1(1))}, \quad (3.7)$$

$$P(S_t = 2 | S_{t-1} = 2) = \frac{\exp(\gamma_1(2))}{1 + \exp(\gamma_1(2))}. \quad (3.8)$$

Note that regime switches are assumed to affect the intercept of the conditional variance, that we have postulated two alternative states for the regime-switching component of volatility, and that the transition probabilities are parameterized using the logistic function. This hybrid of an ARCH and a MS component is intended to capture volatility clustering as well as occasional discrete shifts in volatility. The addition of an ARCH structure in this model presents no significant changes to the basic MS model, and therefore construction of the likelihood and filter follow the usual methods as detailed in Hamilton (1994). In this application the unconditional probabilities were used to startup the filter.

⁵Estimation of a switching GARCH model is intractable since the entire history of state variables enters the likelihood. Gray (1996) proposes a feasible switching GARCH model that avoids this problem. Due to the complexity of our extensions discussed below, we use an MS-ARCH structure in this paper.

⁶In general the stochastic volatility model requires simulation methods to evaluate the likelihood. A recent survey is Ghysels, Harvey, and Renault (1996).

3.3 Regime Switching with Duration Dependence

A first-order Markov chain combined with an ARCH structure should be adequate to capture volatility clustering.⁷ However, there may be benefits to exploring high-order Markov chains. In the duration-dependent Markov switching (DDMS) model (Durland and McCurdy (1994), Maheu and McCurdy (2000), Lam (1997)) the probability of a regime change is a function of the previous state S_{t-1} as well as the duration of the previous state S_{t-1} .⁸

Besides the discrete-valued variable S_t , define duration as a discrete-valued variable D_t which measures the length of a run of realizations of a particular state. To make estimation tractable we set the memory of duration to τ .⁹ This implies that the duration of S_t is

$$D_t = \min(D_{t-1}I(S_t, S_{t-1}) + 1, \tau) \quad (3.9)$$

where the indicator function $I(S_t, S_{t-1})$ is 1 for $S_t = S_{t-1}$ and 0 otherwise. That is, D_t is unobserved but is determined from the history of S_t . Therefore, both will be inferred by the filter which is summarized below. Realizations of the random variables S_t and D_t are referred to as s_t and d_t respectively.

This model allows both state variables, S_t and D_t , to affect the transition probabilities between volatility states. The probabilities are,

$$\begin{aligned} P_{11}(D_{t-1}) &\equiv P(S_t = 1 | S_{t-1} = 1, D_{t-1} = d_{t-1}) \\ &= \frac{\exp(\gamma_1(1) + \gamma_2(1)d_{t-1})}{1 + \exp(\gamma_1(1) + \gamma_2(1)d_{t-1})} \end{aligned} \quad (3.10)$$

as the conditional probability of staying in state 1 given that we have been in state 1 for d_{t-1} periods; and

$$\begin{aligned} P_{22}(D_{t-1}) &\equiv P(S_t = 2 | S_{t-1} = 2, D_{t-1} = d_{t-1}) \\ &= \frac{\exp(\gamma_1(2) + \gamma_2(2)d_{t-1})}{1 + \exp(\gamma_1(2) + \gamma_2(2)d_{t-1})} \end{aligned} \quad (3.11)$$

as the conditional probability of staying in state 2 given that we have been in state 2 for d_{t-1} periods. Note that $(\gamma_1(1), \gamma_2(1))$ and $(\gamma_1(2), \gamma_2(2))$ are parameters associated with states 1 and 2 respectively.

The conditional probability of a state change, given the state has achieved a duration d , is the hazard function. Since there are two states in this

⁷Hamilton (1988), Pagan and Schwert (1990), and Timmermann (2000) (footnote 11), suggest that a first-order MS model without the ARCH structure may not be adequate to capture serial dependence in the conditional variance.

⁸Kim and Nelson (1998) allow for duration dependence in a state space model. Filardo (1994) and Perez-Quiros and Timmermann (1999) allow transition probabilities to be a function of exogenous processes.

⁹While it is possible to estimate τ , for example using a grid search, in this paper we set $\tau = 25$, a number large enough to ensure that all the duration effects on transition probabilities have been captured.

application, this conditional probability of switching from state i to state j , given that we have been in state i for d_{t-1} periods, can be written

$$\begin{aligned} P_{ij}(D_{t-1}) &\equiv P(S_t = j | S_{t-1} = i, D_{t-1} = d_{t-1}) = 1 - P_{ii}(D_{t-1}) \\ &= \frac{1}{1 + \exp(\gamma_1(i) + \gamma_2(i)d_{t-1})}, \quad i, j = 1, 2, i \neq j. \end{aligned} \quad (3.12)$$

A decreasing hazard function is referred to as negative duration dependence while an increasing hazard function is positive duration dependence. The effect of duration on the hazard function is uniquely summarized by the parameters $\gamma_2(i)$ $i = 1, 2$. In particular, for state i , $\gamma_2(i) < 0$ implies positive duration dependence, $\gamma_2(i) = 0$ implies no duration effect and $\gamma_2(i) > 0$ implies negative duration dependence. For example, if state 2 displays negative duration dependence and the market persists in state 2, then the probability of staying in state 2 increases over time.

Given the state dynamics in Equations 3.9, 3.10, 3.11 and 3.12, for this discrete-state parameterization of volatility, log-differences of exchange rates are assumed to follow

$$y_t = \mu + \phi y_{t-1} + \epsilon_t \quad (3.13)$$

$$\epsilon_t = \sigma_t(S_t, D_t) z_t, \quad z_t \sim N(0, 1), \quad S_t = 1, 2 \quad (3.14)$$

$$\sigma_t(S_t, D_t) = (\omega(S_t) + \zeta(S_t)D_t)^2 \quad (3.15)$$

With this parameterization, the latent state affects the level of volatility directly, as indicated by $\omega(S_t)$, while the duration of the state is also allowed to affect the dynamics of volatility within each state through the function $\zeta(S_t)D_t$. For example, if $\zeta(1)$ is positive and we persist in state one, then conditional volatility is increasing since D_t is increasing. Squaring the term in brackets in (3.15) serves two purposes: first the standard deviation is restricted to be non-negative, and second it allows duration to have a nonlinear affect in the second moment.

3.4 The Filter

Volatility in this model is unobservable with respect to the information set. As shown in Hamilton (1994), inference regarding the latent variable S_t can be constructed recursively. In a similar fashion, inference regarding both S_t and D_t can be computed. Define, $f(\cdot|\cdot)$ as the conditional density of the normal distribution. The filter provides optimal inference for the unobserved variables given time t information. For $S_t = 1, 2$ and $1 \leq D_t \leq \tau$ we have

$$P(S_t = s_t, D_t = d_t | \Omega_t) = \frac{f(y_t | S_t = s_t, D_t = d_t, \Omega_{t-1}) P(S_t = s_t, D_t = d_t | \Omega_{t-1})}{P(y_t | \Omega_{t-1})}$$

where

$$\begin{aligned} P(S_t = s_t, D_t = d_t | \Omega_{t-1}) &= \sum_{s_{t-1}, d_{t-1}} P(S_t = s_t, D_t = d_t | S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1}) \\ &\times P(S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1} | \Omega_{t-1}) \end{aligned}$$

and,

$$\begin{aligned}
P(y_t|\Omega_{t-1}) &= \sum_{s_t, s_{t-1}, d_{t-1}} f(y_t|S_t = s_t, D_t = d_t, \Omega_{t-1}) & (3.16) \\
&\times P(S_t = s_t, D_t = d_t|S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1}) \\
&\times P(S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1}|\Omega_{t-2})
\end{aligned}$$

In constructing the likelihood function the unconditional probabilities associated with $S = 1, 2$, $1 \leq D \leq \tau$ were used to startup the filter. For more details see the appendix in Maheu and McCurdy (2000).

The filter plays an important role in forecasts of current as well as future levels of volatility. The significance of the filter and a comparison to alternative models is presented in the next section.

