Regime-Switching and the Estimation of Multifractal Processes*

Laurent Calvet
Harvard University and NYU Stern School of Business

Adlai Fisher
Faculty of Commerce, University of British Columbia

This Version: December 27, 2002
First Draft: March 2, 2002

*Department of Economics, Littauer Center, Cambridge, MA 02138, and Department of Finance, NYU Stern School of Business, 44 West Fourth Street, New York, NY 10012, lcalvet@aya.yale.edu; Finance Division, Faculty of Commerce, 2053 Main Hall, Vancouver, BC, Canada V6T 1Z2, adlai.fisher@commerce.ubc.ca. We are grateful to John Campbell, Guido Kuersteiner and Nour Meddahi for stimulating discussions. We also received helpful comments from T. Andersen, T. Bollerslev, F. Diebold, R. Engle, J. Hamilton, J. MacKinnon, N. Shephard, D. Smith, R. Tsay and seminar participants at the University of Chicago, the 2002 NSF-NBER Time Series Conference, and the 2002 CIRANO Extreme Events in Finance Conference. Excellent research assistance was provided by Xifeng Diao. We are very appreciative of financial support provided for this project by the Social Sciences and Humanities Research Council of Canada under grant 410-2002-0641.

Abstract

We propose a discrete-time stochastic volatility model in which regime-switching serves three purposes. First, changes in regimes capture low frequency variations, which is their traditional role. Second, they specify intermediate frequency dynamics that are usually assigned to smooth autoregressive processes. Finally, high frequency switches generate substantial outliers. Thus, a single mechanism captures three important features of the data that are typically addressed as distinct phenomena in the literature. Maximum likelihood estimation is developed and shown to perform well in finite sample. We estimate on exchange rate data a version of the process with four parameters and more than a thousand states. The estimated model compares favorably to earlier specifications both in- and out-of-sample. Multifractal forecasts slightly improve on GARCH(1,1) at daily and weekly intervals, and provide considerable gains in accuracy at horizons of 10 to 50 days.

JEL Classification: G0, C5.

Keywords: Forecasting, long memory, Markov regime-switching, maximum likelihood estimation, scaling, stochastic volatility, time deformation, volatility component, Vuong test.

1. Introduction

Regime-switching models (Hamilton 1989, 1990) have been extremely useful in capturing the apparent nonstationarity of many macroeconomic and financial series. Despite the theoretical generality of the state space approach, empirical applications most often limit its use to low frequency shifts and combine it with alternative techniques to capture the rich dynamics of economic series.[1] For example in the volatility literature, the class of Markov-switching ARCH and GARCH processes separately specifies three types of fluctuations: regime-switches at very low frequencies, smooth ARCH or GARCH transitions at mid-range frequencies, and a thick-tailed conditional distribution of returns at high-frequency (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; Klaassen, 2002).[2] In contrast, this paper shows that a pure regime-switching process can effectively capture the dynamics at all frequencies.

Previous literature rarely investigates an application with a large number of regimes. This is partly because a small number of states may seem sufficient to capture low frequency regime switches. For a general formulation, a more practical limitation is that the number of parameters grows quadratically in the number of states. A natural solution consists of placing restrictions on the state parameters and switching probabilities. For example, Bollen, Gray, and Whaley (2000) use a set of parameter restrictions to more tightly specify a four regime model.[3] In this paper, we consider a pure regime-switching model with a dense transition matrix by imposing a tight set of restrictions on regime parameters and switching probabilities to parsimoniously specify volatility dynamics across a wide range of frequencies. Our approach permits us to routinely estimate good-performing models with over a thousand states and only four parameters using

[1] One exception is Dueker (1997), who investigates regime switches in the tail-thickness of the conditional density of stock returns and finds a highly leptokurtotic regime with duration of about one week.

[2] In addition to the Markov-switching GARCH processes considered in this paper, Hamilton's model has been extended by letting transition probabilities vary with past returns (Diebold, Lee, and Weinbach, 1994), state occupation times (Durland and McCurdy, 1994), or exogenous variates (Filardo, 1994; Perez-Quiros and Timmerman, 2000). See Hamilton and Raj (2002) for a recent survey of this literature.

[3] Duration-dependent Markov-switching models also use restrictions on state parameters and switching probabilities. For example, Maheu and McCurdy (2000) expand a two-state model by supposing that the volatility level and switching probabilities depend on the occupation time. This results in a sparse transition matrix, in which the system either progresses in the same state or switches to the other state.

standard maximum likelihood methods.

This paper builds on Calvet and Fisher (2001), which develops a time-stationary formulation of multifractal diffusions as well as a weakly convergent sequence of discretizations. Total volatility is specified as the multiplicative product of a large but finite number of random components. It is assumed for simplicity that these components are first order Markov and identical except for time scale. Specifically, the only difference between components is that each has a different switching probability. The specification is completed by assuming that the progression of switching probabilities is approximately geometric. This parsimonious specification delivers long-memory features in volatility and a decomposition into components with heterogenous decay rates.

The present paper develops a toolkit of econometric methods that can be used to estimate and test multifractal processes. Calvet, Fisher, and Mandelbrot (1997) introduce the multifractal model of asset returns (MMAR), a class of diffusions that capture the outliers, moment-scaling and long memory in volatility exhibited by many financial time series. While providing an excellent fit of the data, the MMAR is based on a combinatorial construction that is not well-suited to standard econometric techniques. In particular, regime changes take place at predetermined dates making the model non-stationary. Early efforts to estimate the MMAR have for this reason focused on unconditional moments of returns. More specifically, the multifractal model implies that the moments of returns vary as power functions of the frequency of observation, a property exhibited by a large number of financial series.[4] Calvet and Fisher (2002) use this property to build estimator and diagnostic tests of the MMAR. Lux (2001) similarly develop a GMM estimator based on analytical expressions of high-frequency covariances.

This paper considers instead the pure Markov-switching version of the MMAR. This permits maximum likelihood (ML) estimation, an innovation which, to the best of our knowledge, is new to the literature on multifractal measures and processes. Our model and estimation techniques are easy to use and extend directly from the original Hamilton regime-switching formulation. The main innovation consists of introducing a tight and effective set of restrictions on regimes and transition probabilities. Monte Carlo simulations demonstrate the effectiveness of

[4] Financial economists are already familiar with a scaling property in the second moment through the variance ratio statistic. Calvet, Fisher, and Mandelbrot (1997) and Calvet and Fisher (2002) find evidence of moment-scaling in powers of the absolute value of returns. Further evidence is provided by Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2003). See also LeBaron (2001) for a discussion of robustness.

this technique in typical sample sizes. We apply this method to several exchange rates and show that the multifractal model performs well in comparison with the Student-GARCH(1,1) of Bollerslev (1987), and the Markov-switching GARCH (MS-GARCH) of Klaassen (2002). The choice of these alternative processes is guided by several considerations. Since the focus of the paper is ML estimation, it is natural to choose models for which the likelihood function is readily available. Furthermore, GARCH(1,1) is known to be an excellent contender for volatility forecasting among the large class of ARCH / stochastic volatility processes (e.g. Akgiray, 1989; Andersen and Bollerslev, 1998; Hansen and Lunde, 2001; Pagan and Schwert, 1990; West and Cho, 1998). MS-GARCH processes, on the other hand, combine regime-switching with the smooth weighting of past shocks. They represent a potentially appealing compromise between GARCH and pure regime switching models. Because the original process of Gray (1996) does not conveniently permit multistep forecasting, we instead consider the variant formulation introduced by Klaassen (2002).

The multifractal process compares favorably with GARCH(1,1) both in- and out-of-sample. Our model has a higher likelihood in-sample for all currencies. The statistical significance of this difference is confirmed by a HAC adjusted version of the Vuong (1989) test. Furthermore since GARCH(1,1) and the regime-switching model have the same number of parameters, the multifractal is preferred by standard selection criteria. Analogous results are obtained out-of-sample. While the models' one-day forecasts have similar performances, the multifractal model substantially outperforms GARCH(1,1) at longer horizons. The difference is most pronounced for 20 and 50 days, for which the multifractal substantial outperforms GARCH(1,1) for all currencies. For example in the case of the British Pound, the 50-day forecasting R^2 is 27.3% for our model as compared to -2.6% for GARCH(1,1). Similar improvements are obtained for other currencies. The empirical evidence thus suggests that the multifractal model dominates GARCH(1,1) both in- and out-of-sample.

MS-GARCH provides substantially better fit than GARCH(1,1) in-sample. This property is partly attributable to a larger number of parameters and suggests the possibility of overfitting. Out-of-sample, MS-GARCH roughly matches the performance of GARCH(1,1) at short horizons, which is consistent with the findings of Klaassen (2002). At longer horizons, however, MS-GARCH is dominated by GARCH(1,1), and the multifractal therefore appears as the best performing model.

A pure Markov switching model can thus capture the same dynamics that in previous literature have required not only regime-switching but also linear GARCH transitions and a thick-tailed conditional distribution of returns. It is striking that a single mechanism can play all three of these roles so effectively. The innovation that achieves this surprising economy of modeling technique is based upon scale-invariance, an approach that has long been championed by Mandelbrot (1963, 1997). In our regime-switching formulation, scale invariance provides a specific set of restrictions on the regime parameters and transition probabilities of a high dimensional state space.

Section 2 presents the discrete-time model. Section 3 develops the ML estimator and assesses its accuracy in Monte Carlo simulations. Section 4 provides ML empirical results for four exchange rates. Section 5 compares our model with alternative processes both in- and out-of-sample. Section 6 concludes. All proofs are given in the Appendix.

2. The Multifrequency Regime-Switching Model

This section develops a discrete-time Markov process with multi-frequency stochastic volatility. The process has a finite number $\overline{k}$ of latent volatility state variables, each of which corresponds to a different frequency.

2.1. Stochastic Volatility

We consider an economic series X_t defined in discrete time on the regular grid $t = 0, 1, 2, ..., \infty$. In applications, X_t will be the log-price of a financial asset or exchange rate. Define the innovations $x_t \equiv X_t - X_{t-1}$. A common modeling methodology assumes that every period, the system is hit by a single shock that progressively phases out over time (e.g. Engle, 1982). We consider instead an economy with $\overline{k}$ components $M_{1,t}, M_{2,t}, ..., M_{\overline{k},t}$, which decay at heterogeneous frequencies $\gamma_1, .., \gamma_{\overline{k}}$. Such a model could be very unwieldy as the number $\overline{k}$ becomes very large. We will see, however, that the process can be nonetheless parsimoniously described by a small set of parameters.

We model the innovations $x_t \equiv X_t - X_{t-1}$ by

$$x_t = \sigma(M_{1,t}M_{2,t}...M_{\overline{k},t})^{1/2}\varepsilon_t. \tag{2.1}$$

The parameter σ is a positive constant. The random variables ε_t are IID standard Gaussians $\mathcal{N}(0, 1)$. The random *multipliers* or *volatility components* $M_{k,t}$ are persistent, non-negative and satisfy $\mathbb{E}(M_{k,t}) = 1$. We consider for simplicity that the multipliers $M_{1,t}, M_{2,t}...M_{\bar{k},t}$ at a given time t are statistically independent. The parameter σ is then equal to the unconditional standard deviation of the innovation x_t.

