# NONLINEAR FEATURES OF REALIZED FX VOLATILITY

## John M. Maheu and Thomas H. McCurdy*

*Abstract*—This paper investigates nonlinear features of FX volatility dynamics using estimates of daily volatility based on the sum of intraday squared returns. Measurement errors associated with using realized volatility to estimate ex post latent volatility imply that standard time series models of the conditional variance become variants of an ARMAX model. We explore nonlinear departures from these linear specifications using a doubly stochastic process under duration-dependent mixing. This process can capture large abrupt changes in the level of volatility, time-varying persistence, and time-varying variance of volatility. The results have implications for forecast precision, hedging, and pricing of derivatives.

## I. Introduction

VOLATILITY is a prevailing feature of financial markets, and its presence implies both risks and opportunities. Although asset returns are often represented as a martingale difference series, volatility displays persistence and should therefore be predictable. However, the data-generating process (DGP) for volatility is unobservable. Therefore, most studies have used a fully specified conditional mean and conditional variance model to estimate and analyze that latent volatility. What has been learned from past studies is dependent upon the maintained statistical model and estimation procedures.

An alternative would be to use an observable proxy for latent volatility. A standard proxy for ex post daily volatility has been daily squared returns or squared residuals from a regression model for returns. However, these measures of volatility can be very noisy.[1] This issue has been analyzed by Andersen and Bollerslev (1998) who evaluate GARCH(1,1) forecasts of latent volatility using the sum of high-frequency squared returns as a proxy for ex post daily volatility. Based on a simulation of integrated volatility implied by the GARCH(1,1) diffusion limit, they find that the sum of high-frequency squared returns is a less noisy measure of latent volatility than is squared daily returns.

[1] An alternative proxy for daily ex post volatility that is efficient and easy to calculate is the price range over a day. This has been used by Hsieh (1993) and Alizadeh, Brandt, and Diebold (2002), among others.

This superior proxy for ex post volatility facilitates a more accurate assessment of the GARCH forecasts.

A theoretical motivation for using the sum of high-frequency squared returns to compute measures of volatility at lower frequencies is provided by Merton (1980).[2] Andersen, Bollerslev, Diebold, and Labys (2001) extend the theoretical justification for this measure to the class of semi-martingales and, using a ten-year sample of five-minute returns for exchange rates, provide stylized facts associated with the distributions of volatility and correlation at daily (and lower) frequencies. They refer to the sum of intraday squared returns as realized volatility.

Rather than using realized volatility to gauge the potential performance of GARCH forecasts (Andersen & Bollerslev, 1998) or characterizing distributions of ex post volatility (Andersen, Bollerslev, Diebold, & Labys, 2001), our paper uses this nonparametric measure to analyze the time series behavior of FX volatility. Following the earlier literature, we use the sum of intraday squared returns (realized volatility) as an observable proxy for daily ex post volatility. However, we explicitly incorporate the fact that, for practical sampling frequencies, realized volatility will not be an error-free measure of ex post latent volatility. Although very few assumptions are necessary to ensure that the error is an innovation process, a model of latent volatility is required to fully characterize the distribution of that error. Specifying a functional relationship between our estimate of volatility and latent daily volatility allows the latter to be parameterized in terms of realized volatility, other variables in the information set, and an error term. In this framework, standard latent volatility models become variants of an ARMAX time series model.

In addition to proposing a structure that links time series models of volatility to their functional form in term of realized volatility, this paper tests for nonlinear departures from the linear (in parameters) ARMAX representation of realized volatility. To do this, we introduce a new model that permits regime switching within an ARMAX model and that does not require simulation methods for estimation. In particular, our semi-Markov ARMAX alternative allows for regime switching in the conditional mean and in the conditional variance of realized volatility. The transition probabilities governing the discrete unobserved state variables and the conditional mean of volatility are both functions of the duration of the state. This framework allows us to investigate the importance of time-varying parameters and time-varying persistence, in contrast to standard volatility models. In addition to capturing large abrupt changes in the

[2] Examples of this practice include Poterba and Summers (1986), French, Schwert, and Stambaugh (1987), Schwert (1989, 1990), Schwert and Seguin (1990), Hsieh (1991), and Taylor and Xu (1997).

level of realized volatility, our structure models time variation in the conditional variance of volatility.

The importance of nonlinear effects in volatility are gauged by in-sample statistical tests and by out-of-sample forecasts. These analyses show the semi-Markov model to be a good description of FX volatility dynamics. In particular, we find strong statistical evidence of regime changes in both the conditional mean and conditional variance of realized volatility. With respect to shifts in the mean, our results support the usefulness of recent research exploring the combination of volatility clustering models with nonlinearities such as structural changes or regime-switching dynamics.[3] For example, in keeping with the results of Lamoureux and Lastrapes (1990) and Kim and Kon (1999), capturing shifts in the mean of volatility lowers estimates of the persistence of shocks to volatility. However, in addition, we also find evidence of regime switches in the conditional variance of FX volatility. This new result highlights the issue of time variation in the volatility of volatility which is important for vega hedging and for pricing derivatives.

Switches between states in our model are associated with discrete changes in realized volatility. One of the states captures periods of high volatility in FX rates. Switches from the normal state to that state are characterized by an abrupt increase in the level of realized volatility combined with a large increase in the variance of realized volatility. The estimates indicate that this state is not initially persistent unless the high levels of volatility continue for several periods, in which case the state becomes persistent. Thus, extremely high levels of volatility are not expected to occur often, but periodically explosions in volatility may persist for some time. Note that this flexibility of our parameterization, which incorporates time-varying persistence, allows that state to capture observations in the tail of the unconditional distribution of volatility (for example, due to news announcements) as well as more-sustained episodes of high volatility.

Our evidence of nonlinearity in realized volatility has several important implications. Firstly, the nonlinear model provides superior out-of-sample forecasts. By identifying periods of high volatility, forecasts during normal periods can be held with more confidence. Secondly, modeling large abrupt changes in volatility will have important effects on derivative security pricing. For example, Jorion (1988) finds evidence of jumps in FX markets and shows that they explain some of the mispricing in currency options. The time-varying variance of volatility found in this paper is important for hedging and for pricing derivatives for which the underlying is a volatility measure, such as a volatility index.

This paper is organized as follows: section II reviews the theoretical and practical issues involved in measuring daily volatility. Data sources are detailed in section III. The form of GARCH and stochastic volatility models in terms of realized volatility is explained in section IV as well as the semi-Markov nonlinear model. Details on the model estimates, including out-of-sample evaluation, are found in section V. Section VI provides a discussion of the results, and section VII concludes.

## II. Volatility Measures

In this section, we introduce notation for instantaneous and integrated latent volatilities, for daily realized volatilities, and for potential errors associated with using realized volatility to measure integrated latent volatility. In particular, daily realized volatility is defined as the sum of intraday squared returns. The latter are fully observable and provide a nonparametric estimate of latent volatility over the same time interval. We briefly discuss the conditions for realized volatilities to be consistent and unbiased estimates of the latent volatility associated with possible data-generating processes (DGPs).

### A. Constructing Daily Volatility Measures

We label latent volatility as $\sigma^2$ and use $(t)$ to denote instantaneous variables and subscripts to denote discrete quantities. Let $\tilde{\sigma}_t^2$ be latent volatility for day $t$ which has been aggregated over the interval $(t, t + 1)$ where a full 24-hour day is represented by the time interval 1. To illustrate these concepts, consider the following simple diffusion process with no conditional mean dynamics,

$$dp(t) = \sigma(t)dW(t) \tag{1}$$

where $p(t)$ is the instantaneous logarithm of price,
$dW(t)$ is a standard Brownian process, and
$\sigma(t)$ is a stochastic process independent of $dW(t)$.

For this diffusion, latent volatility associated with day $t$ is the integral of the instantaneous variances over the day,

$$\tilde{\sigma}_t^2 = \int_t^{t+1} \sigma^2(\omega)d\omega. \tag{2}$$

This is an ex post measure of latent volatility associated with day $t$.