3.5 Features of the DDMS

Many popular volatility parameterizations, such as GARCH or stochastic volatility (SV), are continuous-state models. The DDMS has discrete states. However, conditioning on duration D_t in the conditional variance of the DDMS parameterization permits a smoother change between volatility levels than that allowed for by a simple two-state MS model.

In contrast to standard GARCH, SV, or MS models, persistence in volatility levels is time-varying in the DDMS model.¹⁰ For a simple MS model, the latent state, and therefore the volatility level, follows a linear autoregressive process with an innovation that is heteroskedastic.¹¹

To illustrate the *nonlinear* autoregressive process for the DDMS model, first note that knowledge of the state S_t , and its duration D_t , implies knowledge of the volatility level, and therefore it is sufficient to consider the AR process governing S_t . That is,

$$\begin{aligned}
S_t &= 3 - 2P_{11}(D_{t-1}) - P_{22}(D_{t-1}) \\
&+ (P_{11}(D_{t-1}) + P_{22}(D_{t-1}) - 1)S_{t-1} + \eta_t & (3.17)
\end{aligned}$$

$$D_t = \min(D_{t-1}I(S_t, S_{t-1}) + 1, \tau) \quad (3.18)$$

$$S_t = 1, 2 \quad 1 \leq D_t \leq \tau \quad (3.19)$$

Both the level and the persistence of volatility are time-varying, unlike the standard MS model. Equation (3.17) also shows that the DDMS structure is a discrete-time, discrete-state SV model.

To complete the description of the dynamics of conditional volatility, consider the properties of the innovations associated with (3.17). Since the

¹⁰Note that the GARCH volatility function (3.1) can be re-arranged as $\sigma_t^2 = \omega + (\alpha + \beta)\sigma_{t-1}^2 + \alpha(\epsilon_{t-1}^2 - \sigma_{t-1}^2)$ so that α measures the extent to which the period $t-1$ shock affects period t volatility while $\alpha + \beta$ measures the rate at which this effect dies out over time.

¹¹See Hamilton (1989) and Pagan (1996).

state variables are discrete, so are the innovations η_t which are a martingale difference sequence. That is, conditional on $S_{t-1} = 1$,

$$\eta_t = \begin{cases} P_{11}(D_{t-1}) - 1 & \text{with probability } P_{11}(D_{t-1}) \\ P_{11}(D_{t-1}) & \text{with probability } 1 - P_{11}(D_{t-1}) \end{cases} \quad (3.20)$$

and conditional on $S_{t-1} = 2$,

$$\eta_t = \begin{cases} -P_{22}(D_{t-1}) & \text{with probability } 1 - P_{22}(D_{t-1}) \\ 1 - P_{22}(D_{t-1}) & \text{with probability } P_{22}(D_{t-1}). \end{cases} \quad (3.21)$$

As in the standard MS model, these innovations are heteroskedastic.

Forecasts of future volatility make use of the filter and the time-varying transition probabilities. For example,

$$\text{Var}_t(\epsilon_{t+i}) = \sum_{s_{t+i}, d_{t+i}} \sigma_{t+i}^2(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i}) P(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i} | \Omega_t) \quad (3.22)$$

where $\sigma(S_{t+i}, D_{t+i})$ is from Equation (3.15) and $P(S_{t+i}, D_{t+i} | \Omega_t)$ is the filter based on time t information. If volatility is high (low) today, the forecast of future volatility will decrease (increase) towards the unconditional volatility level. However, the dynamics of the forecast will depend critically on the filter. To see this note that,

$$\begin{aligned} P(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i} | \Omega_t) &= \\ &\sum_{s_t, \dots, s_{t+i-1}, d_t} P(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i} | S_{t+i-1} = s_{t+i-1}, D_{t+i-1} = d_{t+i-1}) \\ &\times \dots \times P(S_{t+1} = s_{t+1}, D_{t+1} = d_{t+1} | S_t = s_t, D_t = d_t) P(S_t = s_t, D_t = d_t | \Omega_t) \end{aligned}$$

Since τ is finite, and therefore the Markov chain is ergodic, the product term on the right hand side of this equation, except for the filter, will converge to the unconditional probability of S_{t+i} and D_{t+i} , as $i \rightarrow \infty$.

Similarly, the conditional future volatility of y_{t+i} is

$$\text{Var}_t(y_{t+i}) = \sum_{k=1}^i \phi^{i-k} E_t \sigma_{t+k}^2(S_{t+k}, D_{t+k}) \quad (3.23)$$

In contrast to the standard GARCH and SV models, a unit change in z_t^2 has an effect on volatility forecasts (Equation 3.22) which is a highly nonlinear function involving the filter and the conditional density assumption. The effect is,

$$\begin{aligned} &\sum_{s_t, \dots, s_{t+i}, d_t} \sigma_{t+i}^2(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i}) \\ &\times P(S_{t+i} = s_{t+i}, D_{t+i} = d_{t+i} | S_t = s_t, D_t = d_t) \frac{\partial P(S_t = s_t, D_t = d_t | \Omega_t)}{\partial z_t^2}. \end{aligned}$$

Finally, it is useful to note that the uncertainty inherent in ϵ_t in our regime switching model comes from two sources, z_t and $\sigma(S_t, D_t)$. As a result, the DDMS decomposes uncertainty in volatility into an unpredictable component z_t , and a predictable, albeit stochastic, component $\sigma(S_t, D_t)$. The latter permits within-regime dynamics and will cause *regime uncertainty* since the state is unobserved.

Regime uncertainty means investors' future forecasts of volatility can display the *peso problem* effect. For example, forward-looking investors who believe that the market is in the low volatility state today will nevertheless attach a positive probability to the possibility of the higher volatility state occurring in the future. If a high volatility regime is not realized within the time-frame of the forecast, ex-post this forward-looking behavior can mean agents' forecasts will appear to systematically under/over estimate the true volatility.¹² Nevertheless, MS econometric models allow for the extraction of the true regime-dependent volatility levels implied in agents' forecasts.

4 Model Estimates

Full-sample estimates for our three parameterizations of volatility and each exchange rate are reported in Table 2. Standard error estimates are in parenthesis.

Overall, these results imply that duration-dependent mixing adds significantly to the in-sample fit of the volatility functions for both currencies. Although the ARCH switching model and the DDMS are not nested, the difference in the log-likelihood values suggests that the duration-dependent Markov switching model (DDMS) dominates the first-order MS-ARCH model.¹³ This inference is also supported by parameter estimates ($\gamma_2(i), i = 1, 2$), associated with duration dependence of the transition probabilities, which are statistically different from zero in both states for the GBP case and in one state for the DEM case. Furthermore, duration effects in the state-specific conditional variances are highly significant for both currencies (see estimates of parameters $\zeta(i), i = 1, 2$).

The motivation for the duration-dependent specification was to investigate whether or not a discrete-state parameterization of volatility with duration as a conditioning variable in the conditional variance function could capture volatility clustering without resorting to a time-series ARCH structure to account for remaining conditional heteroskedasticity. Consider the parameter estimates for DEM-USD. They imply that when the market is in

¹²This type of expectation mechanism is well known to play an important role in explaining asset pricing dynamics. See Evans (1996) for a recent review.

¹³We also estimated a MS-ARCH model with duration-dependent transition probabilities which nests the first-order MS-ARCH model reported in Table 2. Likelihood-ratio (LR) tests strongly reject the first-order MS-ARCH structure; p-values are .984e - 4 and .463e - 7 for the DEM and GBP cases respectively. Results are available from the authors on request.

state 1, this state becomes more persistent over time (i.e. $\gamma_2(1) > 0$) and the conditional volatility decreases as we persist in this state ($\zeta(1) < 0$). On the other hand, in state 2 conditional volatility increases with duration in that state ($\zeta(2) > 0$). Figures 2 and 3 plot the hazard function and the state-specific volatility levels for the DEM-USD. Similar duration effects are revealed for the GBP-USD case, except that conditional volatility decreases with duration in both states. However, the initial volatility level is higher for state 1 and the persistence associated with that state is stronger.