Equation (2.1) defines a stochastic volatility model $x_t = \sigma_t \varepsilon_t$ with the multiplicative structure $\sigma_t = \sigma(M_{1,t} M_{2,t}...M_{\bar{k},t})^{1/2}$. We conveniently stack the period t volatility components multipliers into a vector

$$M_t = \left(M_{1,t}, M_{2,t}, ..., M_{\bar{k},t}\right).$$

For any $m = (m_1, .., m_{\bar{k}}) \in \mathbb{R}^{\bar{k}}$, let $g(m)$ denote the product $\prod_{i=1}^{\bar{k}} m_i$. Volatility at time t is then $\sigma_t = \sigma[g(M_t)]^{1/2}$.

The properties of volatility are driven by the stochastic dynamics of the vector M_t. We assume for simplicity that M_t is first-order Markov. This facilitates the construction through time of $\{x_t\}$ and permits maximum likelihood estimation. It is then natural to call M_t the *volatility state vector*, and each component $M_{k,t}$ a *state variable*. The econometrician observes the returns $x_t = \sigma[g(M_t)]^{1/2} \varepsilon_t$, but not the vector M_t itself. The vector M_t is therefore latent, and must be inferred recursively by Bayesian updating.

Each $M_{k,t}$ follows a process that is identical except for time scale. Assume that volatility state variables have been constructed up to date $t-1$. For each $k \in \{1, .., \bar{k}\}$, the next period multiplier $M_{k,t}$ is drawn from a fixed distribution M with probability γ_k, and is otherwise equal to its current value: $M_{k,t} = M_{k,t-1}$. The dynamics of $M_{k,t}$ can be summarized as:

$$\begin{array}{ll} M_{k,t} \text{ drawn from distribution } M & \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \text{with probability } 1-\gamma_k. \end{array}$$

We assume for simplicity that switches and new draws from M are IID across k and t. The volatility components $M_{k,t}$ thus differ in their transition probabilities γ_k but not in their marginal distribution M. These features greatly contribute to the parsimony of the model.

The transition probabilities $\gamma \equiv (\gamma_1, \gamma_2, ..., \gamma_{\bar{k}})$ are specified as

$$\gamma_k = 1 - (1-\gamma_1)^{\left(b^{k-1}\right)}, \tag{2.2}$$

where $\gamma_1 \in (0,1)$ and $b \in (1, \infty)$. This specification is introduced in Calvet and Fisher (2001) in connection with the discretization of a Poisson arrival process. Since $1-\gamma_k = (1-\gamma_1)^{\left(b^{k-1}\right)}$, the logarithms of staying probabilities are exponentially decreasing with k. Consider a process with very persistent components, or equivalently a very small parameter γ_1. For small values of k, the quantity $\gamma_1 b^{k-1}$ remains small and the transition probability satisfies:

$$\gamma_k \sim \gamma_1 b^{k-1}.$$

The transition probabilities of low frequency components thus grow approximately at rate b. At higher frequencies ($\gamma_k \sim 1$), the rate of increase slows down and condition (2.2) guarantees that the parameter γ_k remains lower than 1. In empirical applications, it is numerically convenient to estimate parameters with the same order of magnitude. Since $\gamma_1 < ... < \gamma_{\bar{k}} < 1 < b$, we therefore use $(\gamma_{\bar{k}}, b)$ to specify the set of transition probabilities.

The parameter $\overline{k}$ determines the number of volatility frequencies in the model. We will take this number as fixed, and view the choice of $\overline{k}$ as a model selection problem. This approach is consistent with the methodology employed for ARMA(p, q) or GARCH(p, q), where estimation is developed for a fixed number of lags and the determination of p and q is treated as a model selection problem. The multifractal construction imposes only minimal restrictions on the marginal distribution of the multipliers: $M \geq 0$ and $\mathbb{E}(M) = 1$. While the model allows flexible parametric or even nonparametric specifications of M, this paper only considers the most parsimonious setup. We assume that M is drawn from a family of unit mean distributions that are specified by a single parameter m_0. Useful examples include log-normal distributions: $\ln M \sim \mathcal{N}(m_0, -m_0^2/2)$, or binomial random variables taking values m_0 or $2 - m_0$ with equal probability. The full parameter vector is then

$$\psi \equiv (m_0, \sigma, b, \gamma_{\bar{k}}) \in \mathbb{R}^4,$$

where m_0 characterizes the distribution of the multipliers, σ is the unconditional standard deviation of returns, and b and $\gamma_{\bar{k}}$ define the set of switching probabilities. In Section 3, we will develop and empirically implement the maximum likelihood estimation of this vector.

We find it convenient to call this construct the Markov-Switching Multifractal (or alternatively Markov-Switching Multifrequency) process. The notation MSM($\overline{k}$) will refer to versions of the model with $\overline{k}$ frequencies. Economic intuition and earlier work suggest that the multiplicative structure (2.1) is appealing to model the high variability and high volatility persistence exhibited by financial time series. When a low level multiplier changes, volatility varies very discontinuously and has strong degree of persistence. In addition, high frequency multipliers introduce substantial outliers.

2.2. Properties

The MSM construction permits low frequency regime shifts, and thus long volatility cycles in sample paths. We will see that in exchange rate series, the duration of the most persistent component, $1/\gamma_1$, is typically of the same order as the length of the data. Estimated processes thus tend to generate volatility cycles with periods proportional to the sample size, a property also apparent in the sample paths of long memory processes. Long memory is often defined as the hyperbolic decline in the autocovariance function as the lag goes to infinity. Fractionally integrated processes generate such patterns by assuming that an innovation linearly affects future periods at a hyperbolically declining weight. As a result, fractionally integrated processes tend to be very smooth, and thus do not capture the highly irregular volatility patterns exhibited by financial series. In contrast, our approach generates both substantial variability and long cycles in the size of returns. It thus represents an appealing modeling technique for financial econometrics. Furthermore as $\bar{k} \to \infty$, the growing number of regime switches in the construction implies that the limiting continuous time process lies outside the class of Itô diffusions. The sample paths are continuous but exhibit a high degree of heterogeneity in local behavior, which is characterized by a continuum of local Hölder exponents in any finite time interval (Calvet and Fisher, 2002).

In earlier work, we proposed a definition of long memory that applies to continuous time processes defined on a bounded time domain. This definition is based on increments over progressively smaller intervals. Our discrete-time process can also generate a hyperbolic decline in autocovariances for a range of lags. For every integer n and every $q \geq 0$, the autocovariance in levels is defined as $\gamma_q(n) = Cov(|x_t|^q, |x_{t+n}|^q)$. We show in the Appendix:

Proposition. *For every n, $1 << n << b^{\bar{k}}$, the autocovariance in levels satisfies $\gamma_q(n) \propto n^{2d(q)-1}$, where $d(q) = \log_b\{\mathbb{E}(M^q)/[\mathbb{E}(M^{q/2})]^2\}$.*

Our discrete-time construction thus mimics the hyperbolic autocovariograms exhibited by many financial series (e.g. Dacorogna *et al.*, 1993; Ding and Granger, 1993). The connection between low frequency regime shifts and long memory has been recently discussed in the literature (e.g. Diebold and Inoue, 2001).

A representative return series is illustrated in Figure 1. The graph reveals that multiple switching frequencies help generate large heterogeneity in volatility levels and substantial outliers. This is notable since the return process has by construction finite moments of every order. We will see that it is nonetheless sufficient to capture the tail properties of the exchange rate data. In Calvet and Fisher (2001), we introduced Paretian tails by considering IID shocks of the form $\varepsilon_t = Z_t \Omega_t$, where $\{Z_t\}$ are IID standard Gaussians and $\{\Omega_t\}$ are IID random variables drawn from a Paretian distribution Ω. The random variable Ω is the limit distribution of the squared variation of the price process over a fixed horizon, and is thus fully specified by the vector $(\sigma, m_0, b, \gamma^*)$. We find it computationally more convenient to consider the Gaussian case because the likelihood is then available in closed-form. The discrete-time process with $\varepsilon_t = Z_t \Omega_t$ is a promising direction for future research.

3. Maximum Likelihood Estimation

When the multiplier M has a discrete distribution, there exist a finite number of volatility states. Standard filtering methods then provide the likelihood function in closed-form.

3.1. Updating the State Vector

We assume in the rest of this paper that the multiplier M takes a finite number of values b_m. The vector $M_t = \left(M_{1,t}, M_{2,t}, ..., M_{\overline{k},t}\right)$ can therefore take $d = b_m^{\overline{k}}$ values $m^1, ..., m^d \in \mathbb{R}^{\overline{k}}$. The dynamics of the Markov chain M_t are characterized by the transition matrix $A = (a_{i,j})_{1 \le i,j \le d}$ with components $a_{ij} = \mathbb{P}(M_{t+1} = m^j | M_t = m^i)$.[5]

Let $n(.)$ denote the density of a standard normal. The density of the innovation x_t conditional on M_t is $f_{x_t}(x | M_t = m^i) = [\sigma g(m^i)]^{-1} n[x/\sigma g(m^i)]$. Consider the conditional probabilities

$$\Pi_t^j \equiv \mathbb{P}\left(M_t = m^j | x_1, .., x_t\right)$$

over the unobserved states $m^1, ..., m^d$. We can stack these conditional probabilities in the row vector $\Pi_t = \left(\Pi_t^1, .., \Pi_t^d\right) \in \mathbb{R}_+^d$. Let $\iota = (1, ., 1) \in \mathbb{R}^d$. We know that $\Pi_t \iota' = 1$.

[5] We note that $a_{ij} = \prod_{k=1}^{\overline{k}} \left[(1-\gamma_k) 1_{\{m_k^i = m_k^j\}} + \gamma_k \mathbb{P}(M = m_k^j)\right]$, where m_k^i denotes the mth component of vector m^i, and $1_{\{m_k^i = m_k^j\}}$ is the dummy variable equal to 1 if $m_k^i = m_k^j$, and 0 otherwise. In Calvet and Fisher (2001) the transition matrix is different because there it is assumed that an innovation to a lower frequency multiplier causes innovations in all higher frequency multipliers. Here, we assume that arrival times are independent across frequencies.

Using Bayes' rule, we update the new belief Π_{t+1} from the old belief Π_t and the new innovation x_{t+1}. Consider the function

$$\omega(x_t) = \left[\frac{n\left(x_t/\sigma g\left(m^1\right)\right)}{\sigma g\left(m^1\right)}, ..., \frac{n\left(x_t/\sigma g\left(m^d\right)\right)}{\sigma g\left(m^d\right)} \right].$$

and let $x * y$ denote the Hadamard product $(x_1 y_1, .., x_d y_d)$ of any $x, y \in \mathbb{R}^d$. We show in the Appendix that

$$\Pi_{t+1} = \frac{\omega(x_{t+1}) * (\Pi_t A)}{\left[\omega(x_{t+1}) * (\Pi_t A)\right] \iota'}. \tag{3.1}$$

This formula expresses the conditional probability Π_{t+1} as a function of the observation x_{t+1} and the probability Π_t calculated in period t. These results imply that we can recursively calculate Π_t. Since the multipliers $(M_{1,1}, .., M_{\bar{k},1})$ are independent, the initial vector Π_0 is uniquely determined by $\Pi_0^j = \prod_{l=1}^{\bar{k}} \mathbb{P}(M = m_l^j)$ for all j.

3.2. The Likelihood Function

Having solved the conditioning problem, we show in the Appendix that the log likelihood function is

$$\ln L\left(x_1, ..., x_T; \psi\right) = \sum_{t=1}^{T} \ln[\omega(x_t) \cdot (\Pi_{t-1} A)].$$

For a fixed $\bar{k}$, we know that the maximum likelihood estimator (ML) is consistent and asymptotically efficient as $T \rightarrow \infty$. An important difference between our model and standard Markov switching models stems from the parsimonious parameterization of the transition matrix A. This allows us to estimate the model with reasonable precision even in the presence of a very large state space. While the Expectation Maximization (EM) algorithm proposed in Hamilton (1990) is not directly applicable with constrained transition probabilities, we find that numerical optimization of the likelihood function produces good results.