Merton (1980) showed that integrated volatility (2) can be approximated to an arbitrary precision using the sum of

[3] These approaches include Markov-switching volatility models such as Turner, Startz, and Nelson (1989) and Pagan and Schwert (1990), threshold ARCH models such as Gourieroux and Monfort (1992), and switching ARCH and stochastic volatility models such as Cai (1994), Hamilton and Susmel (1994), Gray (1996), Klaassen (1998) and So, Lam, and Li (1998).

the intraday squared returns. In our case, these nonparametric estimates of $\tilde{\sigma}_t^2$ are labelled $\hat{\sigma}_t^2$ and are computed as[4]

$$\hat{\sigma}_t^2 = \sum_{j=0}^{\delta-1} r_{t+j/\delta}^2, \tag{3}$$

where $r_{t+j/\delta} = p_{t+j/\delta} - p_{t+(j-1)/\delta}$ defines continuously compounded returns, sampled $\delta$ times per day. Note that the subscript $t$ indexes the day while $j$ indexes the time within day $t$ (that is, $\{(t, j) | t = 1, \ldots, T; j = 0, \ldots, \delta - 1\}$). For example, in our empirical application, the interval is one day and the price process is sampled every five minutes in order to compute five-minute returns. This implies that $\delta = 288$. Andersen, Bollerslev, Diebold, and Labys (2001) refer to $\hat{\sigma}_t^2$ as daily realized volatility.

The quadratic variation of the diffusion process implies that equation (3) will provide a consistent estimate of latent volatility over day $t$ because

$$\text{plim}_{\delta \to \infty} \sum_{j=0}^{\delta-1} r_{t+j/\delta}^2 = \tilde{\sigma}_t^2. \tag{4}$$

As the sampling frequency from a diffusion is increased, the sum of squared returns converges to the integrated volatility over the fixed time interval. Andersen, Bollerslev, Diebold, and Labys (2001) generalized this result to the class of special (finite mean) semi-martingales. This class encompasses processes used in standard arbitrage-free, asset-pricing applications such as Ito diffusions, jump processes, and mixed jump diffusions.

This approach could also be applied to the case of a discrete high-frequency price process. For example, consider a class of discrete-time DGPs for intraday FX returns,

$$r_{t+j/N} = \sigma_{t+j/N} z_{t+j/N}, \quad z_{t+j/N} \sim IID(0, 1) \tag{5}$$

where $r_{t+j/N} = p_{t+j/N} - p_{t+(j-1)/N}$,
$p_t$ is the log price, and
$N$ is the total number of price quotes in a day.[5]

Specification (5) encompasses discrete-time stochastic volatility as well as GARCH processes.

### B. Measurement and Estimation Error

In this paper, we explicitly account for various sources of measurement and estimation error that can arise from using realized volatility as a measure for latent volatility over a

day. Estimation error in this context refers to discretization error and/or possible model misspecification.[6] On the other hand, measurement error can result from institutional and behavioral features of the trading process (for example, market microstructure effects such as the bid-ask bounce) and interpolation of prices to an equally spaced grid. This subsection elaborates on these issues and discusses the properties of the realized volatility estimator in the presence of these errors.

One source of error could arise from discretization ($\delta$ is finite) because the true stochastic volatility process may generate quotes at frequencies higher than our five-minute sampling frequency. Secondly, actual price quotes are irregularly spaced in time. Interpolation of observed prices to an equally spaced (five-minute) grid will introduce measurement error with respect to the true price process. Thirdly, the price quotes may be influenced by market microstructure dynamics that can induce autocorrelation in high-frequency FX returns. Bai, Russell, and Tiao (1999) show that autocorrelation and nonnormal innovations can have important effects on the accuracy of the estimator.

To deal with such sources of error, we define $\epsilon_t$ as the difference between $\tilde{\sigma}_t^2$ (the latent volatility for day $t$) and its estimate $\hat{\sigma}_t^2$ (daily realized volatility). That is,

$$\hat{\sigma}_t^2 = \tilde{\sigma}_t^2 + \epsilon_t. \tag{6}$$

Equation (6) makes it clear that we are not treating realized volatility as an error-free measure of latent volatility because $\epsilon_t$ is composed of measurement errors as well as estimation error.

As noted, the nonparametric estimator $\hat{\sigma}_t^2$, given by equation (3), will provide consistent estimates of latent volatility for a broad class of DGPs. However, the accuracy of daily realized volatilities as measures of ex post daily latent volatilities will depend on the true DGP. For example, in the presence of market microstructure effects, we would expect realized volatility constructed as in equation (3) to be a biased estimator for latent volatility ($\epsilon_t$ is not mean zero). In this situation, one must either account for the market microstructure in the construction of the estimator or, alternatively, filter the returns such that they conform to the conditions necessary to ensure that equation (3) provides unbiased estimates of latent volatility. Therefore, in this paper, we filter the five-minute returns to remove high-frequency sources of bias. (See section III.) In this case, when the diffusion in equation (1) is driftless, $\hat{\sigma}_t^2$ will be an unbiased estimator for $\tilde{\sigma}_t^2$, and $\epsilon_t$ will be a martingale difference sequence ($E_{t-1}\epsilon_t = 0$).[7] Although unbiasedness will not hold in the presence of a drift in the underlying DGP, $\hat{\sigma}_t^2$ will still be a consistent estimator of $\tilde{\sigma}_t^2$.

---

[4] Andreou and Ghysels (2001) evaluate the impact of window length, data frequency, weighting scheme, and estimation methods (block versus rolling sample) for the relative efficiency of estimates of integrated volatility associated with diffusions.

[5] The ex post latent volatility of returns for day $t$ associated with this discrete-quote DGP is, $\tilde{\sigma}_t^2 = \sum_{j=0}^{N-1} \sigma_{t+j/N}^2$, and an estimate of $\tilde{\sigma}_t^2$ is equation (3) where $\delta = N$.

[6] Model misspecification would result in equation (3) being an inconsistent estimator.

[7] See Barndorff-Nielsen and Shephard (2002). Additional assumptions on the underlying DGP may allow one to further characterize the distribution of $\epsilon_t$.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Unfiltered | $-0.033\ (0.316\text{e-}02)$ | $-0.016\ (0.250\text{e-}02)$ | $-0.024\ (0.224\text{e-}02)$ | $-0.013\ (0.226\text{e-}02)$ | $-0.357\text{e-}02\ (0.223\text{e-}02)$ |
| Filtered | $0.373\text{e-}04\ (0.322\text{e-}02)$ | $0.248\text{e-}04\ (0.254\text{e-}02)$ | $-0.667\text{e-}04\ (0.224\text{e-}02)$ | $-0.126\text{e-}03\ (0.226\text{e-}02)$ | $-0.322\text{e-}02\ (0.224\text{e-}02)$ |

This table reports the first five autocorrelations of the five-minute returns data before and after a MA(4) filter was applied. Estimates are from GMM and heteroskedastic and autocorrelation consistent standard errors appear in parenthesis.

For example, consider one candidate DGP from equation (5) for which $\epsilon_t$ takes the form

$$\epsilon_t = \hat{\sigma}_t^2 - \tilde{\sigma}_t^2 = \sum_{j=0}^{N-1} \sigma_{t+j/N}^2 (z_{t+j/N}^2 - 1). \qquad (7)$$

Under the null of that DGP, $z_{t+j/N} \sim IID(0, 1)$ and is independent of $\sigma_{t+j/N}^2$. Therefore, although $\epsilon_t$ is heteroskedastic, it is mean zero and $\hat{\sigma}_t^2$ will be an unbiased estimator for $\tilde{\sigma}_t^2$ for this DGP.

Equation (6) links our volatility estimate $\hat{\sigma}_t^2$ to the contemporaneous measure of daily latent volatility $\tilde{\sigma}_t^2$. However, these measures of volatility are not always synonymous with conditional variance. One example is when the conditional variance is defined to be measurable with respect to a prior information set. An example of the latter is the popular GARCH specification of conditional variance. To illustrate this case, we define $\tilde{\sigma}_{t|t-1}^2$ as the GARCH conditional variance that is measurable with respect to information at $t - 1$, consisting of past daily returns, so that

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_{t|t-1}^2 + w_t \qquad (8)$$

in which $w_t$ captures the difference between conditioning information associated with $\tilde{\sigma}_{t|t-1}^2$ as opposed to $\tilde{\sigma}_t^2$. Substituting equation (8) into (6) gives us

$$\hat{\sigma}_t^2 = \tilde{\sigma}_{t|t-1}^2 + w_t + \epsilon_t = \tilde{\sigma}_{t|t-1}^2 + \epsilon_t^*, \qquad (9)$$

where $\epsilon_t^*$ incorporates contemporaneous measurement and estimation error as well as conditioning information innovations.