Unconditional probabilities for S_t from the DDMS model can be computed as $P(S) = \sum_d P(S, D = d)$, where $P(S, D)$, $S = 1, 2$, $1 \leq D \leq \tau$ are the joint unconditional probabilities associated with S_t and D_t . $P(S)$ for the DEM-USD are .39 ($S_t = 1$) and .61 ($S_t = 2$), and those for for the GBP-USD are .80 ($S_t = 1$) and .20 ($S_t = 2$).

Ljung-Box test statistics for autocorrelation (10 and 20 lags) in the squared standardized residuals appear at the bottom of the table. According to this diagnostic, all of the models appear to capture serial correlation in the squared standardized residuals. This suggests that using duration as an instrument in the conditional variance and transition matrix is a substitute for ARCH.

Figure 4 plots the estimates of volatility for the DDMS specification. To calculate the conditional standard deviations implied by the DDMS model, we use Equation (3.22) along with the parameter estimates reported in Table 2. Note that the DDMS model is able to capture abrupt discrete changes in volatility.

5 Adequacy of the conditional distributions

To properly manage short-term risk, the conditional distribution of returns must be correctly specified. The maintained statistical model will influence a risk assessment, such as VaR, primarily through the time-varying dynamics of conditional volatility.¹⁴ As discussed in Section 1, a risk manager will not only be concerned with whether or not VaR assessments are correct on average, but also with the adequacy of such predictions at each point in time.

This section evaluates the models' conditional distributions. We begin by assessing out-of-sample interval forecasts associated with a particular coverage level. We then use a density forecast test which has more statistical power for tail areas of the distribution since the test exploits the level of the realization as well as the indicator of whether or not it falls in the desired interval. This increased statistical power is particularly useful for relatively short data samples for which realizations in the tail may be few in number. Finally we report results for the density test using the entire distribution and the full sample of available data.

¹⁴Other conditional moments such as skewness and kurtosis may also be important for measuring risk.

5.1 Interval Forecasts

Traditionally, the most common approach to assessing a time-series volatility model has been to compare the out-of-sample forecast to some proxy of latent volatility. Volatility is sometimes measured by squared forecast errors or squared returns. For a correctly specified model this measure of volatility is consistent but noisy. The result is that conventional tests have low power to reject constant volatility models in favour of time-varying ones. As Andersen and Bollerslev (1998) show, the noise associated with using squared returns to measure latent volatility can be substantial.

An alternative approach is to consider out-of-sample interval forecasts. This is attractive because no latent measure of volatility is needed. In addition, interval forecasts depend not only on the volatility dynamics but also on the conditional mean specification and the conditional density. Thus, analyses of a model's interval forecasts provide information about the suitability of the maintained conditional distribution of returns.

We follow the testing methodology of Christoffersen (1998) and test for correct conditional coverage associated with out-of-sample interval forecasts. Tests associated with conditional coverage involve a joint test of coverage plus independence. That is, correct unconditional coverage only assesses the total number of realizations in the desired interval. It does not preclude the possibility of temporal dependence in the realization of hits in or out of the desired interval. Clustering of a particular realization would indicate neglected dependence in the conditional volatility model, or more generally, in the maintained conditional distribution. The joint test is derived in a maximum likelihood framework using the appropriate Likelihood Ratio (LR) as a test statistic.

First, define

$$(L_{t|t-1}(p), U_{t|t-1}(p)) \tag{5.1}$$

as the 1-step-ahead interval with desired coverage probability p . This 1-step-ahead interval forecast, for a particular model and information set Ω_{t-1} , is computed as

$$L_{t|t-1}(p) = E_{t-1}y_t - \Phi^{-1}\left(\frac{1-p}{2}\right)\sqrt{\text{Var}_{t-1}(y_t)} \tag{5.2}$$

$$U_{t|t-1}(p) = E_{t-1}y_t + \Phi^{-1}\left(\frac{1-p}{2}\right)\sqrt{\text{Var}_{t-1}(y_t)} \tag{5.3}$$

where $\Phi^{-1}(\frac{1-p}{2})$ is the inverse cumulative distribution function of the normal distribution evaluated at $\frac{1-p}{2}$. That is, we focus on symmetric interval forecasts.¹⁵

¹⁵Using the normal distribution for critical values would be consistent with the modeling assumption for a GARCH model but should be considered an approximation for the forecasted density associated with a MS model. More computationally oriented interval forecast methods, including the bootstrap, are detailed in Chatfield (1993).

A model with correct coverage, given information at $t-1$, has the property that the realized y_t falls in the desired interval with probability p . That is,

$$P(L_{t|t-1}(p) < y_t < U_{t|t-1}(p) | \Omega_{t-1}) = p. \quad (5.4)$$

The parameters of the models discussed in Section 3 are estimated with the information set Ω_{N-1} , $N-1 < T$, and thereafter assumed fixed, while the data from N to T is used to evaluate the out-of-sample forecasts. We set $N-1 = 1000$ leaving 304 observations for out-of-sample interval tests.¹⁶

Given a sample path $\{y_t\}_{t=N}^T$ of the time series of returns y_t , and a corresponding sequence of 1-step ahead interval forecasts, $(L_{t|t-1}(p), U_{t|t-1}(p))_{t=N}^T$, compute the binary indicator variable,

$$I_t = \begin{cases} 1 & \text{if } y_t \in (L_{t|t-1}(p), U_{t|t-1}(p)) \\ 0 & \text{otherwise} \end{cases} \quad (5.5)$$

for $t = N, \dots, T$. This indicator variable records, for each t , whether or not the forecast interval (5.1) contained the realized value y_t .

Now consider a test for correct conditional coverage, which is defined as,

$$E[I_t | I_{t-1}, I_{t-2}, \dots] = p \quad (5.6)$$

for all t . Christoffersen (1998) shows that correct conditional coverage implies that $\{I_t\} \sim IID \text{ Bernoulli}(p)$ under the null hypothesis. This test can be decomposed into two individual tests. The first is a test for correct unconditional coverage $E[I_t] = p$, versus the alternative $E[I_t] \neq p$, while the second test is for independence of the binary sequence $\{I_N, \dots, I_T\}$.

The likelihood under the null hypothesis of correct conditional coverage is

$$L(p; I_N, \dots, I_T) = (1-p)^{n_0} p^{n_1} \quad (5.7)$$

in which n_0 and n_1 are the number of zeros and ones respectively, from the data $\{I_N, \dots, I_T\}$. Following Christoffersen (1998), we consider a first-order Markov chain alternative for which the likelihood is

$$L(\Pi; I_N, \dots, I_T) = \pi_{00}^{n_{00}} (1 - \pi_{00})^{n_{01}} \pi_{11}^{n_{11}} (1 - \pi_{11})^{n_{10}}, \quad (5.8)$$

where n_{ij} denotes the number of observations where the value i is followed by j from the data $\{I_N, \dots, I_T\}$, and the π_{ij} are the Markov transition probabilities associated with the transition matrix Π .

The joint test (coverage and independence) can be assessed using the LR statistic,

$$LR_{cc} = -2 \log \left[L(p; I_N, \dots, I_T) / L(\hat{\Pi}; I_N, \dots, I_T) \right] \quad (5.9)$$

¹⁶The out-of-sample tests reported in this section were computed prior to the full-sample estimation results in Table 2.

which will be distributed $\chi^2(2)$ under the null hypothesis. Note that the restrictions implied by that null hypothesis of correct conditional coverage are $\pi_{11} = p$ and $\pi_{00} = (1 - p)$.

Table 3 reports, for each model and currency, p-values associated with the joint test for correct conditional coverage (labelled IF) for a desired symmetric coverage level of 0.8. That is, we are evaluating an interval forecast which covers the middle 80% of the distribution. Recall that this test associated with conditional coverage is a joint test of correct unconditional coverage and independence. Although this test rejects a linear AR(1) model with constant variance (p-values are .02 and .00 for the DEM-USD and GBP-USD, respectively), it is unable to reject either of the time-varying volatility parameterizations for the DEM case. On the other hand, all models are rejected for the GBP-USD application. Our results indicate that the rejection of correct conditional coverage for this currency is due to failure to capture average (unconditional) coverage rather than due to violation of independence.