3.3. Small Sample Properties

This section assesses the small sample properties of the ML estimator in Monte Carlo simulations. The models investigated have $\bar{k} = 8$ frequencies, which is large enough to represent specifications that perform well in later empirical sections. We also restrict attention to the simple binomial specification where the multiplier M takes values m_0 and $2 - m_0$ with equal probability.

The four required parameters are thus the binomial value m_0, the unconditional standard deviation σ, the frequency growth rate b, and the high frequency switching probability $\gamma_{\bar{k}}$. The choice of the unconditional standard deviation is not consequential since it is simply a normalization of size, and we therefore choose $\sigma = 1$. The remaining parameters are set to values that are representative of empirical results in later sections. Specifically, all simulations have $b = 3$ and $\gamma_{\bar{k}} = 0.95$, and the binomial parameter takes one of three values $m_0 \in \{1.3, 1.4, 1.5\}$. We also consider three sample lengths $T \in \{T_1, T_2, T_3\}$, where $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$.

Table 1 reports the simulation results. Each of the nine columns is based on a different combination of one of the three values of m_0 with one of the three values of the sample length T. For each column, we simulate 400 independent sample paths of the corresponding model and sample length. Maximum likelihood estimation then provides a set of parameter estimates and asymptotic standard errors for each path.[6] The table has four rows corresponding to each parameter. The first row gives the average point estimate over the simulated paths. The second row is the standard error of these point estimates, or the finite sample standard error (FSSE). The third row gives the root mean squared error (RMSE) of the parameter estimates relative to the true parameter values.[7] Finally, the average asymptotic standard error (AASE) gives the average over the 400 simulations of the asymptotic standard errors calculated from the information matrix. As sample size becomes large, we expect that the AASE and the FSSE become close.

The results of Table 1 show that maximum likelihood estimation works well. For m_0, σ, and b, the biases are small and become negligible as sample size increases. The parameter m_0 has a low standard error relative to its size and is thus well identified, which is important because this parameter largely determines the variability of volatility. By contrast, the unconditional standard deviation σ has standard errors that, although declining as expected with sample length, are

[6] We start the optimizations at the true parameter values and iterate to convergence once. Preliminary work considered searching for multiple local optima, and although these occasionally exist, they do not significantly affect the means or standard deviations of the reported Monte Carlo results.

[7] The RMSE will always be greater than or equal to the FSSE, and equal to the FSSE only when the average point estimate is identical to the true parameter value.

roughly ten percent of the true parameter value. We interpret this result as a strength of the model, as it is consistent with the idea that low-frequency variations create considerable uncertainty about long-run averages. The parameter b shows a moderate degree of uncertainty about the amount of spacing between frequencies. Finally, the high-frequency switching probability $\gamma_{\bar{k}}$ is the only parameter that shows more than a small bias. This disappears quickly as sample size increases, and the standard errors are not generally large. Overall the Monte Carlo simulations show that maximum likelihood estimation produces reliable results given the sample sizes considered in subsequent sections. We also note that the convenience and efficiency of maximum likelihood offers significant advantages relative to previous moment-based estimators for multifractal processes.

4. Empirical Results

Using a binomial specification for the multiplier M, we apply ML estimation to four exchange rate series and obtain preferred specifications with a large number of volatility frequencies.

4.1. Exchange Rate Data

The empirical analysis uses daily exchange rate data for the Deutsche Mark (DEM), Japanese Yen (JPY), British Pound (GBP), and Canadian Dollar (CAD), all against the US Dollar. The data consists of daily prices reported at noon by the Federal Reserve Bank of New York.[8] The fixed exchange rate system broke down in early 1973, and the DEM, JPY and GBP series accordingly begin on 1 June 1973. The CAD series starts a year later (1 June 1974) because the Canadian currency was held essentially at parity with the US Dollar for several months after the demise of Bretton Woods. The Deutsche Mark was replaced by the Euro at the beginning of 1999. The DEM data accordingly ends on 31 December 1998, while the other three series run until 30 June 2002. Overall, the series contains 6,420 observations for the Deutsche Mark, 7,049 observations for the Canadian Dollar, and 7,299 observations for the Yen and the Pound.

Figure 2 plots the daily returns of each series. Consistent with earlier studies, we observe apparent volatility clustering on a range of frequencies. For each series, Table 2 reports the square root of average squared returns computed over

[8] More specifically, the data consist of buying rates for wire transfers at 12:00 PM Eastern time.

the entire sample and over four subsamples of equal length. We observe that the sample standard deviation can vary quite substantially across subperiods, which is consistent with the low-frequency regime shifts in our model.

4.2. ML Estimation Results

Table 3 reports MLE results for all four currencies. The columns of the table correspond to the number of frequencies $\bar{k}$ varying from 1 to 10. The first column is thus a standard Markov-switching model with only two possible values for volatility. As $\bar{k}$ increases, the number of states increases at the rate $2^{\bar{k}}$. There are thus over one thousand states when $\bar{k} = 10$.

We first examine the DEM data, and note that the multiplier value $\hat{m}$ tends to decline with $\bar{k}$. This is because with a larger number of volatility components, less variability is required in each individual component to generate the same aggregate amount of stochastic volatility. The estimates of $\hat{\sigma}$ show variability of a different type across $\bar{k}$, with no particular pattern of increasing or decreasing. This is again consistent with the idea that the long-run average of volatility is difficult to identify in a model that permits long volatility cycles. We finally examine the frequency parameters $\hat{\gamma}_{\bar{k}}$ and $\hat{b}$. When $\bar{k} = 1$, the parameter $\hat{\gamma}_{\bar{k}}$ indicates a switch in the single multiplier a little less than once every two weeks. As $\bar{k}$ increases, the switching probability of the highest frequency multiplier increases until for large values of $\bar{k}$ a switch occurs almost every day. At the same time, the estimated value $\hat{b}$ decreases steadily with $\bar{k}$. We can use (2.2) to calculate that when $\bar{k} = 10$, a switch in the lowest frequency multiplier occurs approximately once every ten years, or about one third the sample size. Thus, as $\bar{k}$ increases, the range of frequencies spreads out while the spacing between frequencies becomes tighter.

The parameter estimates for the other three currencies show many similarities. In all cases, the value of $\hat{m}$ tends to decrease with $\bar{k}$. We also observe that the values of $\hat{m}$, and thus the importance of stochastic volatility, are largest for JPY and GBP and smallest for CAD. Consistent with this, variability across $\bar{k}$ in the estimates of $\hat{\sigma}$ is also greatest for JPY and GBP and least for CAD. As $\bar{k}$ increases all currencies also show the highest frequency multiplier switching more often and the spacing between frequencies becoming tighter. The spacing between frequencies is widest for JPY and GBP and tightest for CAD. We correspondingly observe that the duration of the lowest frequency multiplier is longest for JPY at approximately three times the sample size, and smallest for CAD at approximately one tenth the sample size. We also note that all asymptotic standard errors are roughly consistent with the magnitudes obtained in our Monte Carlo simulations, although in the empirical results somewhat smaller for $\hat{\sigma}$ and slightly larger for the frequency location parameters $\hat{\gamma}_{\bar{k}}$ and $\hat{b}$.

The estimated MSM$(\bar{k})$ processes generate substantial outliers for large values of $\bar{k}$ despite the fact that they have finite moments of every order. To confirm this, for each data set we use the estimated process with $\bar{k} = 10$ frequencies to generate ten thousand paths of the same length as the data. We then compute a Hill tail index α for each simulated path. Basing the index on 100 order statistics, the empirical tail index and the average α in the simulated samples are respectively equal to 4.74 and 4.34 (DEM), 3.91 and 3.75 (JPY), 4.59 and 4.03 (GBP), and 4.40 and 4.79 (CAD). Furthermore, for all currencies we cannot at the 10% level reject equality of the simulated and empirical tail statistics. This result is caused by the high frequency variations in volatility in the MSM model. With the highest frequency multipliers taking new values almost daily, the distribution of returns is strongly affected by being composed of a mixture of distributions. Even though this mixture of distributions has finite moments of every order, it is more than sufficient to capture the tail characteristics of the data, even in sample sizes as large as almost thirty years of daily observations.

We finally examine the behavior of the log-likelihood function. For each currency, as we increase the number of frequencies $\bar{k}$ the likelihood increases substantially when $\bar{k}$ is low. As $\bar{k}$ gets larger, the likelihood continues to increase in most cases, but at a decreasing rate. The only exception to the monotonic increase in likelihood is the DEM series, for which the likelihood reaches a peak at $\bar{k} = 7$. In all other cases the likelihood reaches a maximum at $\bar{k} = 10$. This behavior of the likelihood confirms one of the main premises of the multifractal approach. Specifically, volatility fluctuations have a multiplicity of frequencies, and explicitly incorporating a larger number of frequencies results in a better fit. The next section examines the statistical significance of these increases in the likelihood function.

4.3. Model Selection

This section investigates whether the estimated MSM($\bar{k}$) models have significant differences in likelihoods. We compare two processes MSM($\bar{k}$) and MSM($\bar{k}'$), $\bar{k} \neq \bar{k}'$, with respective likelihoods f and g. The processes are non-nested and have log-likelihood difference:

$$\sqrt{T}(\ln L_T^f - \ln L_T^g) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \ln \frac{f(x_t \,|x_1, ..., x_{t-1})}{g(x_t \,|x_1, ..., x_{t-1})}.$$

Consider the null hypothesis that the models have identical unconditional expected log-likelihoods. When the observations $\{x_t\}$ are IID, Vuong (1989) shows that the difference $\ln L_T^f - \ln L_T^g$ is asymptotically normal under the null.[9] In addition, the variance of this difference is consistently estimated by the sample variance of the addends $\ln[f(x_t \,|x_1, ..., x_{t-1})/g(x_t \,|x_1, ..., x_{t-1})]$. Because it is a strong assumption that the conditional likelihoods are IID, we construct in the Appendix a HAC-adjusted version of the Vuong test. Our discussion is a simplified version of the broader approach recently proposed by Rivers and Vuong (2002).

For each $\bar{k} \in \{1, .., 9\}$, we test in Table 4 the null hypothesis that MSM($\bar{k}$) and MSM(10) fit the data equally well. Since HAC-adjusted tests tend to perform poorly in small samples,[10] we compute t-ratios and one-sided p values using both the original and the HAC-adjusted methods. For $\bar{k} \in \{1, 2, 3\}$, the log-likelihood difference is significant at the 1% level in the non-adjusted case (Table 4A) and at the 5% level in the HAC case (Table 4B). This is strong evidence that MSM(10) significantly outperforms models with 1 to 3 frequencies. For $\bar{k} \in \{4, 5\}$, we reject the null at the 5% (non-adjusted) and 20% (HAC-adjusted) levels in almost all cases. We view these results as rather substantial evidence that MSM(10) outperforms models with 4 or 5 frequencies. Lower significance levels are obtained for larger values of $\bar{k}$, and the overall conclusion is that the MSM model works better for larger numbers $\bar{k}$ of frequencies. For this reason and also to maintain consistency in the remaining analysis, we henceforth focus on the MSM($\bar{k} = 10$) process for all currencies.

5. Comparison with Alternative Models

This Section compares the multifractal model with GARCH(1,1) and Markov-switching GARCH (MS-GARCH) both in- and out-of-sample.

[9] See the Appendix for a more detailed discussion of Vuong tests.