### III. Data Sources and Issues

Ten years of FX prices recorded every five minutes were obtained from Olsen and Associates. The data are constructed from logarithmic middle prices (averages of log bid and ask quotes) for the Deutsche mark/U.S. dollar (DM/USD) exchange rate from the interbank FX market for the period Monday, December 1, 1986, 00:00:00 GMT to Sunday, December 1, 1996, 00:00:00 GMT. These prices are interpolated to a five-minute grid using the nearest quotes that bracket the grid point. The data set also includes a data-hole indicator variable designating whether the quotes used for interpolation are more than five minutes away on both sides of the grid point of the constructed quote. If this is the case, the indicator variable takes the value 1, otherwise it is 0. Days with runs of such indicators equal to 1

indicate data gaps due to a problem with the recording or very slow activity in the FX market.

Currencies trade 24 hours per day, seven days a week. This results in 1,052,064 five-minute prices for the ten-year sample in the Olsen and Associates data set. Using these prices, we computed 1,052,063 continuously compounded five-minute returns (expressed in USD terms) and matched those returns with time, date, and day-of-the-week and holiday indicator variables.

To construct our measure of realized volatility at a daily frequency, we need to define the boundaries of a day. Following Bollerslev and Domowitz (1993), we define day $t$ as the interval from 21:05 GMT the previous day to 21:00 GMT on day $t$. Based on the motivation summarized in section II, the estimate for daily realized volatility is computed by summing squares of the currency's five-minute returns over the day, that is, using equation (3) with $\delta = 288$.

Although it is possible to trade currencies at any time, there will clearly be periods such as weekends (covering the period from 21:05:00 GMT Friday to 21:00:00 GMT Sunday) and holidays, during which there will be very low trading activity. Such periods will generate seasonal effects. Following Andersen, Bollerslev, Diebold, and Labys (2001), we eliminated slow trading days (December 24–26 and 31, January 1–2) associated with fixed holidays, as well as those for the moving holidays Good Friday, Easter Monday, Memorial Day, July Fourth, Labor Day, and Thanksgiving and the day after. Secondly, we calculated the sum of the indicator variable for data "holes" over a day and removed days in which the sum was greater than 144 (more than one-half of the day). In addition to weekends and the holidays already mentioned, this resulted in the removal of an extra eighteen days, almost all of which had runs greater than 120 for the indicator value of 1. Finally, we eliminated all of December 1986 due to frequent holidays and runs of the indicator variable. In sum, deseasonalizing by removing days with abnormally low trading activity resulted in 709,920 five-minute returns (2,465 days) over the period from January 6, 1987 to November 27, 1996.[8]

As reported in table 1, there is autocorrelation in the raw five-minute returns (up to four lags), possibly due to the discrete nature of bid/ask quotes and associated market microstructure dynamics. To conform to the high-frequency DGPs in equation (1) or (5), and to ensure that $\epsilon_t$ defined in

[8] Andersen, Bollerslev, Diebold, and Labys (2001) used 2,445 days over their sample from December 1, 1986 to December 1, 1996.

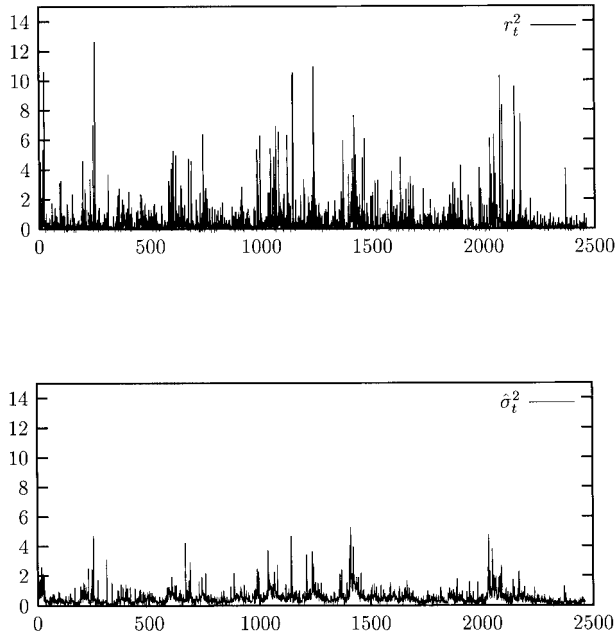FIGURE 1.—SQUARED RETURNS AND REALIZED VOLATILITY, DM/USD
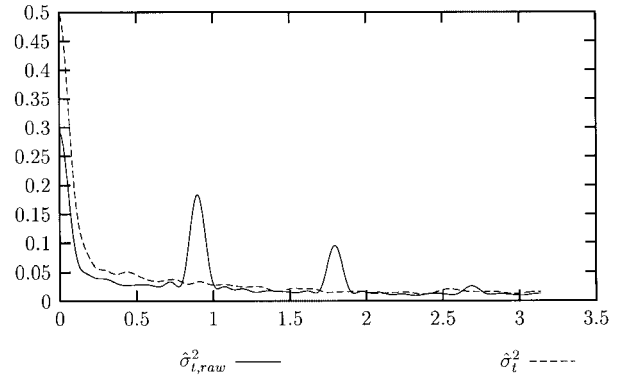


FIGURE 3.—DM/USD SPECTRUM ESTIMATES



Figure 1 displays time series plots of $\hat{\sigma}_t^2$ and daily squared returns, $r_t^2$. This figure shows realized volatility to be a less variable series than squared daily returns. Table 2 also shows that daily realized volatility computed from intra-daily squared returns is less noisy than traditional proxies for ex post daily volatility. In particular, the maximum of $\hat{\sigma}_t^2$ is 5.222 as compared to 12.658 for $r_t^2$, and the standard deviation of $\hat{\sigma}_t^2$ is 0.487 as compared to 1.010 for $r_t^2$.

## IV.    Time-Series Models of Realized Volatility

In section II, we discussed a nonparametric measure of ex post volatility that estimates latent volatility over an interval such as a day. Using equations (6) and (9), we show how popular daily volatility models can be represented in terms of observable realized volatility.

### A.    Linear Time-Series Models

Given the decomposition of realized volatility into daily latent volatility plus error, as specified in equation (6), one can derive the functional form for the time series pattern of realized volatility implied by a particular DGP for latent volatility. For example, consider an autoregressive (AR)

equation (6) is a martingale difference sequence, we filtered the raw five-minute returns data with an MA(4) including a constant.[9] This filter removes the autocorrelations (see filtered data in table 1) and potential trends in the data.

Let $\hat{\sigma}_{t,raw}^2$ denote realized volatility when all days are used and $\hat{\sigma}_t^2$ denote realized volatility for deseasonalized data (abnormal trading days eliminated). Estimates of the autocorrelation, spectrum, and empirical distribution for the two data series ($\hat{\sigma}_{t,raw}^2$ and $\hat{\sigma}_t^2$) of the DM/USD are shown in figures 2 through 4. The spikes in the spectrum plot correspond to the seasonal frequencies of one week and a half week.

---

[9] Andersen, Bollerslev, Diebold, and Ebens (2001) also use a linear filter to remove serial correlation in high-frequency equity returns. Some alternative approaches, including an estimator used by French et al. (1987) that corrects for first-order autocorrelation, are discussed by Bai et al. (1999).