For a well-specified model we would expect the interval forecast tests to pass for a wide range of desired coverage levels. In other words, we want forecasts of the model to be accurate for different regions of the distribution. Although conceptually we could compute the p-values associated with a wide range of desired coverage levels, statistical power will depend on how many realizations fall outside and inside the chosen interval. As the interval approaches the whole distribution, the outside region will disappear so that the test cannot be implemented. On the other hand, choosing a desired coverage corresponding to, for example, the lower tail of the distribution, may result in a relatively small number of realizations inside the interval. In either case, the test will lack statistical power to discriminate between the null and the alternative and consequently between the models. For these cases, an alternative strategy to evaluating the conditional density is required.

5.2 Density Forecasts

A closely related family of tests on a model's distributional assumptions is detailed in Diebold, Gunther, and Tay (1998) and Berkowitz (1999). The tests are based on the integral transformation of Rosenblatt (1952). Suppose that $f(y_t|\Omega_{t-1})$ is the conditional distribution of y_t based on the information set Ω_{t-1} . Then Rosenblatt (1952) shows that

$$u_t = \int_{-\infty}^{y_t} f(v|\Omega_{t-1})dv, \quad t = N, \dots, T, \quad (5.10)$$

is iid and uniformly distributed on $(0, 1)$. Therefore, a researcher can construct $\{\tilde{u}_t\}_{t=N}^T$ based on a candidate model $\tilde{f}(\cdot)$, and perform tests to see whether or not $\tilde{u}_t \sim UID(0, 1)$.

Diebold, Gunther, and Tay (1998) suggest graphical methods in order to assess how $\tilde{f}(\cdot)$ fails in approximating the true unknown density, while

Berkowitz (1999) suggests applying the inverse normal transformation to $\{\tilde{u}_t\}_{t=N}^T$ in order to test whether or not the transformed series $\{z_t\}_{t=N}^T$ is independent standard normal. An attractive feature of the latter approach is that we can use any of the exact tests based on normality. In particular, following Berkowitz (1999), we consider likelihood ratio tests for several applications. By construction, these tests involve a continuous variable and not a discrete indicator variable as in the interval forecast tests (Section 5.1). For this reason, and also due to the use of exact Likelihood Ratio statistics, we expect these tests to have good power properties in finite samples.¹⁷

First, define $z_t = \Phi^{-1}(\tilde{u}_t)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function. Then under the null hypothesis of the correct density, $z_t \sim NID(0, 1)$. The first LR test is based on the following regression,

$$z_t = \mu + \rho(z_{t-1} - \mu) + w_t \quad (5.11)$$

The null hypothesis is $\mu = \rho = 0, Var(w_t) = 1$ whereas the alternative is $\mu \neq 0, \rho \neq 0, Var(w_t) \neq 1$.¹⁸ This test may identify problems in the unconditional distribution and also time dependences not captured by the conditional dynamics.

The first panel of Table 3 reports p-values for this LR test (labelled DT-full) associated with density forecasts using the same out-of-sample period $t = N, \dots, T$ used for the interval forecasts reported in Section 5.1. The relevant integrals were computed numerically utilizing quadrature. Broadly speaking, this test produces conclusions similar to the interval forecast tests with respect to the alternative currencies. However, the p-values are lower so that there are marginal rejections for the DEM-USD as well. The DDMS model is not as strongly rejected as the MS-ARCH, however both models appear to be missing important features in the data. Note however that we did not update parameter estimates after each one-step ahead forecast. Doing so might result in a more favourable result for the postulated distributions.

Finally, we also compute a LR test that focuses on lower-tail behaviour (as in VaR) which can be constructed based on a truncated normal distribution. The loglikelihood for the truncated normal $z_t < \alpha$ is

$$l = \sum_{z_t < \alpha} \left[-.5 \log(2\pi\sigma^2) - .5 \frac{(z_t - \mu - \rho(z_{t-1} - \mu))^2}{\sigma^2} - \Phi\left(\frac{\alpha - \mu - \rho(z_{t-1} - \mu)}{\sigma}\right) \right].$$

Again the null hypothesis is $\mu = \rho = 0, Var(w_t) = 1$ and the alternative is $\mu \neq 0, \rho \neq 0, Var(w_t) \neq 1$.

¹⁷Traditional tests on $\{\tilde{u}_t\}_{t=N}^T$ for IID behaviour generally will not have good power. However, Berkowitz (1999) shows that applying the inverse normal transformation to \tilde{u}_t and using classical tests results in good power properties in finite samples.

¹⁸It should be noted that these tests do not take into account sampling variability in estimating the parameters of $\tilde{f}(\cdot)$.

Since the number of realizations in the tail will be small, we computed this density test using the entire sample ($t = 1, \dots, T$).¹⁹ The in-sample results for the lower tail, ($\alpha = -1.645$), are reported in Table 3 in row DT-tail. Again, these results favour the DDMS parameterization. As a further comparison the AR(1)-GARCH(1,1) model had a p-value of .006 (DEM-USD) and .004 (GBP-USD) for the DT-tail test. These results for the left tail of the distribution concord well with our measures of the unconditional distribution discussed in the next section which show that the duration-dependent parameterization captures excess kurtosis more adequately.

Our results suggest that volatility is forecastable at a weekly frequency and that a constant volatility model for exchange rates can be a poor choice. Moreover, unlike the tests conducted in West and Cho (1995), these interval and density forecast tests show the value of richer models in tracking conditional volatility. We have found that the DDMS parameterization outperforms the GARCH and MS-ARCH model, but is still strongly rejected by the UK weekly data.

6 Features of the Unconditional Distributions

Earlier sections showed that the addition of high-order dependence through duration dependence in the DDMS model allows for rich conditional density dynamics for foreign currency rates. This section shows that the DDMS parameterization also provides a more flexible structure when matching unconditional moments.

Monte Carlo methods are used to compute some summary measures of the unconditional distributions implied by the alternative volatility parameterizations. To estimate the simulated moments, we draw a sample of size 1000000 from the conditional distribution implied by the maintained model, and calculate a battery of summary statistics.²⁰ These simulated moments can then be compared to the corresponding moments (and standard errors) estimated from the actual data. Tables 4 and 5 report these numbers for the DEM-USD and GBP-USD cases, respectively.

The simulated statistics for the MS models match those observed in the data more closely than those from the plain vanilla GARCH model and, in some cases, those for the DDMS model match better than the MS-ARCH model. Consider, for example, the results for the DEM-USD case. The standard deviation from the DDMS and MS-ARCH models is much closer to that found in the data than is the GARCH estimate. In addition, the GARCH model produces too much kurtosis²¹, the MS-ARCH produces too

¹⁹Although these distribution tests were motivated as out-of-sample tests there is no reason they cannot be used as an in-sample diagnostic test.

²⁰The first 20000 draws were dropped to eliminate dependence on startup conditions.

²¹Note that for this case the GARCH parameter estimates are close to the boundary for the existence of the fourth moment. Further simulations show that, in this case, there is

little kurtosis, while the estimate for the DDMS parameterization is within one standard error of that for the data. For the GBP-USD case, the standard deviation of y_t is matched by all models but, in this case, both the GARCH and MS-ARCH models produce too little kurtosis while the DDMS parameterization is within two standard errors of the kurtosis estimate for the data.

The DDMS model also fits better than a simple first-order MS model. For example, the unconditional standard deviation and kurtosis measures²² associated with a first-order MS model are 1.46 and 3.54 for the DEM-USD. As shown in Table 4, the kurtosis associated with the data is 4.822 (standard error .639) and that implied by the DDMS model is 4.340. In the GBP-USD case, the first-order MS model yields an unconditional kurtosis measure of 4.43 while the data has 6.579 (standard error .843) and DDMS has 5.163. These results, as well as other simulation results for the simpler MS models, point to the duration dependence structure in the DDMS model providing a closer fit to the unconditional distribution of the weekly changes in foreign exchange rates.

Many of the summary measures for $|y_t|$ and $\log |y_t|$, which we can interpret as alternative proxies for volatility, are also captured better by the DDMS parameterization. For example, for the DDMS parameterization applied to the DEM-USD, all of the first four moments of volatility as measured by $|y_t|$ are within one standard error of the data estimate whereas none of them are for either the MS-ARCH or GARCH models. For volatility clustering (measured by the autocorrelation in $|y_t|$), there is not much to choose between the alternative parameterizations. However, the MS-ARCH model does slightly better for the DEM-USD case whereas DDMS and GARCH are marginally better for the GBP-USD currency.