[10] See for example Andrews (1991), Andrews and Monahan (1992), and den Haan and Levin (1997).

5.1. In-Sample Comparison

We consider alternative processes of the form $x_t = h_t^{1/2} e_t$, where h_t is the conditional variance of x_t at date $t-1$, and $\{e_t\}$ are IID Student innovations with unit variance and ν degrees of freedom (d.f.'s). GARCH(1,1) assumes the recursion $h_{t+1} = \omega + \alpha\varepsilon_t^2 + \beta h_t$. MS-GARCH combines short-run autoregressive dynamics with low frequency regime shifts (Gray, 1996; Klaassen, 2002). A latent state $s_t \in \{1, 2\}$ follows a first-order Markov process with transition probabilities $p_{ij} = \mathbb{P}(s_{t+1} = j | s_t = i)$. In every period, the econometrician observes the return x_t but not the latent s_t. For $i = \{1, 2\}$, let $h_{t+1}(i) = Var_t(x_{t+1}|s_{t+1} = i)$ denote the variance of x_{t+1} conditional on past returns $\{x_t\}_{s=1}^t$ and $s_{t+1} = i$. The quantity h_t is latent in every period, and the econometrician can similarly define $\mathbb{E}_t\left[\, h_t(s_t) \,|s_{t+1} = i\right]$, the expectation of h_t conditional on $s_{t+1} = i$ and past returns. Klaassen (2002) assumes the conditional dynamics:

$$h_{t+1}(i) = \omega_i + \alpha_i\varepsilon_t^2 + \beta_i\mathbb{E}_t\left[\, h_t(s_t) \,|s_{t+1} = i\right]. \tag{5.1}$$

The equation conditions volatility on a larger information set than the Gray specification: $h_{t+1}(i) = \omega_i + \alpha_i\varepsilon_t^2 + \beta_i\mathbb{E}_{t-1}\, h_t(s_t)$. We prefer (5.1) for two reasons. First, Klaassen shows that his model has better forecasting performance on three of the exchange rates considered in this paper (DEM, JPY and GBP), and attributes these improvements to finer conditioning. Second, the Klaassen version permits multi-step forecasting, which will be a key part of the out-of-sample analysis.

We report in Table 5 the ML estimates of the alternative processes. The coefficient $1/\nu$ is the inverse of the degrees of freedom in the Student distribution. This is a convenient renormalization that has been frequently used in the literature (e.g. Bollerslev, 1987). Each coefficient σ_i, $i = 1, 2$, represents the standard deviation of returns conditional on the volatility state: $\sigma_i^2 = \omega_i/(1 - \alpha_i - \beta_i)$. These coefficients are easier to interpret across models than the intercepts ω_i. As shown in Table 6, the multifractal has a higher likelihood than GARCH(1,1) for all exchange rates in spite of the fact that both processes have the same number of parameters. In fact, the GARCH(1,1) likelihoods approximately match what is obtained by MSM with only about 3 or 4 frequencies. The multifractal model thus gives an improved fit over GARCH(1,1) in-sample.

The MS-GARCH model raises different issues because it uses more parameters (9) than either the GARCH or MSM model (4). Thus, although MS-GARCH has higher likelihoods than either GARCH or MSM, we obtain a different ordering when using the Schwarz BIC criterion. In particular, the multifractal model is then indistinguishable from MS-GARCH in the GBP data, and is preferred for DEM and CAD.

As suggested by Vuong (1989), we can also test the statistical significance of differences in the BIC criterion. The last two columns of Table 6 test the GARCH and MS-GARCH models against MSM under this metric.[11] We again give p-values for both the original version of the test as well as a HAC adjusted variant. Under the original version, the in-sample performance of the MSM model over GARCH(1,1) is highly significant for DEM, JPY and GBP, and somewhat significant for CAD. The HAC adjustments produce analogous but slightly weaker results. In comparing the MSM model to MS-GARCH, there is some evidence that the multifractal model is a better performer for DEM and CAD, but the significance is marginal at best. Overall, this analysis suggests that in-sample the multifractal matches the performance MS-GARCH and significantly outperforms GARCH(1,1).

5.2. Out-of-Sample Forecasts

We now investigate the out-of-sample forecasting performance of the competing models. We use forecasting horizons ranging from 1 to 50 days. For each currency, we estimate the three processes on the beginning of the series, and use the last twelve years of data (or approximately half the sample) for our out-of-sample forecasting comparison.

Table 7 reports results of one-day forecasts. The first two columns correspond to the coefficients γ_0 and γ_1 from the Mincer-Zarnowitz OLS regressions $x_{t+1}^2 = \gamma_0 + \gamma_1 \mathbb{E}_t x_{t+1}^2 + u_t$ of squared returns on a constant and one-day ahead forecasts. These regressions are common in the financial econometrics literature (e.g., Pagan and Schwert, 1990; West and Cho, 1995), and unbiased forecasts would imply $\gamma_0 = 0$ and $\gamma_1 = 1$. We adjust the standard errors of γ_0 and γ_1 for parameter uncertainty as in West and McCracken (1998), and for HAC effects using the weighting and lag selection methodology of Newey and West (1987, 1994).

The MSM results show that for each currency, the estimated intercept $\hat{\gamma}_0$ is slightly positive and the slope $\hat{\gamma}_1$ is slightly lower than 1. These small biases, however, are not statistically significant. In particular, the hypothesis $\gamma_0 = 0$ is accepted at the 5% confidence level for all currencies, and $\gamma_1 = 1$ is accepted at the 5% level for JPY and CAD and at the 1% level for DEM and GBP. The

[11] Note that a BIC test of GARCH against the multifractal model is identical to a likelihood test since both have the same number of parameters.

Mincer-Zarnowitz regressions thus show little evidence of bias in MSM forecasts.

The point estimates $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are slightly worse with GARCH(1,1) than with the multifractal. All intercepts are more positive, and the slopes are further away from 1 for three currencies. The biases are also statistically significant. The hypotheses $\gamma_0 = 0$ and $\gamma_1 = 1$ are rejected at the 5% level in seven out of eight cases. Since $0 < \hat{\gamma}_1 < 1$, these results suggest that GARCH forecasts are too variable and can be improved by the linear smoothing $\hat{\gamma}_0 + \hat{\gamma}_1 \mathbb{E}_t x_{t+1}^2$. In contrast, Markov-switching GARCH improves on the out-of-sample performance of GARCH(1,1). We accept that $\gamma_0 = 0$ at the 5% confidence level for all currencies, and that $\gamma_1 = 1$ at the 1% level for DEM, GBP and CAD. Furthermore, the regression estimates are best with MS-GARCH for two currencies (DEM and GBP), and with the multifractal for the other two. The two processes thus seem to perform quite similarly out-of-sample at the one-day horizon. We also report in Table 7 two standard measures of goodness of fit: the Mean Squared Error (MSE) and the restricted R^2 coefficient.[12] The multifractal produces the best forecasting R^2 for DEM and JPY. On the other hand, GARCH produces better results for the GBP and MS-GARCH for the CAD. To summarize the one-day forecast results, the multifractal model appears to slightly dominate GARCH(1,1) and to give results comparable to MS-GARCH.

Multistep forecasts provide sharper evidence of differences between the three models. We report in Table 8 the results for 20-day forecasts, which are representative of longer horizons. Since our data sets only contains business days, this frequency corresponds to about a month of calendar time. Following Andersen and Bollerslev (1998) and Klaassen (2002), the dependent variable is the sum of squared daily returns $\sum_{s=t}^{t+19} x_s^2$ over the twenty-day period. Because the average size of returns increases with the sampling interval, the estimated intercepts $\hat{\gamma}_0$ are larger in Table 8 than in Table 7. For each currency, the multifractal produces point estimates of γ_0 and γ_1 that are closest to their preferred values. We also accept the hypotheses $\gamma_0 = 0$ and $\gamma_1 = 1$ in all cases at the 5% confidence level. By contrast, for the other models each currency gives a strong rejection of either one hypothesis (MS-GARCH) or both (GARCH) at the 5% confidence level. The reported MSE and R^2 further confirm that the multifractal provides the best 20-day forecasts for all currencies. The difference is particularly large in the case of

[12] The Mean Squared Error (MSE) quantifies forecast errors in the out-of-sample period: $L^{-1} \sum_{t=T-L+1}^{T} (x_t^2 - \mathbb{E}_{t-1} x_t^2)^2$. The coefficient of determination is defined by $R^2 = 1 - MSE/TSS$, where TSS is the out-of-sample variance of squared returns: $TSS = L^{-1} \sum_{t=T-L+1}^{T} (x_t^2 - \sum_{i=T-L+1}^{T} x_i^2/L)^2$.

the DEM and JPY. The R^2 coefficient is 13.5% and 20.5% respectively for DEM and JPY with the multifractal, while negative values are produced by GARCH and MS-GARCH.[13]

Table 9 reports summary forecasting results and tests of statistical significance for horizons of 1, 5, 10, 20 and 50 days. Panel A shows the forecasting R^2 for each model. For the Mark and the Yen, the multifractal model is already quite dominant at the 5-day horizon, and increasingly outperforms other models at longer horizons. For the Pound and the Canadian Dollar, the multifractal is only dominant at horizons of 20 days and higher. Panel B analyzes the statistical significance of these results. At horizons of 50 days, the multifractal model outperforms the other models very significantly for DEM, GBP and JPY, and at moderate or marginal significance levels for CAD. The superior forecasts of the multifractal are also highly significant at horizons of 10 and 20 days for the DEM, and somewhat strong at the 20-day horizons for JPY and GBP.

These results are quite impressive for the multifractal model. Despite the fact that this is the first forecasting evaluation the MSM, and we have investigated only the very simplest binomial specification of the model, it compares quite well with established models. In particular, GARCH(1,1) is often viewed as a standard benchmark that is very difficult to outperform in forecasting exercises (e.g. West and Cho, 1995; Andersen and Bollerslev, 1998; Hansen and Lunde, 2001). Nonetheless, our results have shown the multifractal model to match or slightly outperform GARCH and MS-GARCH at short horizons, and in many cases to substantially dominate these models at longer horizons. The multifractal model thus appears to be a serious contender among volatility models, and well-deserving of both further empirical and theoretical investigation.

6. Conclusion

This paper proposes an expanded role for regime-switching in modeling volatility. Traditional approaches, such as SWARCH (Cai, 1994; Hamilton and Susmel, 1994) and MS-GARCH (Gray, 1996; Klaassen, 2002), consider separately three categories of volatility dynamics. High-frequency variation is captured by a thick-tailed conditional distribution of returns, mid-range frequencies by smooth

[13] We note that the multifractal yields a higher R^2 for 20-day returns than for daily returns. This stems from the fact that our measure of 20-day volatility is a sum of daily squared returns $\sum_{s=t}^{t+19} x_s^2$. As in Andersen and Bollerslev (1998), the reduction in the noisiness of the volatility measure leads to an increase in explanatory power.

ARCH or GARCH components, and only very low frequencies are modeled with regime-switching. We suggest an alternative approach based on regime-switching dynamics at all frequencies. The model is very tightly parameterized in spite of a high dimensional state space. Using four long series of daily exchange rates, we find that the process matches or dominates the performance of traditional models across a range of in-sample and out-of-sample measures of fit. Thus, the primary contribution of the paper is to show that regime-switching can be used in a much wider variety of settings than previously envisioned. In particular, our pure regime switching model provides a viable alternative to traditional approaches that combine regime-switching, linear volatility dynamics, and flexible tail distributions.