FIGURE 2.—DM/USD SAMPLE AUTOCORRELATIONS
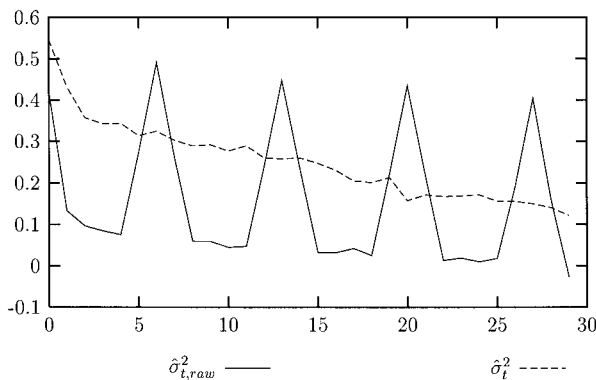


FIGURE 4.—DM/USD EMPIRICAL DISTRIBUTION FUNCTION
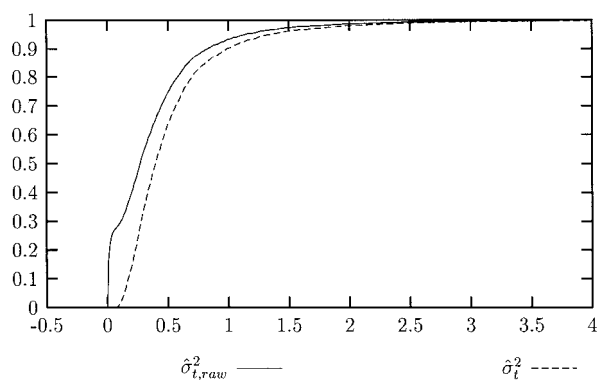
|  | $\hat{\sigma}^2_{t,raw}$ | $\hat{\sigma}^2_t$ | $r_t$ | $r_t^2$ |
|---|---|---|---|---|
| Mean | 0.374 | 0.528 | 0.009 | 0.488 |
| Stdev | 0.464 | 0.487 | 0.699 | 1.010 |
| Skewness | 3.520 | 3.721 | 0.010 | 5.391 |
| Kurtosis | 23.890 | 24.085 | 5.288 | 43.733 |
| Min | 0.000 | 0.051 | −3.558 | 0.000 |
| Max | 5.245 | 5.222 | 3.254 | 12.658 |

discrete-time SV specification for the dynamics of daily latent volatility,[10] that is,

$$\tilde{\sigma}^2_t = \mu + \beta(\tilde{\sigma}^2_{t-1} - \mu) + v_t, \tag{10}$$

where $v_t$ is stochastic and not measurable with respect to an information set consisting of daily returns.

One might be tempted to replace daily latent volatility $\tilde{\sigma}^2_t$ with its nonparametric estimator $\hat{\sigma}^2_t$ and fit equation (10) using least squares. However, realized volatility measures latent volatility with error. Ignoring potential measurement and estimation error may cause an error-in-variables problem. For this reason, we decompose realized volatility into daily latent volatility plus error, as discussed in section II and specified in equation (6). Using equation (6) to replace $\tilde{\sigma}^2_t$ in equation (10) implies that

$$\hat{\sigma}^2_t = \mu + \beta(\hat{\sigma}^2_{t-1} - \mu) - \beta\epsilon_{t-1} + \epsilon_t + v_t. \tag{11}$$

That is, the SV specification implies an (approximate) ARMA(1,1) when expressed in terms of realized volatility $\hat{\sigma}^2_t$, and sources of error.

As emphasized by Andersen, Bollerslev, Diebold, and Labys (2001), volatility dynamics may involve auxiliary state variables. In this case, optimal forecasts could utilize available information in addition to past realized volatility. One candidate conditioning variable would be the most recent daily information about volatility, $r_{t-1}^2$. For example, consider a GARCH(1,1) function specified as[11]

$$\tilde{\sigma}^2_{t|t-1} = \mu + \beta(\tilde{\sigma}^2_{t-1|t-2} - \mu) + \alpha r_{t-1}^2, \tag{12}$$

in which the notation $\tilde{\sigma}^2_{t|t-1}$ is used to emphasize that in the GARCH parameterization the conditional variance is measurable with respect to an information set of past daily returns. In this case, using equation (9), we have

$$\hat{\sigma}^2_t = \mu + \beta(\hat{\sigma}^2_{t-1} - \mu) + \alpha r_{t-1}^2 - \beta\epsilon^*_{t-1} + \epsilon^*_t, \tag{13}$$

which is an ARMA(1,1) for $\hat{\sigma}^2_t$, our measure of realized volatility, with an added conditioning variable, $r_{t-1}^2$, from the agents' information set of returns. We refer to this time series specification for daily realized volatility as an ARMAX(1,1) model.

The ARMAX structure implied by the popular SV and GARCH(1,1) models of latent volatility provides us with a benchmark to gauge the importance of the proposed nonlinear features discussed in the next subsection. The simple SV and GARCH(1,1) parameterizations (10) and (12) have several well-known shortcomings in characterizing time series properties of volatility. In our context, improvements might alternatively be gained by retaining the linear functions but allowing for more-flexible innovation distributions, longer lags in the ARMAX, or long memory.

### B. A Nonlinear Model

The SV and GARCH(1,1) specifications are both linear in the parameters conditional on realized volatility and past returns. A vast literature has developed from the ARCH model of Engle (1982). Part of this literature investigates whether nonlinear effects are important in the conditional variance function.

In this section, we discuss a nonlinear alternative to the linear models. Our model is a discrete mixture of distributions, which can approximate arbitrary densities (Gottschling, Haefke, & White, 1999). A discrete mixture of distributions is capable of capturing the skewness and kurtosis that we observe in the distribution of volatility (table 2). Within this class of models, we consider a parsimonious hidden Markov model directed by duration-dependent mixing.

The nonlinear specification applied to realized volatility is a discrete-state semi-Markov[12] model. In this model, the duration of a state can affect the conditional moments as well as the transition matrix.[13] Our model of realized volatility builds on the results in Maheu and McCurdy (2000b), which investigate a model that captures volatility dynamics exclusively through discrete changes in the level of volatility.[14] Motivated by these findings and the availability of observable volatility in the form of realized volatility, we consider a combination of smooth persistent changes in volatility through an ARMA specification as well as discrete jumps captured by regime switching. This combination of dynamics provides a flexible model that can capture important features in the conditional and unconditional distribu-

---

[10] Meddahi and Renault (2000) provide an AR specification of SV that is closed under temporal aggregation. See also Barndorff-Nielsen and Shephard (2001).

[11] Such functions have stochastic volatility diffusion limits. See, for example, Nelson (1990) and Duan (1997). For discrete-time DGPs, Drost and Nijman (1993) show that weak-GARCH models are closed under temporal aggregation.

[12] In semi-Markov processes, the times at which transitions occur are governed by a separate probability distribution. (See Howard (1971).) Our duration-dependent Markov-switching specification is a variant of a semi-Markov process for which virtual transitions (a move from a state to itself) and actual transitions (a move from state $i$ to state $j$) are directed by a duration function. The hazard function defined by the duration parameterization implicitly characterizes the holding time (and waiting time) distributions.

[13] Other Markov-switching applications that built on Hamilton (1989) to address issues of duration include Durland and McCurdy (1994), Lam (1997), Filardo and Gordon (1998), and Maheu and McCurdy (2000a,b). A Bayesian approach to volatility modeling with duration-dependent mixing is presented by Kim and Nelson (1998).

[14] Maheu and McCurdy (2000b) show that a duration-dependent, Markov-switching (DDMS) model provides a better description of the dynamics of the conditional and unconditional distribution of FX returns compared to GARCH and Markov-switching ARCH models. Their results suggest that discrete changes in volatility are an important ingredient in the time series dynamics of FX returns.

tions of realized volatility. It can also capture nonnormal innovation distributions.

Begin by assuming that realized volatility is governed by a discrete mixture of distributions. The discrete mixing variable $S_t \in \{1, 2\}$ is unobserved and depends only on $S_{t-1}$ and its duration $D_{t-1}$. Duration is defined as the number of consecutive occurrences of a particular state. We record duration up to and including $\tau$ periods in the past. Define the duration of state $S_t$ as

$$D_t = \min (D_{t-1}I(S_t, S_{t-1}) + 1, \tau). \tag{14}$$

$I(S_t, S_{t-1}) = 1$ if $S_t = S_{t-1}$ and otherwise 0. Durations above $\tau$ are recorded as having a duration of $\tau$.[15] Therefore, if duration happens to be greater than the memory of the process (which would be very unlikely), the transition probabilities or any function of $D_t$ would be constant after that point.

The transition probabilities are parameterized using the logistic function so that

$$P_{ii}(D_{t-1}) = P (S_t = i | S_{t-1} = i, D_{t-1} = d_{t-1})$$
$$= \frac{\exp(\gamma_1(i) + \gamma_2(i)d_{t-1})}{1 + \exp(\gamma_1(i) + \gamma_2(i)d_{t-1})} \quad i = 1, 2 \tag{15}$$

where $\gamma_1(1)$, $\gamma_2(1)$ and $\gamma_1(2)$, $\gamma_2(2)$ are parameters associated with states 1 and 2, respectively. $P_{ii}(D_{t-1})$ is the conditional probability of staying in state $i$ given that we have been in state $i$ for $d_{t-1}$ periods where $i = 1, 2$. The probability of switching from state $i$ to state $j$, given that we have been in state $i$ for $d_{t-1}$ periods, is $P_{ij}(D_{t-1}) = 1 - P_{ii}(D_{t-1})$, $i \neq j$, $j = 1, 2$.