To complement the results reported in Tables 4 and 5, Figures 5 and 6 present kernel density plots of the data and of the simulated unconditional distributions for y_t and $\log |y_t|$ implied by the DDMS and GARCH models. 1000000 draws were taken from the respective models. These data were used to estimate the density assuming a Gaussian kernel and a constant bandwidth that is optimal for the normal distribution with the variance calculated from the data. The plots were robust to a wide range of bandwidth parameters.

The figures support the conclusion that the DDMS parameterization provides a good description of the unconditional distributions. For instance, Figure 5 shows that DDMS captures the density of y for the DEM-USD quite well around the origin, while GARCH does not. This is consistent with the GARCH model putting too much mass in the tail and resulting in a higher kurtosis than in the data (Table 4). Similarly, for the GBP-USD (Figure 6), the distribution of $\log |y_t|$ is closer to the data for the DDMS parameteriza-

an upward bias on the kurtosis estimate even for large sample sizes.

²²These were computed using parameter estimates (available on request) for a first-order MS model and formulae in Timmermann (2000).

tion, although neither model captures this log absolute value transformation very well. This evidence is consistent with that presented in earlier sections of the paper, that is, although the DDMS model is preferred, none of the parameterizations fully capture the structure in the GBP-USD case.

7 Concluding Comments

The last section showed that the DDMS can account for many of the properties of the unconditional distribution of y_t and functions of y_t usually associated with measures of volatility. Unlike GARCH parameterizations, the DDMS structure allows time-varying persistence, includes a stochastic component for volatility, and incorporates anticipated discrete changes in the level of volatility. The duration-dependent Markov switching (DDMS) model is an example of a mixture of distributions model. As reviewed in Section 1, MS models are well known to have the ability to produce various shaped distributions including skewness and leptokurtosis. Our plots of unconditional distributions for DDMS confirm those results.

According to our results, including the out-of-sample interval and density forecast tests reported in Section 5, the DDMS parameterization is also a good statistical characterization of the conditional distribution of foreign exchange returns. This is in contrast to earlier studies (for example, Pagan and Schwert (1990)) that have shown that simpler MS models cannot capture all of the volatility dependence.

There are several important differences between the DDMS and those simpler first-order MS models of heteroskedasticity. Firstly, incorporating the discrete-valued random variable summarizing state-dependent duration allows a parsimonious parameterization of potential high-order dependence. We have shown that the duration-dependent MS model (DDMS) is better suited to capturing volatility dependence as compared to the MS-ARCH model which appends an ARCH structure to a first-order MS model. This is evident from our in-sample and out-of-sample analyses. Also, DDMS may be able to capture long memory. Secondly, although the DDMS is only a two-state model, it can capture a broad range of volatility levels through conditioning on duration in the conditional variance. Finally, by including duration in the transition matrix, persistence in a volatility state is permitted to be time varying. These features render the DDMS parameterization particularly suitable for capturing stochastic volatility dependence.

The empirical distribution generated by our proposed structure is a superior match for the samples of data used in this paper. These enhancements may be particularly relevant for forecasts necessary for risk management. However, it is still difficult to fully capture the distributions of log-differences of the GBP-USD exchange rate (see Gallant, Hsieh, and Tauchen (1991)). More work remains to be done.

Table 1: Summary Statistics for Weekly Exchange Rates

	Germany	U.K.
mean	-.036 (.040)	.024 (.039)
stdev	1.462 (.040)	1.423 (.046)
skewness	-.104 (.179)	.267 (.245)
excess kurtosis	1.822 (.639)	3.579 (.843)
modified $Q(1)$	1.157 [.282]	1.472 [.225]
modified $Q(2)$	4.378 [.112]	1.476 [.478]
modified $Q(10)$	11.320 [.333]	11.565 [.315]
$Q^2(10)$	83.349 [.000]	156.416 [.000]

The data are percentage returns from weekly exchange rates with the US dollar from 1974/01/02-1998/12/23. Standard errors robust to heteroskedasticity are in parenthesis, p-values in square brackets. $Q(j)$ is the Ljung-Box statistic for serial correlation in the demeaned series and $Q^2(j)$ is the same in the squared series with j lags. The modified $Q(j)$ allows for conditional heteroskedasticity and follows West and Cho (1995).

Model Summary

GARCH(1,1)

$$y_t = \mu + \phi y_{t-1} + \epsilon_t, \quad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1).$$

MS-ARCH(p)

$$\begin{aligned} y_t &= \mu + \phi y_{t-1} + \epsilon_t \\ \epsilon_t &= \sigma(S_t) z_t, \quad z_t \sim N(0, 1), \quad S_t = 1, 2. \\ \sigma_t^2(S_t) &= \omega(S_t) + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 \\ P(S_t = 1 | S_{t-1} = 1) &= \frac{\exp(\gamma_1(1))}{1 + \exp(\gamma_1(1))}, \\ P(S_t = 2 | S_{t-1} = 2) &= \frac{\exp(\gamma_1(2))}{1 + \exp(\gamma_1(2))}. \end{aligned}$$

DDMS

$$\begin{aligned} y_t &= \mu + \phi y_{t-1} + \epsilon_t, \\ \epsilon_t &= \sigma(S_t) z_t, \quad z_t \sim N(0, 1), \quad S_t = 1, 2. \\ \sigma(S_t) &= (\omega(S_t) + \zeta(S_t) D_t)^2 \\ P(S_t = 1 | S_{t-1} = 1, D_{t-1} = d_{t-1}) &= \frac{\exp(\gamma_1(1) + \gamma_2(1) d_{t-1})}{1 + \exp(\gamma_1(1) + \gamma_2(1) d_{t-1})}, \\ P(S_t = 2 | S_{t-1} = 2, D_{t-1} = d_{t-1}) &= \frac{\exp(\gamma_1(2) + \gamma_2(2) d_{t-1})}{1 + \exp(\gamma_1(2) + \gamma_2(2) d_{t-1})}. \end{aligned}$$

Table 2: Model Estimates

Parameter	Germany		UK	
	MS-ARCH	DDMS	MS-ARCH	DDMS
μ	-.058 (.036)	-0.054 (.037)	-.004 (.012)	-.004 (.011)
ϕ	.055 (.028)	.061 (.029)	.066 (.029)	.033 (.028)
$\omega(1)$	1.705 (.156)	1.133 (.116)	1.642 (.155)	1.573 (.043)
$\zeta(1)$		-.012 (.004)		-.023 (.002)
$\omega(2)$.367 (.069)	1.088 (.106)	.002 (.001)	1.114 (.067)
$\zeta(2)$.020 (.004)		-.036 (.003)
α_1	.039 (.030)		.112 (.032)	
α_2	.096 (.038)		.101 (.038)	
α_3	.152 (.041)		.206 (.041)	
$\gamma_1(1)$	5.662 (.867)	.728 (.626)	1.598 (.609)	.958 (.470)
$\gamma_2(1)$.118 (.047)		.135 (.031)
$\gamma_1(2)$	3.819 (.726)	2.181 (.945)	.036 (.319)	.430 (.395)
$\gamma_2(2)$		-.004 (.046)		.107 (.046)
lgl	-2267.598	-2262.845	-2191.407	-2148.736
$Q^2(10)$	3.292 [.973]	16.363 [.090]	1.806 [.997]	.016 [.999]
$Q^2(20)$	21.209 [.385]	27.886 [.112]	2.177 [.999]	.033 [.999]

$Q^2(10)$ and $Q^2(20)$ are Ljung-Box statistics on the squared standardized residuals.