Researchers often focus on applications of immediate practical value when evaluating statistical models. We have similarly shown that our process does well by several standard measures of performance. Another motivation for statistical models is that good description is the first step in explanation. If we can find a statistical model that represents data well among many different dimensions, we can then seek to identify economic mechanisms that might replicate the observed features. Whereas standard models would require us to explain a triumvirate of statistical phenomena – regime-switching, linear dynamics, and thick tails – our model offers the possibility of much more concise explanation. Specifically, our results invite the theorist to explain the economic mechanism that might cause a self-similar form of regime-switching to provide such a useful description of financial data.

Returning to the more immediate contributions of our work, this is the first paper to develop and use a comprehensive set of standard econometric tools to estimate and test multifractal processes. We develop the first maximum likelihood estimator for multifractal processes, based on a straightforward application of the filter developed by Hamilton (1989). We show that this estimator works well in Monte Carlo simulations. The application to exchange rates leads to several observations. First, the likelihood function of the multifractal model decreases monotonically as we add volatility components to the model. This finding is important because it confirms substantial heterogeneity in the persistence of volatility shocks, which is one of the primary motivations of the multifractal approach. Consistent with our intuition, we also observe that the spacing of volatility components across frequencies becomes tighter and the contribution of individual volatility components becomes smaller as we increase the number of components.

The multifractal model is then compared with Student GARCH and MS-

GARCH. Among the many possibilities, these are chosen because of their demonstrated good performance with exchange rate data under a variety of metrics. Like our multifractal model, these models conveniently permit maximum likelihood estimation and analytical forecasting. The in-sample likelihood is significantly higher for the multifractal than for GARCH, even though these processes have the same number of parameters. On the other hand, there is no significant results in-sample between the multifractal model and MS-GARCH.

Out-of-sample evidence further validates the multifractal process. MSM matches or slightly outperform GARCH and MS-GARCH at short horizons, and in many cases substantially dominates these models at longer horizons. The multifractal model thus appears to be a serious contender among volatility models, and well-deserving of both further empirical and theoretical investigation.

7. Appendix

7.1. Covariogram

Let $K_{2q}(n) = \mathbb{E}(|x_t|^{2q}|x_{t+n}|^{2q})/[\mathbb{E}(|x_t|^{2q})]^2$. Multipliers in different stages of the cascade are statistically independent. Since $x_t = \sigma(M_{1,t}M_{2,t}...M_{\overline{k},t})^{1/2}\varepsilon_t$ and $x_{t+n} = \sigma(M_{1,t+n}M_{2,t+n}...M_{\overline{k},t+n})^{1/2}\varepsilon_{t+n}$, we infer that

$$K_{2q}(n) = \frac{\prod_k \mathbb{E}(M_{k,t}^q M_{k,t+n}^q)}{[\mathbb{E}(M^q)]^{2\overline{k}}}.$$

Bayes' rule implies $\mathbb{E}(M_{k,t}^q M_{k,t+n}^q) = \mathbb{E}(M^{2q})(1-\gamma_k)^n + [\mathbb{E}(M^q)]^2[1-(1-\gamma_k)^n]$ or equivalently

$$\mathbb{E}(M_{k,t}^q M_{k,t+n}^q) = [\mathbb{E}(M^q)]^2 \quad [1 + a_q(1-\gamma_k)^n],$$

where $a_q = Var(M^q)/[\mathbb{E}(M^q)]^2$. Since $1-\gamma_k = (1-\gamma_{\bar{k}})^{b^{k-\bar{k}}}$, we obtain

$$K_{2q}(n) = \prod_{k=1}^{\overline{k}} \left[1 + a_q(1-\gamma_{\bar{k}})^{nb^{k-\bar{k}}}\right].$$

We now assume that $n = b^{\bar{k}-l}$. The quantity $(1-\gamma_{\bar{k}})^{nb^{k-\bar{k}}}$ can be rewritten as $e^{b^{k-l}\ln(1-\gamma_{\bar{k}})}$, and thus

$$\log_b K_{2q}(n) = \sum_{i=-l+1}^{\overline{k}-l} \log_b[1 + a_q e^{b^i \ln(1-\gamma_{\bar{k}})}].$$

We note that the terms $e^{-\ln(1-\gamma_{\bar{k}})b^i}$ are close to 1 when $i \in \{-l, ..., 0\}$ and negligible for larger values of i, suggesting that $\log_b K_{2q}(n) \sim l \log_b(1+a_q)$. This heuristic argument can be proved by dividing the sum into three parts:

$$\sum_{i=-l+1}^{-\sqrt{l}} \log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b^i}] + \sum_{i=-\sqrt{l}+1}^{\sqrt{l}} \log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b^i}] + \sum_{i=\sqrt{l}+1}^{\overline{k}-l} \log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b^{-i}}].$$

We note that the first component is bounded below $(l-\sqrt{l})\log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b^{-l}}]$ and above by $(l-\sqrt{l})\log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b^{-\sqrt{l}}}]$, and is thus equivalent to $l\log_b(1+a_q)$. The second term is bounded above by $2\sqrt{l}\log_b(1+a_q)$ and is therefore $O(\sqrt{l})$. The third term is bounded above by $\sum_{i=\sqrt{l}+1}^{\bar{k}-l}\log_b[1+a_q e^{\ln(1-\gamma_{\bar{k}})b\sqrt{l}}]$ and is therefore negligible. Hence

$$K_{2q}(n)\sim(b^{\bar{k}}/n)^{\log_b(1+a_q)}.$$

We also note that $1+a_q=\mathbb{E}(M^{2q})/[\mathbb{E}(M^q)]^2>1$, and thus $\log_b(1+a_q)=\log_b\mathbb{E}(M^{2q})-2\log_b\mathbb{E}(M^q)>0$.

We can now derive the covariogram of the process. Since $\gamma_{2q}(n)=K_{2q}(n)-[\mathbb{E}(|x_t|^{2q})]^2$ and $[\mathbb{E}(|x_t|^{2q})]^2=\sigma^{4q}[\mathbb{E}(\,|\varepsilon_t|^{2q})]^2[\mathbb{E}(M^q)]^{2\bar{k}}$, we infer that

$$\gamma_{2q}(n)=c_{q,\bar{k}}\,\left\{\prod_{k=1}^{\bar{k}}[1+a_q(1-\gamma_1)^{nb^{k-1}}]-1\right\}.$$

When $1<<l<<\bar{k}$, we know that $\prod_{k=1}^{\bar{k}}[1+a_q(1-\gamma_1)^{nb^{k-1}}]\sim(b^{\bar{k}}/n)^{\log_b(1+a_q)}>>1$ and thus $\gamma_{2q}(n)\sim K_{2q}(n)\propto n^{\tau_\theta(2q)-2\tau_\theta(q)-1}$.

7.2. Bayesian Updating

Denote the set of past observations $\mathcal{I}_t\equiv\{x_s\}_{s=1}^t$. We note that the conditional probability $\Pi_{t+1}^j=\mathbb{P}\left(M_{t+1}=m^j\,|\mathcal{I}_t,x_{t+1}\right)$ satisfies

$$\Pi_{t+1}^j=f_{x_{t+1}}\left(x_{t+1}|M_{t+1}=m^j\right)\mathbb{P}\left(M_{t+1}=m^j\,|\mathcal{I}_t\right)/f_{x_{t+1}}\left(x_{t+1}\,|\mathcal{I}_t\right),$$

which can be rewritten as

$$\Pi_{t+1}^j=\frac{n[x_{t+1}/\sigma g\left(m^j\right)]\left(\sum_{i=1}^d a_{ij}\Pi_t^i\right)}{\sigma g\left(m^j\right)f_{x_t}\left(x_{t+1}\,|\mathcal{I}_t\right)}. \tag{7.1}$$

This implies (3.1) since $\sum_j\Pi_{t+1}^j=1$.

7.3. Likelihood Function

We know that $L\left(x_1,...,x_T;\psi\right)=\sum_{t=1}^T\ln f(x_t\,|x_1,...,x_{t-1})$. Bayes' rule implies that

$$\begin{aligned}f(x_t\,|x_1,...,x_{t-1}) &= \sum_{i=1}^d\mathbb{P}(M_t=m^i|x_1,...,x_{t-1})f(x_t|M_t=m^i)\\ &= \sum_{i=1}^d\mathbb{P}(M_t=m^i|x_1,...,x_{t-1})\frac{1}{\sigma g\left(m^i\right)}n\left(\frac{x}{\sigma g\left(m^i\right)}\right)\end{aligned}$$

and thus $f(x_t\,|x_1,...,x_{t-1}) = \omega(x_t) \cdot (\Pi_{t-1}A)$.

7.4. HAC-Adjusted Vuong Test

We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}^0)$ and a stochastic process $\{x_t\}_{t=-\infty}^{+\infty}$. Each x_t is a random variable taking values on the real line. For every t, it is convenient to consider the vector of past values $X_{t-1} = \{x_s\}_{s=-\infty}^{t-1}$. The econometrician directly observes a finite number of realizations of x_t, but ignores the true data generating process. She instead considers two competing families of models specified by their conditional densities $\mathcal{M}_f = \{f(x_t|X_{t-1}, \theta);\ \theta \in \Theta\}$ and $\mathcal{M}_g = \{g(x_t|X_{t-1}, \gamma);\ \gamma \in \Gamma\}$. These families may or may not contain the true data generating process. The pseudo true value θ^* specifies the model in $\mathcal{M}_f$ with the optimal Kullback-Leibler Information Criterion:

$$\theta^* = \arg\max_{\theta\in\Theta} \quad \mathbb{E}^0[\ln f(x_t|X_{t-1}, \theta)].$$

The pseudo true value γ^* is similarly defined.

Consider the log-likelihood functions:

$$L_T^f(\theta) \equiv \sum_{t=1}^{T} \ln f(x_t|X_{t-1}, \theta), \qquad L_T^g(\gamma) \equiv \sum_{t=1}^{T} \ln g(x_t|X_{t-1}, \gamma).$$

By definition, the MLE estimators $\hat{\theta}_T$ and $\hat{\gamma}_T$ maximize the functions $L_T^f(\theta)$ and $L_T^g(\gamma)$. The corresponding FOCs are

$$\frac{\partial L_T^f}{\partial \theta}(\hat{\theta}_T) = 0, \quad \frac{\partial L_T^g}{\partial \theta}(\hat{\gamma}_T) = 0. \tag{7.2}$$

We now examine the likelihood ratio

$$LR_T(\hat{\theta}_T, \hat{\gamma}_T) = L_T^f(\hat{\theta}_T) - L_T^g(\hat{\gamma}_T) = \sum_{t=1}^{T} \ln \frac{f(x_t|X_{t-1}, \hat{\theta}_T)}{g(x_t|X_{t-1}, \hat{\gamma}_T)}.$$

By equation (7.2), a second order expansion of LR_T implies that $\frac{1}{\sqrt{T}} LR_T(\hat{\theta}_T, \hat{\gamma}_T) = \frac{1}{\sqrt{T}} LR_T(\theta^*, \gamma^*) + o_p(1)$, and thus

$$\frac{1}{\sqrt{T}} LR_T(\hat{\theta}_T, \hat{\gamma}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \ln \frac{f(x_t|X_{t-1}, \theta^*)}{g(x_t|X_{t-1}, \gamma^*)} + o_p(1).$$

Let $a_t = \ln[f(x_t|X_{t-1}, \theta^*)/g(x_t|X_{t-1}, \gamma^*)]$ and $\hat{a}_t = \ln[f(x_t|X_{t-1}, \hat{\theta}_T)/g(x_t|X_{t-1}, \hat{\gamma}_T)]$.