The semi-Markov model of realized volatility (SM-ARMAX(1,1)) extends the linear ARMAX specification with regime switching in the conditional mean and the conditional variance as

$$\hat{\sigma}_t^2 = \mu(S_t, D_t) + \beta(\hat{\sigma}_{t-1}^2 - \mu(S_{t-1}, D_{t-1}))$$
$$+ \alpha r_{t-1}^2 + \theta \tilde{\eta}_{t-1} + \eta_t(S_t)$$
$$\mu(S_t, D_t) = \phi(S_t) + \psi(S_t)D_t \tag{16}$$
$$\tilde{\eta}_{t-1} = \hat{\sigma}_{t-1}^2 - E_{t-2}\hat{\sigma}_{t-1}^2$$
$$\eta_t(S_t) \sim N(0, \lambda^2(S_t)).$$

In general, an ARMA(1,1) with regime switching is intractable due to a state space that increases with the sample size.[16] Our ARMA model augmented with regime switching is an approximation to a true latent-variable, regime-switching ARMA model. The approximation is due to the fact that $\tilde{\eta}_{t-1}$ is not a function of the unobserved state variables but rather represents the error observed by the econometrician.[17] By integrating out the dependence of $\eta_{t-1}$ on $S_{t-1}$ and $D_{t-1}$, $\tilde{\eta}_{t-1}$ represents an estimate of the true (unobserved) innovation $\eta_{t-1}(S_{t-1})$.

The model in equation (16) nests the ARMAX(1,1), which was motivated by a GARCH(1,1) formulation of latent volatility, and extends it to allow for state-dependent variance and regime switching in the conditional mean of realized volatility.[18] The conditional mean term $\mu(S_t, D_t)$ is a function of $S_t$ and its duration $D_t$ through the coefficients $\phi(S_t)$ and $\psi(S_t)$, respectively.

The earlier linear ARMAX(1,1) specification is a parsimonious model that imposes strong assumptions on volatility. For example, the unconditional variance and persistence of volatility levels are both assumed to be constant. Furthermore, that parameterization is not well suited to explaining or predicting discrete jumps in volatility levels. However, the SM-ARMAX(1,1) is a nonlinear extension that allows the levels of volatility to change continuously and discretely over time, provides a flexible structure governing persistence of volatility levels through the transition probabilities and duration dependence, and allows for conditional heteroskedasticity in realized volatility.

To discuss duration effects on realized volatility, it is useful to define the hazard function. The hazard function is the conditional probability of a state change given the state has achieved a duration $d$. In terms of the transition probabilities, the hazard function is

$$P_{ij}(D_{t-1}) = \frac{1}{1 + \exp(\gamma_1(i) + \gamma_2(i)d_{t-1})}, \quad i = 1, 2. \tag{17}$$

A decreasing hazard function is referred to as negative duration dependence whereas 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 FX market persists in state 2, then the probability of staying in state 2 increases.

---

[15] Although it is possible to estimate $\tau$ (for example, using a grid search), we set $\tau = 25$, which is large enough to ensure that all the duration effects have been captured. In other words, the likelihood is flat with respect to values of $\tau$ greater than the chosen value. See Maheu and McCurdy (2000a) for these issues.

[16] Billio, Monfort, and Robert (1999) use Bayesian simulation methods to deal with the exploding state space in switching ARMA models.

[17] The accuracy of the approximation depends on the forecast precision for future states. To illustrate, consider a simple two-state MS model $y_t = \mu(S_t) + e_t$, and let $p_{1,t-1}$ be the inferred probability of state 1 given time $t - 1$ information. Then, if $S_t = 1$, $\tilde{e}_t = y_t - E_{t-1}y_t = (\mu(1) - \mu(2))(1 - p_{1,t-1}) + e_t$. Therefore, the accuracy of $\tilde{e}_t$ in measuring $e_t$ depends on how close the states are ($|\mu(1) - \mu(2)|$) and/or how close $p_{1,t-1}$ is to 1. An analogous example applies if $S_t = 2$. Therefore, models that predict well will provide good estimates of the underlying innovations; that is, $\tilde{e}_t$ will be close to $e_t$. In our more general case, $\tilde{\eta}_{t-1}$ in equation (16) is an innovation with respect to the information set $\{\hat{\sigma}_{t-2}^2, \ldots, \hat{\sigma}_1^2\}$ of the econometrician, not with respect to an information set that includes $S_{t-1}$.

[18] This is a slight abuse in terminology because, in order to nest the earlier linear specifications, the conditional mean of realized volatility is also a function of $\alpha r_{t-1}^2$ and the MA(1) term.

TABLE 3.—VOLATILITY MODEL ESTIMATES, DM/USD

| | ARMA(1,1) | ARMA(1,1)-$r_{t-1}^2$ | SM-ARMA(1,1)-$r_{t-1}^2$ | |
| | | | $S_t = 1$ | $S_t = 2$ |
|---|---|---|---|---|
| $\mu$ | 0.553 (0.038) | 0.400 (0.041) | | |
| $\phi(S_t)$ | | | 0.299 (0.036) | 0.560 (0.052) |
| $\psi(S_t)$ | | | $-0.004$ (0.002) | 0.045 (0.020) |
| $\beta$ | 0.902 (0.021) | 0.861 (0.024) | 0.757 (0.027) | |
| $\alpha$ | | 0.041 (0.007) | 0.058 (0.007) | |
| $\theta$ | $-0.615$ (0.043) | $-0.613$ (0.042) | $-0.425$ (0.035) | |
| $\lambda(S_t)$ | 0.410 (0.007) | 0.408 (0.007) | 0.132 (0.005) | 0.797 (0.039) |
| $\gamma_1(S_t)$ | | | 1.744 (0.178) | $-0.502$ (0.208) |
| $\gamma_2(S_t)$ | | | 0.035 (0.016) | 0.181 (0.053) |
| $\bar{Q}(10)$ | 15.131 [0.127] | 11.251 [0.338] | 14.092 [0.169] | |
| $Q^2(10)$ | 52.772 [0.000] | 53.916 [0.000] | 6.524 [0.769] | |
| NL | 54.600 [0.000] | 48.077 [0.000] | | |
| lgl | $-948.282$ | $-938.770$ | 10.279 | |

Standard errors appear in parenthesis and $p$-values in square brackets. The maintained model for the ARMA(1,1), and ARMA(1,1)-$r_{t-1}^2$ is

$$\hat{\sigma}_t^2 = \mu + \beta(\hat{\sigma}_{t-1}^2 - \mu) + \alpha r_{t-1}^2 + \theta\epsilon_{t-1} + \epsilon_t, \ \epsilon_t \sim iid(0, \lambda^2).$$

The SM-ARMA(1,1)-$r_{t-1}^2$ is

$$\hat{\sigma}_t^2 = \mu(S_t, D_t) + \beta(\hat{\sigma}_{t-1}^2 - \mu(S_{t-1}, D_{t-1})) + \alpha r_{t-1}^2 + \theta\tilde{\eta}_{t-1} + \eta_t(S_t)$$

$$\mu(S_t, D_t) = \phi(S_t) + \psi(S_t) D_t$$

$$\tilde{\eta}_{t-1} = \hat{\sigma}_{t-1}^2 - E_{t-2}\hat{\sigma}_{t-1}^2, \ \eta(S_t) \sim N(0, \lambda^2(S_t)),$$

with state transition probabilities from equation (15). $\bar{Q}(10)$ is a heteroskedastic robust version of the Ljung-Box statistic for serial correlation in the residuals with ten lags, while $Q^2(10)$ is the standard Ljung-Box statistic on the squared residuals. NL is a test for neglected nonlinearity in the residuals based on a heteroskedastic-robust Wald test of $\pi_2 = 0$, in

$$\epsilon_t = z_t\pi_1 + w_t\pi_2 + error$$

$$z_t = [1, \epsilon_{t-1}, \dots, \epsilon_{t-p}], \ w_t = [\hat{\epsilon}_{t-1}^2, \epsilon_{t-1}\epsilon_{t-2}, \dots, \epsilon_{t-p}^2]$$

which is a $0.5p(p + 1)$ degree-of-freedom test. The NL test sets $p = 4$. Data observations are from 1 to 1,800.