Table 3: Distribution Test Results
Out-of-Sample Forecasts

Test	Germany		UK	
	MS-ARCH	DDMS	MS-ARCH	DDMS
IF	.209	.137	.0005	.0011
DT-full	.012	.030	.1e-5	.0065

In-Sample Forecasts

Test	Germany		UK	
	MS-ARCH	DDMS	MS-ARCH	DDMS
DT-tail	.075	.348	.011	.029

The IF (interval forecast) test is a joint LR test for correct coverage and independence based on Christoffersen (1998) with a desired coverage level of .8. The DT (density tests) are LR tests (Berkowitz (1999)) for the adequacy of the maintained model's distribution – DT-full for the entire density and DT-tail for the lower 5% tail of the distribution. p-values are reported for the null hypotheses: for IF that the time-series of interval forecasts for the middle 80% of the distribution have correct conditional coverage; and for DT that the maintained forecast density is the true density.

Table 4: Selected Statistics for Models and Data, Germany

Statistic	DATA	MS-ARCH	DDMS	GARCH(1,1)
Stdev y_t	1.462 (.040)	1.465	1.469	1.619
Skewness y_t	-.104 (.179)	.200e-2	.179e-2	.978e-2
Kurtosis y_t	4.822 (.639)	3.546	4.340	6.091
corr($ y_t , y_{t-1} $)	.107 (.027)	.108	.140	.223
corr($ y_t , y_{t-2} $)	.132 (.036)	.142	.129	.222
corr($ y_t , y_{t-3} $)	.168 (.030)	.182	.121	.221
Mean $ y_t $	1.103 (.027)	1.139	1.116	1.220
Stdev $ y_t $.961 (.036)	.924	.958	1.065
Skewness $ y_t $	1.722 (.233)	1.249	1.584	2.106
Kurtosis $ y_t $	8.401 (1.756)	4.904	6.715	13.297
Mean log $ y_t $	-.379 (.033)	-.314	-.353	-.256
Stdev log $ y_t $	1.174 (.031)	1.144	1.157	1.147
Skewness log $ y_t $	-1.157 (.089)	-1.414	-1.346	-1.363
Kurtosis log $ y_t $	4.738 (.382)	6.510	6.308	6.369

DATA is 100 times the log first-difference in the exchange rate. For each model one draw of sample size 1000000 was used to calculate the sample statistic. Standard errors robust to heteroskedasticity appear in parenthesis. The parameter values in Table 2 are assumed to be the true model parameters.

Table 5: Selected Statistics for Models and Data, UK

Statistic	DATA	MS-ARCH	DDMS	GARCH(1,1)
Stdev y_t	1.423 (.046)	1.458	1.432	1.411
Skewness y_t	.267 (.245)	.572e-2	.012	.011
Kurtosis y_t	6.579 (.843)	4.244	5.163	3.691
$\text{corr}(y_t , y_{t-1})$.156 (.045)	.188	.159	.131
$\text{corr}(y_t , y_{t-2})$.117 (.037)	.165	.132	.127
$\text{corr}(y_t , y_{t-3})$.168 (.036)	.225	.121	.124
Mean $ y_t $	1.024 (.027)	1.094	1.049	1.102
Stdev $ y_t $.989 (.044)	.964	.975	.882
Skewness $ y_t $	2.258 (.234)	1.463	1.844	1.339
Kurtosis $ y_t $	11.376 (1.940)	6.189	7.815	5.736
Mean $\log y_t $	-.500 (.034)	-.427	-.476	-.330
Stdev $\log y_t $	1.216 (.034)	1.241	1.245	1.124
Skewness $\log y_t $	-1.131 (.103)	-1.294	-1.347	-1.453
Kurtosis $\log y_t $	5.009 (.454)	5.749	6.158	6.709