When the observations x_t are IID, the addends a_t are also IID. If the models f and g have equal Kullback-Leibler Information criterion, the CLT implies

$$\frac{1}{\sqrt{T}} LR_T(\hat{\theta}_T, \hat{\gamma}_T) \xrightarrow{d} \mathcal{N}(0, \sigma_*^2),$$

where $\sigma_*^2 = Var(a_t)$. The variance is consistently estimated by the sample variance of $\{\hat{a}_t\}$.

In the non-IID case, we need to adjust for the correlation in the addends a_t. Let

$$\sigma_T^2 = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}(a_s a_t).$$

We know that $\frac{1}{\sqrt{T}} LR_T(\hat{\theta}_T, \hat{\gamma}_T) = \sigma_T Z + o_p(1)$, where Z is a standard Gaussian. Following Newey-West (1987), we estimate σ_T by

$$\hat{\sigma}_T^2 = \hat{\Omega}_0 + 2 \sum_{j=1}^{m_T} w(j, m) \hat{\Omega}_j,$$

where $\hat{\Omega}_j = \sum_{t=j+1}^{T} \hat{a}_t \hat{a}_{t-j}/T$ denotes the sample covariance of $\{\hat{a}_t\}$, and $w(j, m) = 1 - j/(m + 1)$ is the Bartlett weight. We choose m_T using the automatic lag selection method of Newey and West (1994).

References

[1] Akgiray, V. (1989), Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts, *Journal of Business* **62**, 55-80.

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

[3] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association* **96**,42-55.

[4] Andrews, D. W. K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* **59**, 817-854.

[5] Andrews, D. W. K., and Monahan, J. C. (1992), An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* **60**, 953-966.

[6] Baillie, R. T., (1996), Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics* **73**, 5-59.

[7] Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **74**, 3-30.

[8] Barndorff-Nielsen, O., and Shephard, N. (2003), Realized Power Variation and Stochastic Volatility, *Bernoulli*, forthcoming.

[9] Bollen, N., Gray, S. and Whaley, R. (2000), Regime Switching in Foreign Exchange Rates: Evidence from Currency Option Prices, *Journal of Econometrics* **94**, 239-276.

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

[11] Bollerslev, T. (1987), A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics* **69**, 542-547.

[12] Breidt, J., Crato, N., and DeLima, P. (1998), On the Detection and Estimation of Long-Memory in Stochastic Volatility, *Journal of Econometrics* **83**, 325-348.

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

[14] Calvet, L. (2001), Incomplete Markets and Volatility, *Journal of Economic Theory* **98**, 295-338.

[15] Calvet, L., and Fisher, A. (2001), Forecasting Multifractal Volatility, *Journal of Econometrics* **105**, 27-58.

[16] Calvet, L., and Fisher, A. (2002), Multifractality in Asset Returns: Theory and Evidence, *Review of Economics and Statistics* **84**, 381-406.

[17] Calvet, L., Fisher, A., and Mandelbrot, B. B. (1997), Cowles Foundation Discussion Papers No. 1164-1166, Yale University. Papers available from http://www.ssrn.com.

[18] Campbell, J., Lo, A., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

[19] Dacorogna, M., Müller, U., Nagler, R., Olsen, R., and Pictet, O. (1993), A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market, *Journal of International Money and Finance* **12**, 413-438.

[20] den Haan, W. J., and Levin, A. (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation", in G. Maddala and C. Rao Eds., *Handbook of Statistics: Robust Inference*, Vol. 15, New York: Elsevier.

[21] Diebold. F. X., and Inoue, A. (2001), Long Memory and Regime Switching, *Journal of Econometrics* **105**, 27-58.

[22] Diebold, F. X., Lee, J. H., and Weinbach, G. C. (1994), Regime Switching with Time-Varying Transition Probabilities, in C. Hargreaves ed. *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press.

[23] Ding, Z., Granger, C., and Engle, R. (1993), A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* **1**, 83-106.

[24] Dueker, M. (1997), Markov Switching in GARCH Processes and Mean Reverting Volatility, *Journal of Business and Economic Statistics* **15**, 26-34.

[25] Durland, J., and McCurdy, T. (1994), Duration-Dependent Transitions in a Markov Model of US GNP Growth, *Journal of Business and Economic Statistics* **12**, 279-288.

[26] Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* **50**, 987-1007.

[27] Filardo, A. J. (1994), Business Cycle Phases and their Transitional Dynamics, *Journal of Business and Economic Statistics* **12**, 299-308.

[28] Frisch, U., and Parisi, G. (1985), Fully Developed Turbulence and Intermittency, in: M. Ghil ed., *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, 84-88, Amsterdam: North-Holland.

[29] Garcia, R., and Perron, P. (1996), An Analysis of the Real Interest Rate under Regime Shifts, *Review of Economics and Statistics* **78**, 111-125.

[30] Ghysels, E., Harvey, A. C., and Renault, E. (1996), Stochastic Volatility, in: G. S. Maddala and C. R. Rao Eds., *Handbook of Statistics*, Vol. 14, 119-191. Amsterdam: North-Holland.

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

[32] Hamilton, J. D. (1989), A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* **57**, 357-84.

[33] Hamilton, J. D. (1990), Analysis of Time Series Subject to Change in Regime, *Journal of Econometrics* **45**, 39-70.

[34] Hamilton, J. D. and B. Raj (2002), New Directions in Business Cycle Research and Financial Analysis, *Empirical Economics* **27**, 149-162.

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

[36] Hansen, P., and Lunde, A. (2001), "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?", Working Paper, Brown University.

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

[38] LeBaron, B. (2001), Stochastic Volatility as a Simple Generator of Apparent Financial Power Laws and Long Memory, *Quantitative Finance* **1**, 621-631.

[39] Lux, T. (2001), "The Multifractal Model of Asset Returns: Simple Moment and GMM Estimation", Working Paper, Kiel University.

[40] Maheu, J., and McCurdy, T. (2000), Volatility Dynamics under Duration-Dependent Mixing, *Journal of Empirical Finance* **7**, 345-372

[41] Mandelbrot, B. B. (1963), The Variation of Certain Speculative Prices, *Journal of Business* **36**, 394-419.

[42] Mandelbrot, B. B. (1997), *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, New York: Springer Verlag.

[43] Mariano, R., Schuerman, T., and M. Weeks (1999), *Simulation-Based Inference in Econometrics*, Cambridge University Press.

[44] Newey, W., and West, K. (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703-708.

[45] Newey, W., and West, K. (1994), Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies* **61**, 631-654.

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

[47] Perez-Quiros, G., and Timmerman, A. (2000), Firm Size and Cyclical Variations in Stock Returns, *Journal of Finance* **55**, 1229-1262.

[48] Rivers, D., and Vuong, Q. (2002), Model Selection Tests for Nonlinear Dynamic Models, *Econometrics Journal* **5** (1), 1-39.

[49] Stock, J. H. (1987), Measuring Business Cycle Time, *Journal of Political Economy* **95**, 1240-1261.

[50] Stock, J. H. (1988), Estimating Continuous-Time Processes Subject to Time Deformation, *Journal of the American Statistical Association* **83**, 77-85.

[51] Vuong, Q. (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica* **57**, 307-333.

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

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

TABLE 1. – MONTE CARLO MLE RESULTS

	$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\bar{m}_{sim}$	1.288	1.293	1.297	1.392	1.393	1.397	1.494	1.494	1.497
FSSE	(0.026)	(0.018)	(0.012)	(0.031)	(0.019)	(0.015)	(0.032)	(0.025)	(0.015)
RMSE	(0.029)	(0.019)	(0.012)	(0.032)	(0.021)	(0.015)	(0.033)	(0.025)	(0.016)
AASE	(0.019)	(0.013)	(0.010)	(0.018)	(0.014)	(0.011)	(0.019)	(0.014)	(0.011)
$\bar{\sigma}_{sim}$	1.014	1.004	0.999	1.031	1.011	1.005	1.026	1.017	1.006
FSSE	(0.167)	(0.102)	(0.073)	(0.221)	(0.147)	(0.091)	(0.255)	(0.224)	(0.105)
RMSE	(0.167)	(0.102)	(0.073)	(0.223)	(0.148)	(0.092)	(0.256)	(0.225)	(0.105)
AASE	(0.072)	(0.056)	(0.046)	(0.091)	(0.075)	(0.057)	(0.111)	(0.088)	(0.067)
$\bar{\gamma}_{sim}$	0.867	0.908	0.934	0.907	0.935	0.940	0.924	0.938	0.944
FSSE	(0.164)	(0.113)	(0.068)	(0.111)	(0.069)	(0.047)	(0.087)	(0.055)	(0.037)
RMSE	(0.184)	(0.120)	(0.070)	(0.119)	(0.070)	(0.048)	(0.091)	(0.056)	(0.037)
AASE	(0.181)	(0.114)	(0.071)	(0.110)	(0.068)	(0.048)	(0.078)	(0.052)	(0.036)
$\bar{b}_{sim}$	2.853	2.942	2.988	3.052	2.938	2.973	3.054	2.987	2.979
FSSE	(0.934)	(0.670)	(0.416)	(0.963)	(0.480)	(0.363)	(0.735)	(0.565)	(0.316)
RMSE	(0.945)	(0.673)	(0.416)	(0.964)	(0.484)	(0.364)	(0.737)	(0.565)	(0.317)
AASE	(0.677)	(0.471)	(0.331)	(0.599)	(0.367)	(0.264)	(0.476)	(0.323)	(0.225)

Notes: This table is based on $J = 400$ simulated paths for each column. All simulations are based on a multifractal process with $\bar{k} = 8$. The columns are distinguished by combinations of $m_0 \in \{1.3, 1.4, 1.5\}$ and sample lengths of $T_1 = 2500$, $T_2 = 5000$, and $T_3 = 10000$. Parameters that are fixed across all simulations are $\sigma = 1$, $\gamma_k = 0.95$, and $b = 3$. These parameters provide a reasonable approximation to values that are estimated on exchange rate data in future sections of the paper. In each set of four rows, the first row is the average MLE parameter value over the J simulated paths. In the remaining rows, FSSE denotes finite sample standard error, RMSE the root mean squared error, and AASE the average asymptotic standard error. To derive AASE, the asymptotic variance is calculated for each $j \in \{1, .., J\}$ path from the inverse of the information matrix and the average is taken over the J simulations.

TABLE 2. – FX Return Variability

	Root of Average Squared Return				
	Entire	By Subperiod			
	Sample	1	2	3	4
DEM	0.664	0.587	0.717	0.709	0.635
JPY	0.657	0.545	0.641	0.646	0.775
GBP	0.607	0.486	0.725	0.699	0.473
CAD	0.274	0.220	0.255	0.284	0.327

Notes: For each data set, the first column shows the square root of the average of squared returns over the entire sample period. In the next four columns, each data series is broken into quarters and the same statistic is calculated for each subperiod. The results show that the variability of return variance is substantial even at very low frequencies.