Hamilton (1989) shows that 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 for the semi-Markov model. Define $f(\cdot|\cdot)$ as the conditional density of the normal distribution, and $\Phi_t$ as the time $t$ information set available to the econometrician. The filter provides optimal inference for the unobserved variables given time $t$ information. For $S_t = 1, 2$ and $1 \leq D_t \leq \tau$ and using $s_t$ and $d_t$ to denote realized values of $S_t$ and $D_t$, we have

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

where

$$P(S_t = s_t, D_t = d_t|\Phi_{t-1})$$
$$= \sum_{s_{t-1}, d_{t-1}} P(S_t = s_t, D_t = d_t|S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1})$$
$$\times P(S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1}|\Phi_{t-1})$$

and

$$P(y_t|\Phi_{t-1}) = \sum_{s_t, s_{t-1}, d_{t-1}} f(y_t|S_t = s_t, D_t = d_t, \Phi_{t-1})$$
$$\times P(S_t = s_t, D_t = d_t|S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1})$$
$$\times P(S_{t-1} = s_{t-1}, D_{t-1} = d_{t-1}|\Phi_{t-1}).$$

The filter plays an important role in forecasts of current as well as future levels of realized volatility. One-step-ahead forecasts from this model are[19]

$$E_t\hat{\sigma}_{t+1}^2$$
$$= \sum_{s_{t+1}, s_t, d_{t+1}, d_t} E_t[\hat{\sigma}_{t+1}^2|S_{t+1} = s_{t+1}, S_t = s_t, D_{t+1}$$
$$= d_{t+1}, D_t = d_t] \times P(S_{t+1} = s_{t+1}, S_t = s_t, D_{t+1} \quad (18)$$
$$= d_{t+1}, D_t = d_t|\Phi_t).$$

## V. Estimates

Table 3 presents unrestricted ARMA(1,1), ARMAX(1,1), and SM-ARMAX(1,1) estimates for the in-sample observations 1–1800.[20] Recall that these correspond to SV, GARCH(1,1) and semi-Markov volatility models, respectively, expressed in terms of realized volatility. A Ljung-Box (LB) statistic robust to heteroskedasticity is reported for the

[19] Because the regime switching in the conditional mean affects only the intercept and the memory of duration is truncated at $\tau$, multiperiod forecasts of $\hat{\sigma}_{t+i}^2$ are a straightforward generalization of the one-period forecast. However, an assumed data-generating process for the high-frequency returns is needed to forecast daily squared returns that appear in equation (16). One possibility is to use equation (5) to derive the daily squared return to show $E_t r_{t+i}^2 = E_t(\sum_{j=0}^{N-1} r_{t+i+j/N})^2 = E_t\hat{\sigma}_{t+i}^2 = E_t\hat{\sigma}_{t+i}^2$, which ensures a closed-form solution for multiperiod forecasts of the SM-ARMAX model.

[20] Observations 1801–2465 are saved for out-of-sample evaluation of the models.

TABLE 4.—OUT-OF-SAMPLE FORECAST RESULTS, DM/USD

| | ARMA(1,1) | ARMA(1,1)-$r_{t-1}^2$ | SM-ARMA(1,1)-$r_{t-1}^2$ |
|---|---|---|---|
| $c_0$ | −0.069 (0.046) | −0.074 (0.044) | −0.050 (0.042) |
| $c_1$ | 1.094 (0.110) | 1.112 (0.107) | 1.156 (0.109) |
| $R^2$ | 0.417 | 0.446 | 0.459 |
| $PL$ | −0.237 | −0.224 | −0.131 |
| $\tilde{R}^2$ | 0.039 | 0.039 | 0.041 |
| $\widetilde{PL}$ | −1.828 | −1.816 | −1.722 |

The test regression is

$$\hat{\sigma}_t^2 = c_0 + c_1 E_{t-1}\hat{\sigma}_t^2 + \text{error},$$

where $E_{t-1}\hat{\sigma}_t^2$ is the particular model-based one-step-ahead forecast. Model estimates are obtained using observation 1–1800, and thereafter fixed. Out-of-sample observations are from 1801–2465. Heteroskedastic consistent standard errors appear in parenthesis. The proportional loss function is

$$PL = \frac{1}{T}\sum_{t=1}^{T} \log\left(\frac{\hat{\sigma}_t^2}{E_{t-1}\hat{\sigma}_t^2}\right).$$

$\tilde{R}^2$ is the $R^2$ from the test regression using daily squared returns

$$r_t^2 = c_0 + c_1 E_{t-1}\hat{\sigma}_t^2 + \text{error},$$

and

$$\widetilde{PL} = \frac{1}{T}\sum_{t=1}^{T} \log\left(\frac{r_t^2}{E_{t-1}\hat{\sigma}_t^2}\right).$$

residuals and a standard LB test for the squared residuals. Results from a heteroskedastic-robust test for neglected nonlinearity (NL) in the residuals from the linear models are also reported in the table.[21]

On the basis of the likelihood, the ARMAX(1,1) specification for realized volatility suggests a marginal improvement in explaining realized volatility compared to the ARMA(1,1) specification. In an ideal setting, we would like to use our ARMAX estimates to back out parameter estimates of the SV and GARCH models. However, this is not directly possible from table 3 estimates for two reasons. First, the SV specification of realized volatility in equation (11) does not imply an exact ARMA model, and secondly the GARCH parameterization in equation (13) puts restrictions on the ARMAX parameters ($\beta = -\theta$). For example, estimates from an ARMAX(1,1) with the restriction, $\beta = -\theta$ imposed gave $\hat{\alpha} = 0.103$ and $\hat{\beta} = 0.831$.[22] However, the unrestricted ARMAX estimates reported in table 3 represent a more flexible parameterization of the time series dynamics of realized volatility. We discuss this issue further in section VI.

The specification tests for both linear ARMA models reported in table 3 suggest conditional heteroskedasticity as well as the presence of nonlinear effects in the residuals (NL test). This motivates the SM-ARMAX(1,1) model, which is also reported in table 3. According to the specification tests, this model captures the conditional heteroskedasticity and shows a vast improvement in the likelihood. One might expect that this improvement is due to the relaxation of the assumption of a constant conditional variance of volatility

imposed in the ARMAX specifications. However, even with no state dependence for the conditional variance of volatility, the log likelihood is improved from −938.770 to −486.031. That is, approximately one-half of the increase in log likelihood for SM-ARMAX relative to ARMAX can be attributed to conditional mean dynamics and the other half to dynamics of the conditional variance of volatility. The SM-ARMAX(1,1) uncovers a number of nonlinear dynamics in realized volatility, and these issues are explored further in subsection VB.

### A. Forecasts

Although the nonlinear model of realized volatility dominates the linear alternatives in-sample, these results may not extend out-of-sample. Out-of-sample evaluation through one-step-ahead forecasts are detailed in tables 4 and 5. In each case, estimates for a particular model are fixed at the reported values in table 3.

Table 4 displays the estimates and $R^2$ from a regression of realized volatility at time $t$ on a constant and a model estimate based on time $t - 1$ information. For all models, the intercept and slope coefficient are individually insignificant from 0 and 1, respectively (a sufficient condition for unbiased forecasts). However, based on $R^2$ statistics, the ARMAX model shows a 0.03 increase (compared to ARMA) in explanatory power for one-step-ahead realized volatility, whereas the SM-ARMAX(1,1) shows a 0.042 increase. Squared returns are another proxy (although a noisy one) of latent daily volatility. The $\tilde{R}^2$ in table 4 is the $R^2$ from a regression of daily squared returns on a constant and a model forecast of realized volatility. This provides a check on the results that use realized volatility as the forecast target. The apparent reduction in forecasting power as measured by the $\tilde{R}^2$ in this regression is consistent with the results in Andersen and Bollerslev (1998).