See notes to Table 4

Figure 1: Time Series of Exchange Rates

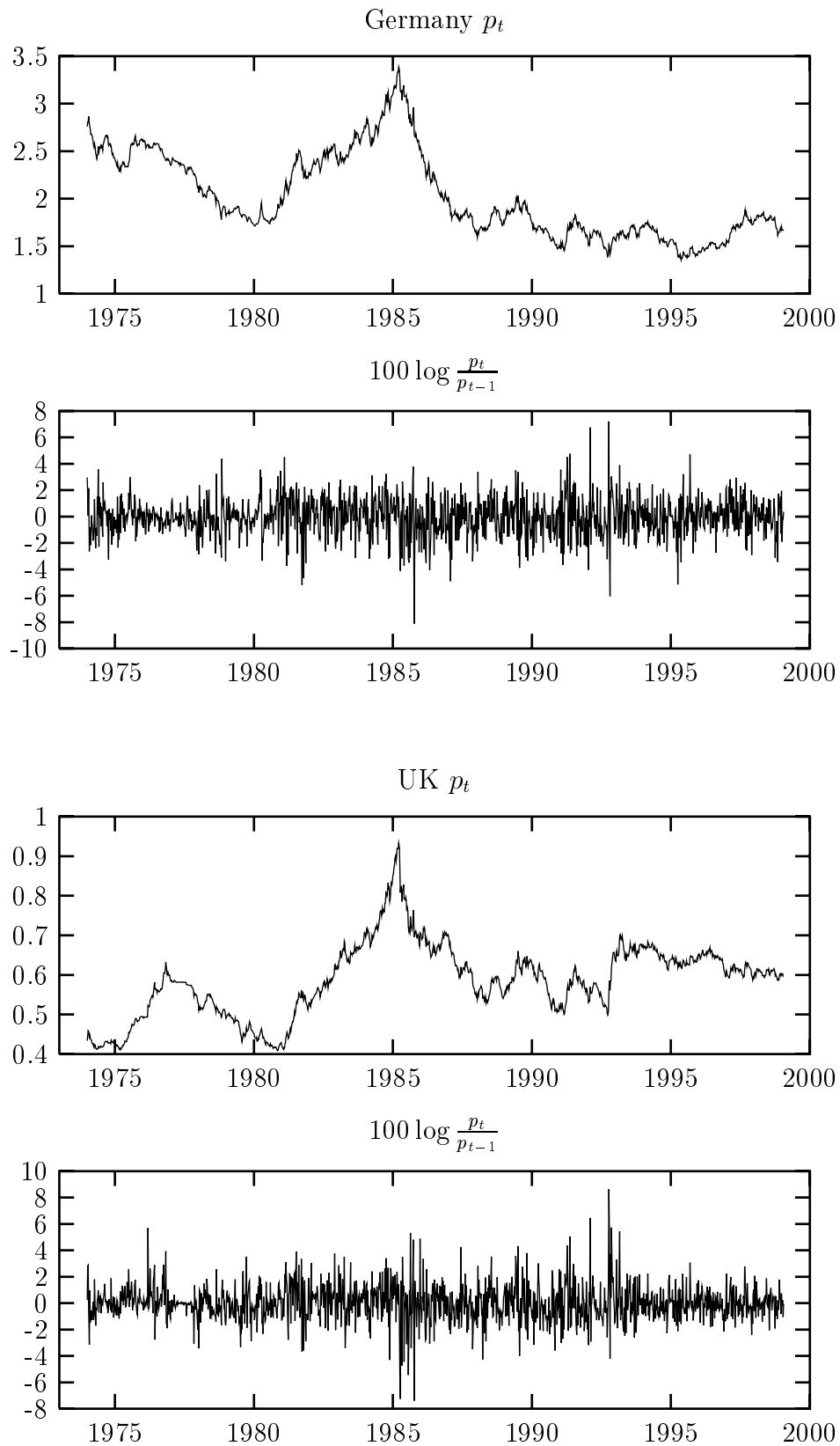


Figure 2: Hazard Function, DDMS, DEM-USD

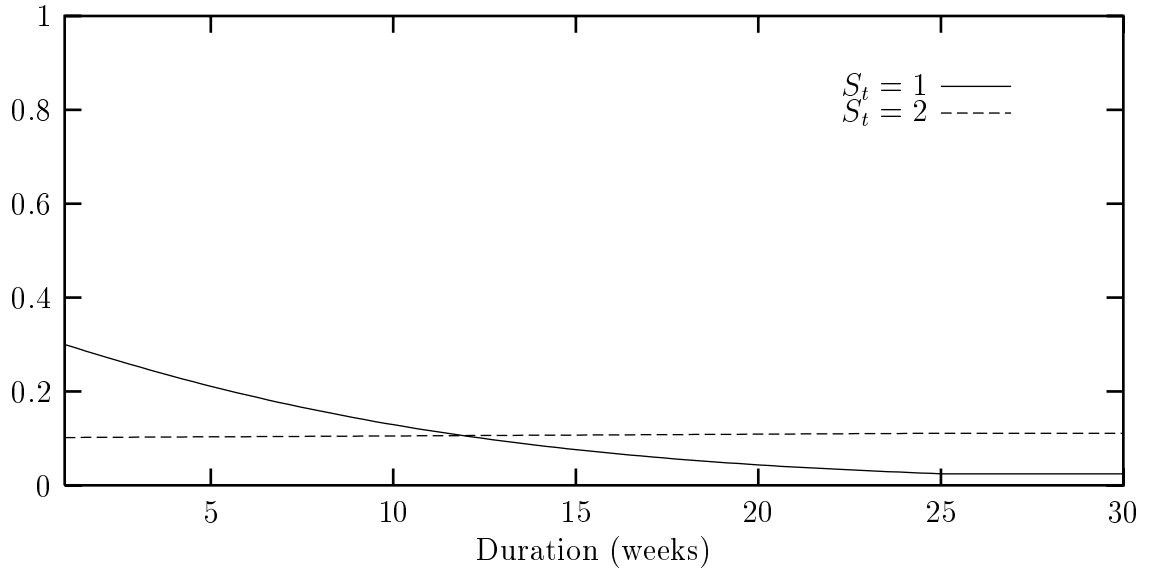


Figure 3: Variance Function, DDMS, DEM-USD

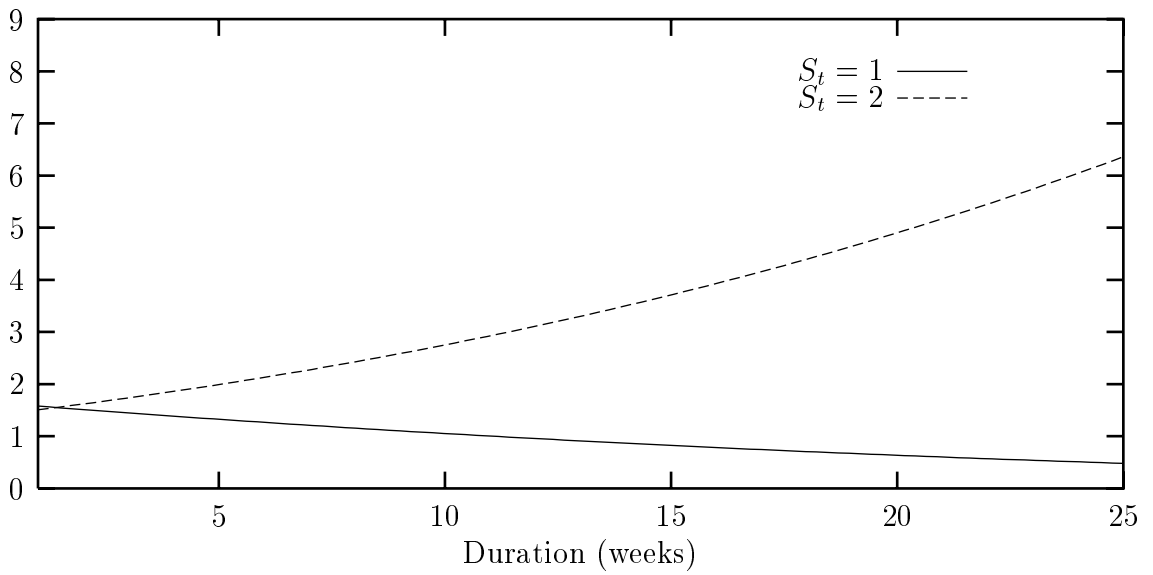


Figure 4: Estimated Volatility from the DDMS Model

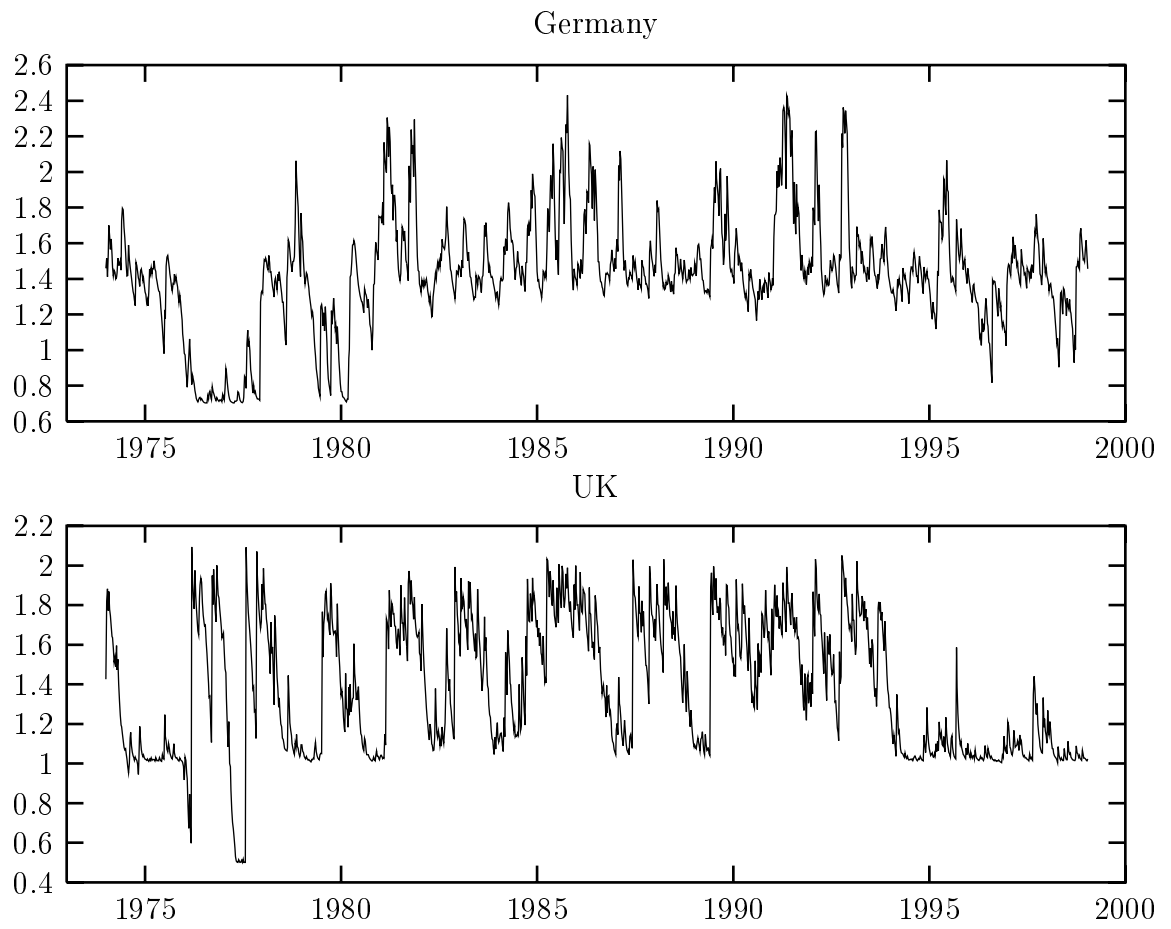


Figure 5: y_t , Models versus Data, Germany

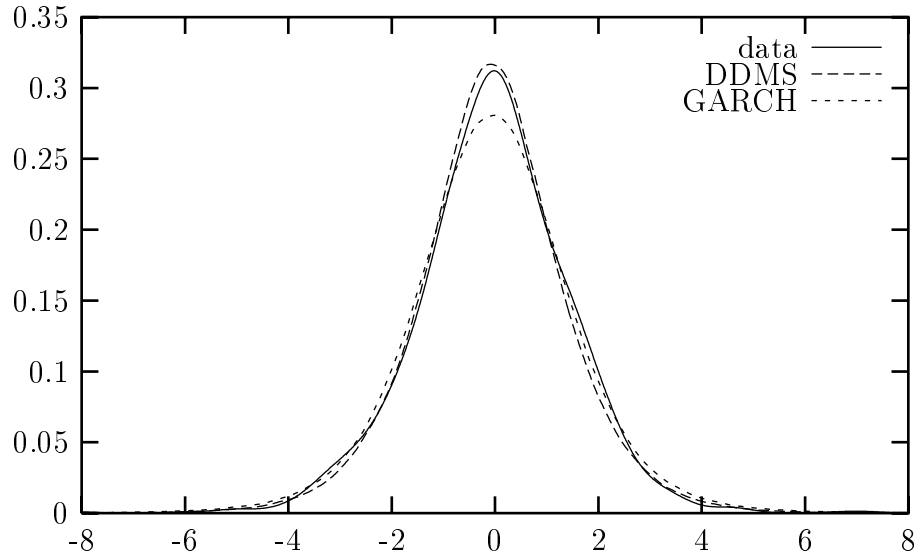
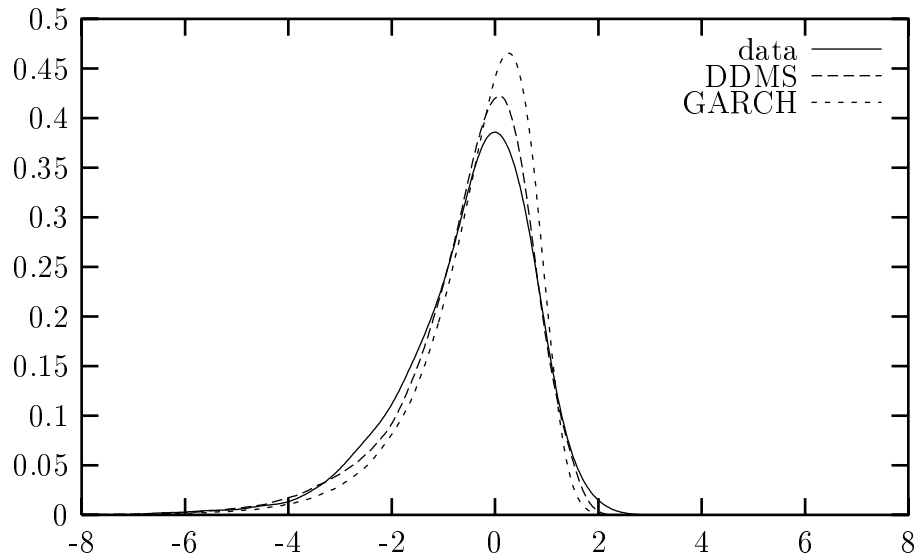


Figure 6: $\log |y_t|$, Models versus Data, UK



References

- ANDERSEN, T. G., AND T. BOLLERSLEV (1998): “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” *International Economic Review*, 39(4), 885–905.
- BERKOWITZ, J. (1999): “Evaluating the Forecasts of Risk Models,” Federal Reserve Board.
- BOLLERSLEV, T. (1986): “Generalized Autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 309–328.
- BOOTHE, P., AND D. GLASSMAN (1987): “The Statistical Distribution of Exchange Rates,” *Journal of International Economics*, 22, 297–319.
- CAI, J. (1994): “A Markov Model of Switching-Regime ARCH,” *Journal of Business & Economic Statistics*, 12(3), 309–316.
- CHATFIELD, C. (1993): “Calculating Interval Forecasts,” *Journal of Business & Economic Statistics*, 11, 121–135.
- CHRISTOFFERSEN, P. F. (1998): “Evaluating Interval Forecasts,” *International Economic Review*, 39(4), 841–862.
- DIEBOLD, F. X. (1988): *Empirical Modeling of Exchange Rate Dynamics*. Springer, Berlin.
- DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (1998): “Evaluating Density Forecasts with Application to Financial Risk Management,” *International Economic Review*, 39(4), 863–883.
- DURLAND, J. M., AND T. H. MCCURDY (1994): “Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth,” *Journal of Business & Economic Statistics*, 12(3), 279–288.
- ENGLE, 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 Heteroskedasticity with Estimates of the UK inflation,” *Econometrica*, 50, 987–1008.
- EVANS, M. D. D. (1996): *Peso Problems: Their Theoretical and Empirical Implications* chap. 21, Handbook of Statistics vol. 14. Elsevier Science.
- FILARDO, A. J. (1994): “Business Cycle Phases and Their Transitional Dynamics,” *Journal of Business & Economic Statistics*, 12(3), 299–308.

- GALLANT, A. R., D. A. HSIEH, AND G. E. TAUCHEN (1991): *Nonparametric and Semiparametric Methods in Econometrics and Statistics and Econometrics*chap. On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974-1983, pp. 199–240. Cambridge University Press.
- GHYSELS, E., A. C. HARVEY, AND E. RENAULT (1996): *Stochastic Volatility* chap. 5, Handbook of Statistics vol. 14. Elsevier Science.
- GRAY, S. F. (1996): “Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process,” *Journal of Financial Economics*, 42, 27–62.