TABLE 3. – Maximum Likelihood Results

	$\bar{k}=1$	2	3	4	5	6	7	8	9	10
					Deutsche Mark / US Dollar					
$\hat{m}$	1.654	1.590	1.555	1.492	1.462	1.413	1.380	1.353	1.351	1.326
	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.012)	(0.011)	(0.013)	(0.015)
$\hat{\sigma}$	0.682	0.651	0.600	0.572	0.512	0.538	0.547	0.550	0.674	0.643
	(0.012)	(0.018)	(0.014)	(0.016)	(0.018)	(0.026)	(0.021)	(0.025)	(0.035)	(0.073)
$\hat{\gamma}_{\bar{k}}$	0.075	0.107	0.672	0.714	0.751	0.858	0.932	0.974	0.966	0.959
	(0.011)	(0.022)	(0.151)	(0.096)	(0.106)	(0.128)	(0.071)	(0.042)	(0.065)	(0.066)
$\hat{b}$	-	8.01	21.91	10.42	7.89	5.16	4.12	3.38	3.29	2.70
		(2.58)	(7.30)	(1.92)	(1.31)	(0.76)	(0.48)	(0.36)	(0.47)	(0.36)
ln L	-5920.86	-5782.96	-5731.78	-5715.31	-5708.25	-5706.91	-5704.48	-5704.77	-5704.86	-5705.09
					Japanese Yen / US Dollar					
$\hat{m}$	1.797	1.782	1.693	1.654	1.640	1.573	1.565	1.513	1.475	1.448
	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
$\hat{\sigma}$	0.630	0.538	0.566	0.462	0.709	0.642	0.518	0.514	0.486	0.461
	(0.011)	(0.009)	(0.017)	(0.013)	(0.023)	(0.023)	(0.018)	(0.020)	(0.026)	(0.036)
$\hat{\gamma}_{\bar{k}}$	0.199	0.345	0.312	0.697	0.778	0.899	0.897	0.975	0.995	0.998
	(0.019)	(0.033)	(0.054)	(0.080)	(0.076)	(0.060)	(0.057)	(0.034)	(0.010)	(0.006)
$\hat{b}$	-	134.20	12.46	15.58	16.03	8.07	7.46	5.65	4.43	3.76
		(48.27)	(2.18)	(2.67)	(2.67)	(1.03)	(0.89)	(0.78)	(0.53)	(0.45)
ln L	-6451.80	-6102.18	-5959.72	-5900.67	-5882.93	-5871.35	-5867.88	-5863.20	-5863.01	-5862.68
					British Pound / US Dollar					
$\hat{m}$	1.716	1.671	1.648	1.609	1.579	1.534	1.503	1.461	1.428	1.403
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.009)
$\hat{\sigma}$	0.609	0.590	0.513	0.467	0.421	0.468	0.389	0.384	0.374	0.370
	(0.009)	(0.011)	(0.016)	(0.016)	(0.017)	(0.019)	(0.014)	(0.015)	(0.022)	(0.022)
$\hat{\gamma}_{\bar{k}}$	0.110	0.222	0.278	0.645	0.637	0.784	0.811	0.958	0.964	0.982
	(0.017)	(0.034)	(0.052)	(0.080)	(0.075)	(0.078)	(0.083)	(0.052)	(0.043)	(0.031)
$\hat{b}$	-	19.90	14.29	12.51	11.02	8.32	6.72	5.23	4.08	3.45
		(5.19)	(2.58)	(2.00)	(1.74)	(1.15)	(0.91)	(0.69)	(0.41)	(0.32)
ln L	-5960.18	-5724.37	-5622.73	-5570.02	-5537.80	-5523.64	-5516.89	-5515.37	-5515.28	-5514.94
					Canadian Dollar / US Dollar					
$\hat{m}$	1.646	1.556	1.474	1.435	1.386	1.374	1.338	1.319	1.296	1.278
	(0.012)	(0.012)	(0.014)	(0.015)	(0.012)	(0.013)	(0.012)	(0.016)	(0.013)	(0.012)
$\hat{\sigma}$	0.280	0.278	0.293	0.263	0.251	0.295	0.282	0.262	0.259	0.262
	(0.005)	(0.006)	(0.014)	(0.009)	(0.010)	(0.011)	(0.013)	(0.017)	(0.015)	(0.021)
$\hat{\gamma}_{\bar{k}}$	0.064	0.109	0.129	0.171	0.441	0.524	0.593	0.594	0.631	0.644
	(0.009)	(0.016)	(0.040)	(0.062)	(0.153)	(0.128)	(0.145)	(0.151)	(0.155)	(0.158)
$\hat{b}$	-	10.92	4.76	3.95	4.02	4.08	3.11	2.72	2.35	2.11
		(3.12)	(1.15)	(0.83)	(0.76)	(0.58)	(0.39)	(0.39)	(0.25)	(0.18)
ln L	-271.01	-129.80	-105.16	-91.32	-88.41	-84.73	-84.03	-83.40	-83.06	-83.00

Notes: This table shows maximum likelihood estimation results for the binomial multifractal model for all four exchange rate series. Columns correspond to the number of frequencies $\bar{k}$ in the estimated model. The likelihood function increases monotonically in the number of volatility frequencies for all data sets except DEM, which obtains a maximum at $\bar{k}=7$. Asymptotic standard errors are in parenthesis.

TABLE 4. – Multifractal Model Selection

	$\bar{k}=1$	2	3	4	5	6	7	8	9
				A. Vuong (1989) Test					
Mark	-8.655	-5.523	-2.972	-1.858	-0.688	-0.733	0.341	0.204	0.337
	(0.000)	(0.000)	(0.001)	(0.032)	(0.246)	(0.232)	(0.633)	(0.581)	(0.632)
Yen	-13.067	-8.406	-5.342	-3.154	-2.156	-1.192	-1.108	-0.180	-0.162
	(0.000)	(0.000)	(0.000)	(0.001)	(0.016)	(0.117)	(0.134)	(0.429)	(0.436)
Pound	-11.810	-8.337	-6.267	-4.360	-2.984	-1.334	-0.408	-0.149	-0.236
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.089)	(0.342)	(0.441)	0.407)
Canada	-8.475	-4.421	-3.289	-1.795	-2.108	-0.862	-0.825	-0.472	-0.158
	(0.000)	(0.000)	(0.000)	(0.036)	(0.017)	(0.194)	(0.205)	(0.318)	(0.437)
				B. HAC Adjusted Vuong Test					
Mark	-4.285	-3.033	-1.683	-1.101	-0.402	-0.424	0.197	0.120	0.194
	(0.000)	(0.001)	(0.046)	(0.135)	(0.344)	(0.336)	(0.578)	(0.548)	(0.577)
Yen	-5.219	-4.262	-2.865	-1.645	-1.224	-0.648	-0.663	-0.105	-0.098
	(0.000)	(0.000)	(0.002)	(0.050)	(0.111)	(0.259)	(0.254)	(0.458)	(0.461)
Pound	-3.788	-2.804	-2.803	-2.195	-1.759	-0.779	-0.242	-0.088	-0.137
	(0.000)	(0.003)	(0.003)	(0.014)	(0.039)	(0.218)	(0.404)	(0.465)	(0.446)
Canada	-4.237	-2.383	-1.789	-1.019	-1.150	-0.480	-0.445	-0.276	-0.091
	(0.000)	(0.009)	(0.037)	(0.154)	(0.125)	(0.316)	(0.328)	(0.391)	(0.464)

Notes: This table reports t-ratios and one-sided p-values for the log-likelihood difference of the model in each column against the multifractal with ten frequencies. Panel A uses the Vuong (1989) methodology and Panel B adjusts for heteroskedasticity and autocorrelation using Newey and West (1987, 1994). A low p-value indicates that the corresponding model would be rejected in favor of the multifractal with ten frequencies.

TABLE 5. – Alternative Processes

	$1/\nu$	Regime 1 σ_1	α_1	β_1	p_{11}	Regime 2 σ_2	α_2	β_2	p_{22}	$\ln L$
					Deutsche Mark / US Dollar					
GARCH	0.1929	1.5539	0.0879	0.9108						-5730.52
	(0.011)	(0.405)	(0.009)	(0.009)						
MS-GARCH	0.2041	1.0749	0.2048	0.7896	0.9998	1.3145	0.0718	0.9241	0.9999	-5694.78
	(0.011)	(0.288)	(0.023)	(0.024)	(0.0003)	(0.282)	(0.010)	(0.011)	(0.0002)	
					Japanese Yen / US Dollar					
GARCH	0.2290	0.1638	0.0652	0.9348						-5965.07
	(0.0002)	(0.059)	(0.006)	(0.006)						
MS-GARCH	0.2632	0.4443	0.3420	0.6500	0.9999	0.9639	0.0650	0.9227	0.9999	-5833.59
	(0.012)	(0.137)	(0.040)	(0.040)	(0.0002)	(0.121)	(0.010)	(0.013)	(0.0002)	
					British Pound / US Dollar					
GARCH	0.2007	0.2365	0.0681	0.9319						-5562.00
	(0.0077)	(0.070)	(0.0046)	(0.0046)						
MS-GARCH	0.2202	0.8423	0.3653	0.6051	0.9860	0.9343	0.0587	0.9365	0.9986	-5492.44
	(0.009)	(0.013)	(0.053)	(0.056)	(0.005)	(0.012)	(0.008)	(0.008)	(0.0003)	
					Canadian Dollar / US Dollar					
GARCH	0.1528	0.3108	0.0810	0.9108						-96.03
	(0.037)	(0.008)	(0.008)	(0.010)						
MS-GARCH	0.1385	0.2046	0.0584	0.9361	0.9896	0.2972	0.2587	0.2925	0.9415	-73.51
	(0.011)	(0.035)	(0.009)	(0.010)	(0.004)	(0.025)	(0.074)	(0.215)	(0.023)	

Notes: This table shows maximum likelihood estimation results for alternative processes for the four exchange rate series. Asymptotic standard errors are in parenthesis. For the GARCH(1,1) model, the parameter estimates for JPY/USD and GBP/USD are on the boundary of the restriction $\alpha + \beta \leq 1 - \epsilon$, where $\epsilon = 10^{-5}$.

TABLE 6. – In-Sample Model Comparison

	No. of Parameters	$\ln L$	BIC	BIC p-value vs. Multifractal: Vuong (1989)	BIC p-value vs. Multifractal: HAC Adj
		Deutsche Mark / US Dollar			
Binomial Multifractal	4	-5705.09	1.7830		
GARCH	4	-5730.52	1.7910	0.005	0.071
MS-GARCH	9	-5694.78	1.7866	0.140	0.248
		Japanese Yen / US Dollar			
Binomial Multifractal	4	-5862.68	1.6115		
GARCH	4	-5965.07	1.6396	0.000	0.008
MS-GARCH	9	-5833.59	1.6097	0.619	0.572
		British Pound / US Dollar			
Binomial Multifractal	4	-5514.94	1.5162		
GARCH	4	-5562.00	1.5291	0.004	0.070
MS-GARCH	9	-5492.44	1.5162	0.505	0.503
		Canadian Dollar / US Dollar			
Binomial Multifractal	4	-83.00	0.0286		
GARCH	4	-96.03	0.0323	0.072	0.200
MS-GARCH	9	-73.51	0.0322	0.092	0.235

Notes: This table summarizes information about in-sample goodness of fit for the three models. The Bayesian Information Criterion is given by $BIC = T^{-1}(-2\ln L + NP\ln T)$. The sample lengths are 6419 for DEM/USD, 7298 for JPY/USD and GBP/USD, and 7048 for CAD/USD. The last two columns give p-values from a test that the corresponding model dominates the multifractal model by the BIC criterion. The first value uses the Vuong (1989) methodology, and the second value adjusts the test for heteroskedasticity and autocorrelation. A low p-value indicates that the corresponding model would be rejected in favor of the multifractal model.