As a comparison, a GARCH(1,1) model estimated from daily returns data obtained an out-of-sample $R^2 = 0.384$. This represents a 0.033 decrease in explanatory power compared to the ARMA(1,1) model applied to realized volatility estimates and a 0.075 decrease relative to the SM-ARMAX(1,1) specification of the latter. In other words, using realized volatility estimates results in superior forecasts than do those based on latent volatility estimates derived from fitting a GARCH model to returns themselves.

Table 5 displays estimates from the out-of-sample regression of realized volatility on one-step-ahead forecasts from

TABLE 5.—FORECAST COMPARISONS, DM/USD

| $c_0$ | $c_1$ | $c_2$ | $R^2$ |
|---|---|---|---|
| −0.049 (0.041) | 1.172 (0.561) | −0.016 (0.522) | 0.460 |

The test regression is

$$\hat{\sigma}_t^2 = c_0 + c_1 E_{t-1}^1\hat{\sigma}_t^2 + c_2 E_{t-1}^2\hat{\sigma}_t^2 + \text{error},$$

where $E_{t-1}^1\hat{\sigma}_t^2$ is the one-step-ahead forecast from the SM-ARMA(1,1)-$r_{t-1}^2$ and $E_{t-1}^2\hat{\sigma}_t^2$ is the one-step-ahead forecast from the ARMA(1,1)-$r_{t-1}^2$. Out-of-sample observations are from 1801–2465. Heteroskedastic consistent standard errors appear in parenthesis.

---

[21] This is a heteroskedastic-robust version of the Tsay (1986) test.

[22] On the other hand, latent volatility estimates using innovations to daily returns (instead of realized volatility) give $\hat{\alpha} = 0.0497$ and $\hat{\beta} = 0.914$.

TABLE 6.—FULL-SAMPLE SM-ARMA(1,2)-$r_{t-1}^2$ ESTIMATES, DM/USD

| | SM-ARMA(1,2)-$r_{t-1}^2$ | |
| | $S_t = 1$ | $S_t = 2$ |
|---|---|---|
| $\phi(S_t)$ | 0.250 (0.037) | 0.454 (0.056) |
| $\psi(S_t)$ | −0.004 (0.160e-2) | 0.056 (0.018) |
| $\beta$ | 0.784 (0.030) | |
| $\alpha$ | 0.057 (0.005) | |
| $\theta_1$ | −0.427 (0.033) | |
| $\theta_2$ | −0.056 (0.017) | |
| $\lambda(S_t)$ | 0.119 (0.003) | 0.746 (0.028) |
| $\gamma_1(S_t)$ | 1.491 (0.140) | −0.495 (0.182) |
| $\gamma_2(S_t)$ | 0.052 (0.012) | 0.176 (0.042) |
| $\tilde{Q}(10)$ | 16.716 [0.081] | |
| $Q^2(10)$ | 7.977 [0.631] | |
| lgl | 147.161 | |

The SM-ARMA(1,2)-$r_{t-1}^2$ is

$\hat{\sigma}_t^2 = \mu(S_t, D_t) + \beta(\hat{\sigma}_{t-1}^2 - \mu(S_{t-1}, D_{t-1})) + \alpha r_{t-1}^2 + \theta_1\tilde{\eta}_{t-1} + \theta_2\tilde{\eta}_{t-2} + \eta_t(S_t)$

$\mu(S_t, D_t) = \phi(S_t) + \psi(S_t) D_t$

$\tilde{\eta}_{t-1} = \hat{\sigma}_{t-1}^2 - E_{t-2}\hat{\sigma}_{t-1}^2$

$\eta_t(S_t) \sim N(0, \lambda^2(S_t)),$

with state transition probabilities from equation (15) a $\tilde{Q}(10)$ is a heteroskedastic robust version of the Ljung-Box statistic for serial correlation in the residuals with ten lags. $Q^2(10)$ is the standard Ljung-Box statistic on the squared residuals.

both the SM-ARMAX(1,1) and ARMAX(1,1). These results suggest that the nonlinear model provides a marginal improvement in forecasting power. The heteroskedastic $t$-statistic on the SM-ARMAX(1,1) forecast is 2.09; it is −0.03 on the ARMAX(1,1) forecast.

To further investigate the out-of-sample forecasting power of the ARMAX and SM-ARMAX models, we considered the following proportional loss (PL) function.

$$PL = \frac{1}{T} \sum_{t=1}^{T} \log\left(\frac{\hat{\sigma}_t^2}{E_{t-1}\hat{\sigma}_t^2}\right).$$

Unlike the $R^2$ measure just reported, this loss function weights small and large forecast errors equally. If this measure is positive (negative), it means the forecasting model on average underpredicts (overpredicts) future volatility. The perfect measure for this loss function is 0. As shown in table 4, the out-of-sample loss for the ARMAX was −0.224, and for the SM-ARMAX it was −0.131. A similar ranking of the models is obtained from $\widetilde{PL} = \frac{1}{T}\Sigma_{t=1}^{T}\log\left(\frac{r_t^2}{E_{t-1}\hat{\sigma}_t^2}\right)$, which uses squared returns as the forecast target instead of realized volatility. These results indicate that all models tend to overpredict realized volatility (and squared returns), with the linear models doing this to a greater degree. One potential explanation for this difference is that linear ARMA models imply an exponential decay associated with innovations to realized volatility, whereas the nonlinear SM-ARMAX specification has the ability to make immediate abrupt adjustments in volatility as a result of a new innovation.

FIGURE 5.—HAZARD FUNCTION FOR SM-ARMA(1,2)-$r_{t-1}^2$
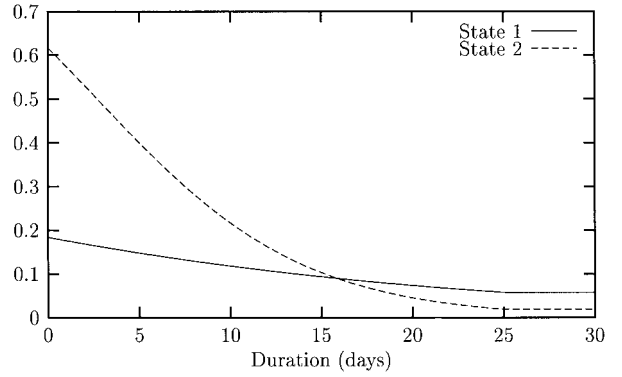


### B. Full-Sample Estimates

In this section, we present and discuss the full-sample estimates of the semi-Markov ARMA model. Estimates for the entire sample are found in table 6. Based on specification tests, we found it necessary to include a second MA term for this sample to eliminate residual serial correlation. Figure 5 displays the hazard function for states 1 and 2, and figure 6 shows the conditional mean of realized volatility as a function of duration.

The estimates provide evidence of regime switching. The test of no regime switching in the conditional mean of realized volatility ($\phi(1) = \phi(2)$, $\psi(1) = \psi(2) = 0$) and also the test of no regime switching in the conditional variance of realized volatility ($\lambda(1) = \lambda(2)$) are both strongly rejected. Moreover, duration is important in explaining within-state changes in the conditional mean. For instance, the $t$-statistic on $\psi(2)$ is 3.11 and indicates the conditional mean is increasing with the duration of state 2. (See figure 6.) The duration of a state also significantly affects the transition probabilities of both states through $\gamma_2(i)$, $i = 1, 2$.