- 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 Non-stationary Time Series and the Business Cycle,” *Econometrica*, 57, 357–384.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.
- HAMILTON, J. D., AND R. SUSMEL (1994): “Autoregressive Conditional Heteroskedasticity and Changes in Regime,” *Journal of Econometrics*, 64, 307–333.
- HSIEH, D. (1989): “Testing for Nonlinearity in Daily Foreign Exchange Rate Changes,” *Journal of Business*, 62, 339–368.
- KAEHLER, J., AND V. MARNET (1993): “Markov-Switching Models for Exchange-Rate Dynamics and the Pricing of Foreign-Currency Options,” ZEW Discussion paper No. 93-03.
- KIM, C. J., AND C. R. NELSON (1998): “Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching,” *Review of Economics and Statistics*, 80(2), 188–201.
- KIM, C. J., C. R. NELSON, AND R. STARTZ (1998): “Testing for Mean Reversion in Heteroskedastic Data Based on Gibbs-Sampling-Augmented Randomization,” *Journal of Empirical Finance*, 5(2), 131–154.
- KLAASSEN, F. (1998): “Improving GARCH Volatility Forecasts,” Tilburg University.
- LAM, P. (1997): “A Markov Switching Model of GNP Growth with Duration Dependence,” Department of Economics, Ohio State University.

- LINDGREN, G. (1978): “Markov Regime Models for Mixed Distributions and Switching Regressions,” *Scandinavian Journal of Statistics*, 5, 81–91.
- MAHEU, J. M., AND T. H. MCCURDY (2000): “Identifying Bull and Bear Markets in Stock Returns,” *Journal of Business & Economic Statistics*, 18(1), 100–112.
- PAGAN, A. (1996): “The Econometrics of Financial Markets,” *Journal of Empirical Finance*, 3, 15–102.
- PAGAN, A., AND G. W. SCHWERT (1990): “Alternative Models for Conditional Stock Volatility,” *Journal of Econometrics*, 45(1-2), 267–290.
- PALM, F. C. (1996): *GARCH Models of Volatility* chap. 7, Handbook of Statistics vol. 14. Elsevier Science.
- PEREZ-QUIROS, G., AND A. TIMMERMANN (1999): “Firm Size and the Cyclical Variations in Stock Returns,” Forthcoming in *Journal of Finance*.
- ROSENBLATT, M. (1952): “Remarks on a Multivariate Transformation,” *Annals of Mathematical Statistics*, 23, 470–472.
- RYDÉN, T., T. TERÄSVIRTA, AND S. ÅSBRINK (1998): “Stylized Facts of Daily Returns Series and the Hidden Markov Model,” *Journal of Applied Econometrics*, 13, 217–244.
- TAYLOR, S. J. (1999): “Markov Processes and the Distribution of Volatility: A Comparison of Discrete and Continuous Specifications,” Lancaster University, Management School, working paper 99/001.
- TIMMERMANN, A. (2000): “Moments of Markov Switching Models,” *Journal of Econometrics*, 96, 75–111.
- WEST, K. D., AND D. CHO (1995): “The Predictive Ability of Several Models of Exchange Rate Volatility,” *Journal of Econometrics*, 69, 367–391.