TABLE 7. – ONE-DAY FORECASTS

	Mincer-Zarnowitz		Restricted $\gamma_0 = 0, \gamma_1 = 1$	
	γ_0	γ_1	MSE	R^2
Deutsche Mark / US Dollar				
Binomial Multifractal	0.098	0.703	0.7263	0.041
	(0.072)	(0.126)		
GARCH	0.153	0.622	0.7304	0.035
	(0.061)	(0.105)		
MS-GARCH	0.042	0.740	0.7296	0.037
	(0.080)	(0.130)		
Japanese Yen / US Dollar				
Binomial Multifractal	0.028	0.772	1.6053	0.053
	(0.090)	(0.117)		
GARCH	0.172	0.668	1.6137	0.048
	(0.075)	(0.105)		
MS-GARCH	0.080	0.709	1.6141	0.048
	(0.084)	(0.109)		
British Pound / US Dollar				
Binomial Multifractal	0.053	0.715	0.5081	0.057
	(0.049)	(0.100)		
GARCH	0.085	0.751	0.4980	0.076
	(0.044)	(0.098)		
MS-GARCH	0.017	0.814	0.4997	0.072
	(0.051)	(0.108)		
Canadian Dollar / US Dollar				
Binomial Multifractal	0.015	0.905	0.0345	0.051
	(0.016)	(0.156)		
GARCH	0.033	0.679	0.0348	0.042
	(0.012)	(0.111)		
MS-GARCH	0.025	0.785	0.0344	0.055
	(0.013)	(0.124)		

Notes: This table gives out of sample forecasting results for the three models. The first two columns correspond to parameter estimates from the Mincer-Zarnowitz OLS regression $e_{t+1}^2 = \gamma_0 + \gamma_1 E_t(e_{t+1}^2) + u_t$. For an unbiased forecast we expect $\gamma_0 = 0$ and $\gamma_1 = 1$. Asymptotic standard errors in parenthesis are corrected for heteroskedasticity and autocorrelation using the method of Newey and West(1987,1994) and for parameter uncertainty using the method of West and McCracken(1998). MSE is the mean square forecast error, and R^2 is one less the MSE divided by the sum of squared demeaned squared returns in the out of sample period.

TABLE 8. – Twenty-Day Forecasts

	Mincer-Zarnowitz		Restricted $\gamma_0 = 0, \gamma_1 = 1$	
	γ_0	γ_1	MSE	R^2
		Deutsche Mark / US Dollar		
Binomial Multifractal	1.749	0.706	37.12	0.135
	(1.649)	(0.150)		
GARCH	4.474	0.443	49.24	-0.147
	(1.108)	(0.092)		
MS-GARCH	1.934	0.568	50.66	-0.180
	(1.577)	(0.118)		
		Japanese Yen / US Dollar		
Binomial Multifractal	-1.248	0.909	76.95	0.205
	(2.160)	(0.155)		
GARCH	5.311	0.488	99.15	-0.024
	(1.233)	(0.086)		
MS-GARCH	2.148	0.573	103.29	-0.067
	(1.776)	(0.108)		
		British Pound / US Dollar		
Binomial Multifractal	0.330	0.792	27.35	0.250
	(1.114)	(0.120)		
GARCH	2.702	0.606	29.61	0.188
	(0.760)	(0.085)		
MS-GARCH	0.641	0.730	29.08	0.203
	(1.021)	(0.105)		
		Canadian Dollar / US Dollar		
Binomial Multifractal	-0.038	1.179	1.6339	0.217
	(0.385)	(0.221)		
GARCH	0.676	0.707	1.6615	0.204
	(0.243)	(0.121)		
MS-GARCH	0.630	0.754	1.6719	0.199
	(0.270)	(0.140)		

Notes: This table gives out of sample forecasting results for the three models. The first two columns correspond to parameter estimates from the Mincer-Zarnowitz OLS regression $e_{t+1}^2 = \gamma_0 + \gamma_1 E_t(e_{t+1}^2) + u_t$. For an unbiased forecast we expect $\gamma_0 = 0$ and $\gamma_1 = 1$. Asymptotic standard errors in parenthesis are corrected for heteroskedasticity and autocorrelation using the method of Newey and West(1987,1994) and for parameter uncertainty using the method of West and McCracken(1998). MSE is the mean square forecast error, and R^2 is one less the MSE divided by the sum of squared demeaned squared returns in the out of sample period.

TABLE 9. – Forecast Summary, Multiple Horizons

	Horizon (Days)				
	1	5	10	20	50
A. Restricted R^2					
Deutsche Mark / US Dollar					
Binomial Multifractal	0.041	0.124	0.160	0.135	0.038
GARCH	0.035	0.069	0.033	-0.147	-0.761
MS-GARCH	0.039	0.072	0.030	-0.180	-1.137
Japanese Yen / US Dollar					
Binomial Multifractal	0.053	0.113	0.142	0.205	0.213
GARCH	0.048	0.054	0.011	-0.024	-0.358
MS-GARCH	0.048	0.044	-0.009	-0.067	-0.569
British Pound / US Dollar					
Binomial Multifractal	0.057	0.165	0.235	0.250	0.273
GARCH	0.076	0.191	0.244	0.188	-0.026
MS-GARCH	0.072	0.165	0.238	0.203	0.038
Canadian Dollar / US Dollar					
Binomial Multifractal	0.051	0.172	0.221	0.217	0.111
GARCH	0.042	0.154	0.205	0.204	0.070
MS-GARCH	0.055	0.181	0.229	0.199	0.036
B. MSE Test vs. Multifractal (p-value)					
Deutsche Mark / US Dollar					
GARCH	0.307	0.040	0.009	0.001	0.000
MS-GARCH	0.314	0.004	0.000	0.000	0.000
Japanese Yen / US Dollar					
GARCH	0.426	0.208	0.144	0.117	0.063
MS-GARCH	0.415	0.143	0.071	0.021	0.000
British Pound / US Dollar					
GARCH	0.906	0.824	0.606	0.156	0.016
MS-GARCH	0.857	0.499	0.547	0.108	0.000
Canadian Dollar / US Dollar					
GARCH	0.294	0.3590	0.410	0.447	0.292
MS-GARCH	0.597	0.603	0.565	0.380	0.065

Notes: This table summarizes out of sample forecasting results across multiple horizons. Panel A gives the restricted forecasting R^2 for each model and horizon. Panel B gives p-values from testing that the corresponding model has a lower out of sample forecasting MSE than the binomial multifractal. The tests are corrected for autocorrelation and heteroskedasticity using Newey and West (1987, 1994). A low p-value indicates that forecasts from the corresponding model would be rejected in favor of multifractal forecasts.

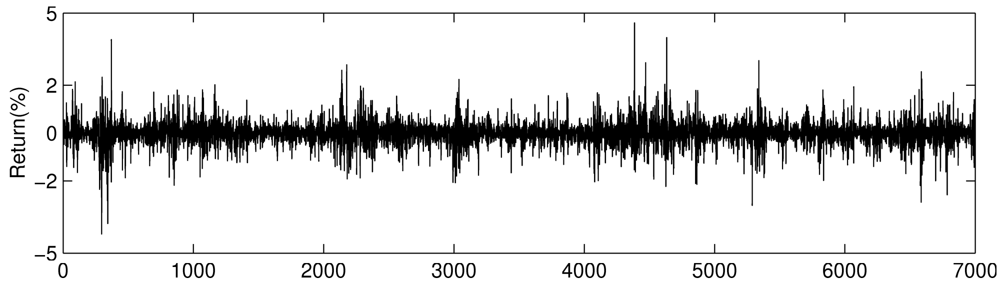

Figure 1: Simulated Multifractal Process. This figure shows simulated log price differences of a multifractal process. The process has $\bar{k} = 8$ frequencies and parameter values $m = 1.4$, $\sigma = 0.5$, $\gamma_{\bar{k}} = 0.95$, and $b = 3$. These parameter values are roughly consistent with estimates that are found to provide a good description of several exchange rate series in later sections of the paper.

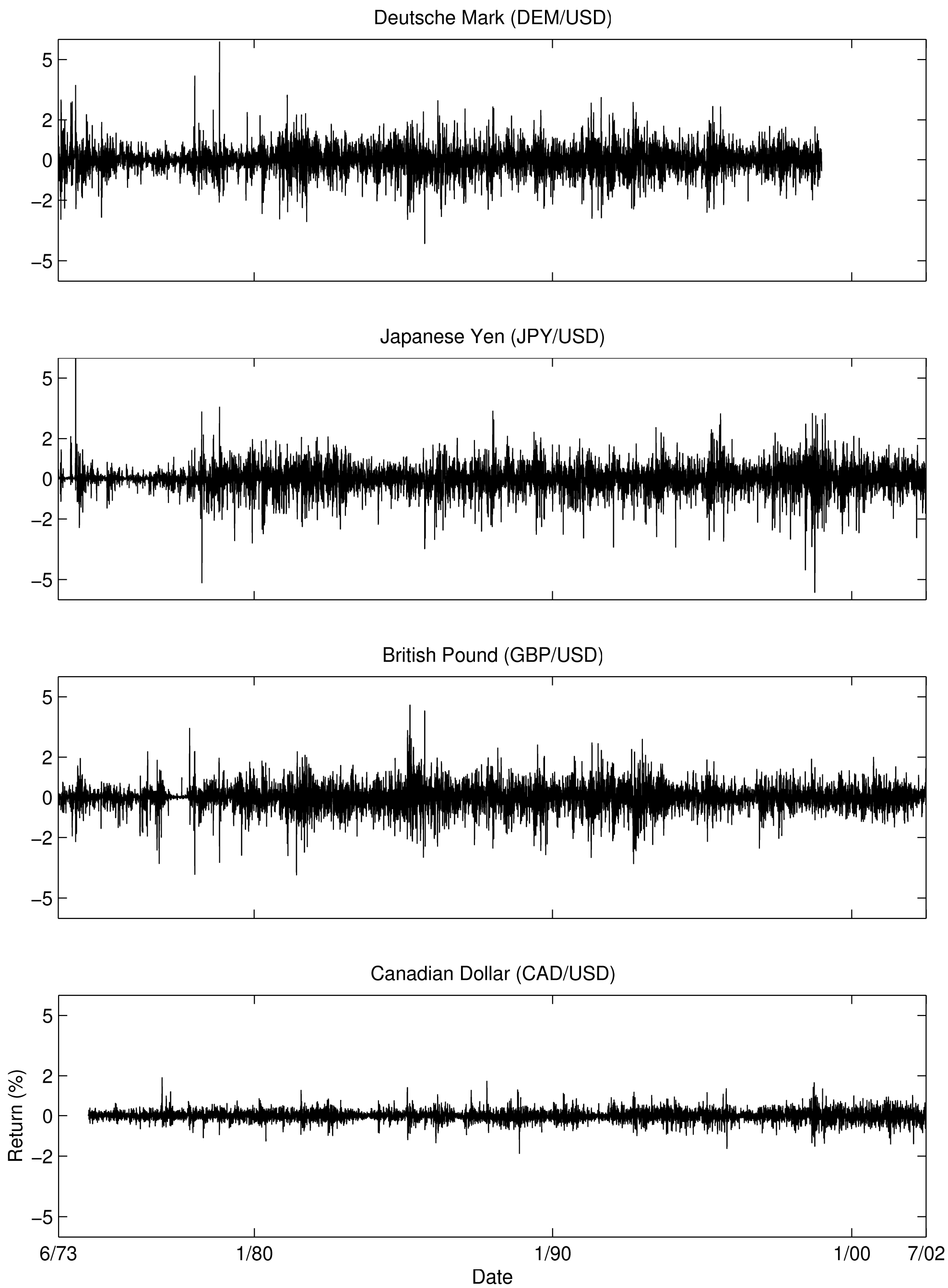

Figure 2: Exchange Rate Data. This figure shows the daily log price differences of the four exchange rate series.