State 1 of realized volatility is characterized by normal levels of realized volatility ($\mu(1, 1) = 0.246$, $\mu(1, 25) = 0.150$) coupled with a low conditional standard deviation ($\lambda(1) = 0.119$). State 2 displays a large and increasing level

FIGURE 6.—CONDITIONAL MEAN AS A FUNCTION OF DURATION $\phi(S_t) + \psi(S_t)D_t$
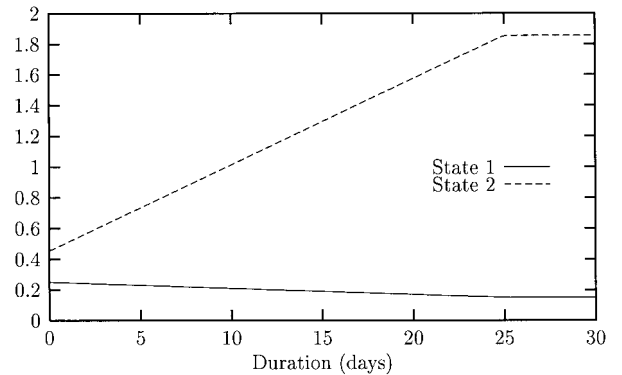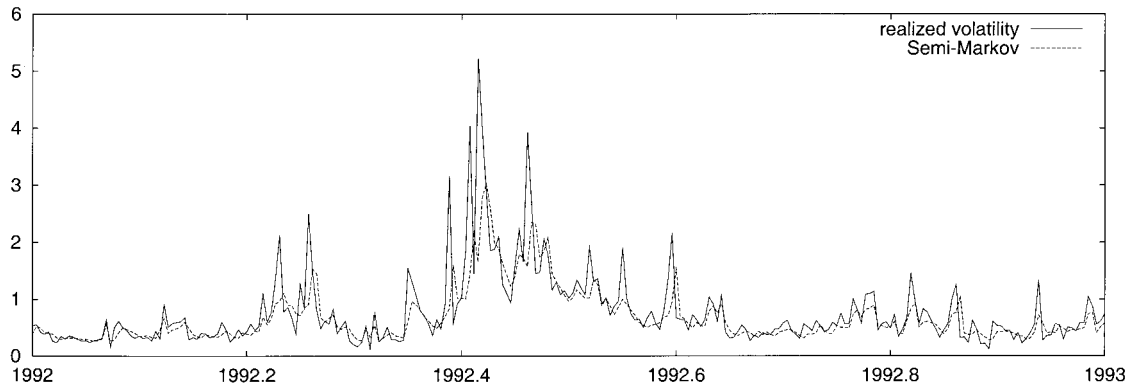
FIGURE 7.—IN-SAMPLE FITTED VALUES AND REALIZED VOLATILITY



of realized volatility as a function of duration ($\mu(2, 1) = 0.510$, $\mu(2, 25) = 1.854$), as well as a high conditional standard deviation ($\lambda(2) = 0.746$). The unconditional probabilities for states 1 and 2 are 0.81 and 0.19, respectively.

The hazard function for both states is shown in figure 5. Recall that this is the probability of exiting a state as a function of duration. Both states have a declining hazard function that is indicative of negative duration dependence. This figure suggests that the high variance in state 2 is not initially persistent and on average results in a switch to the lower-variance state. However, if realized volatility persists in state 2 for three or more periods, then it becomes persistent. Therefore, prolonged episodes of high variance in realized volatility, on average, lead to more of the same.

Finally, figure 7 shows realized volatility with the in-sample fitted values from the semi-Markov model, and figure 8 plots the probability of the low-volatility state (state 1) based on the full-sample smoother for a particularly volatile period in FX markets. Figure 8 indicates that the nonlinear model accurately identifies the sharp increases in the level and the volatility of realized volatility.

## VI.    Discussion

The SM-ARMAX model is an example of a doubly stochastic process. In this case, the two sources of random-ness are $\eta_t(S_t)$ conditional on the state and the latent state itself, $(S_t, D_t)$. As such, uncertainty about $\hat{\sigma}_t^2$ comes from these two sources. Although the largest component of uncertainty comes from the innovation variance, $\lambda^2(S_t)$ conditional on $S_t = 1$ or 2, regime changes and/or regime uncertainty about the conditional mean $\mu(S_t, D_t)$ are also an important component of the conditional variance of realized volatility. These two sources of randomness play an important role in the confidence we attach to future forecasts of realized volatility.

There are important differences between the nonlinear model and the linear ARMA models. Contrary to the evidence from the linear models in table 3, the semi-Markov model indicates that the standard deviation of the innovations to realized volatility is approximately 0.13 most of the time[23] and not 0.4 as the ARMA(1,1) and ARMAX(1,1) estimates imply. Therefore, based on the semi-Markov model, if we are very certain that next period will be normal (that is, state 1), then from a forecasting perspective we are much more confident about the behavior of realized volatility next period compared to what the linear models suggest. However, our estimates for the nonlinear model

[23] Eighty-one percent of the time, realized volatility is in state 1 according to the unconditional probabilities.

FIGURE 8.—PROBABILITY OF STATE 1
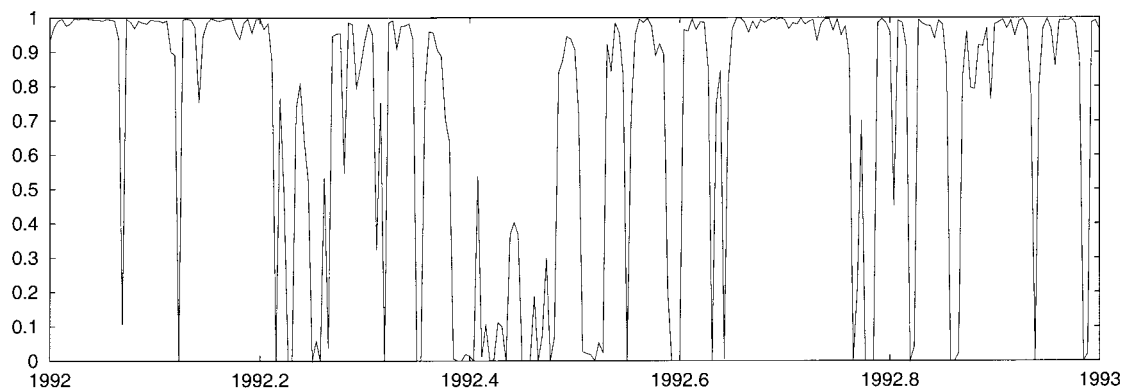
show that, 19% of the time, realized volatility undergoes a large increase in its conditional standard deviation to 0.792. This represents an increase of more than 600% in the standard deviation.

It is natural to interpret state 2 as capturing periods characterized by high volatility of FX rates. Switches between that state and the normal state will involve discrete jumps in volatility. This is an important issue in derivative asset pricing. Empirical evidence suggests that some form of discrete jumps may be necessary to fully describe speculative returns (Andersen, Benzoni, & Lund, 1999). The usual assumption in continuous-time models is that a Poisson process describes the discrete jumps, which implies that jumps are independent and not affected by the most recent information set. However, our results suggest that dependence in the probability of discrete jumps may be an important component of FX volatility dynamics.

Although we have presented out-of-sample forecasts for the various models using a particular sample period, we also explored a few other sample periods. In summary, we found that forecasting performance of the linear ARMAX models varied greatly, depending on the proportion of extremely high levels of realized volatility in the forecasting sample. However, the semi-Markov model consistently produced good forecast results relative to the linear models for different out-of-sample periods. Nevertheless, for small forecasting samples found at the end of our data set, both the ARMAX(1,1) and the semi-Markov model had out-of-sample $R^2$ in excess of 0.5.

Although several issues arise when discussing persistence, we can ask how today's volatility affects tomorrow's volatility. As noted, using realized volatility and unrestricted ARMA or ARMAX parameterizations makes it difficult to compare the coefficient estimates in table 3 to the SV-motivated approximate ARMA function in equation (11) or the GARCH-motivated restricted ARMAX function in equation (13). However, estimates from an ARMAX(1,1) with the restriction, $\beta = -\theta$ imposed gave $\hat{\alpha} + \hat{\beta} = 0.934$.[24] This estimate is relatively close to what we would expect for daily return data.[25] The unrestricted ARMAX estimates reported in table 3 represent a more flexible model than the GARCH(1,1), and therefore $\alpha + \beta$ is not necessarily a meaningful measure of persistence in latent daily volatility.

The results from the semi-Markov model indicate that the linear ARMAX(1,1) in realized volatility neglects several nonlinear features. First, our evidence points to a time-varying intercept parameter. Second, our results suggest that the stochastic process governing the time-varying parameters shows dependence. Finally, we find evidence of conditional heteroskedasticity for realized volatility. This study

points to the usefulness of recent research that explores the combination of GARCH or SV models with structural changes or regime switching dynamics.

In this paper, we have purposely studied levels of realized volatility to build on the large GARCH (Palm, 1996) and SV (Bates, 2000; Chernov et al., 1999) literature that investigates the importance of nonlinearities and asymmetries in levels of volatility. However, transformations of realized volatility may be useful. For example, density plots in Andersen, Bollerslev, Diebold, and Labys (2001) show that the unconditional distribution of log transforms of realized volatility is approximately normal. A starting point for such investigations in our context might be to apply a Box-Cox transformation to our ARMAX specification. In fact, this transformation has been used in the GARCH literature (Higgins & Bera, 1992). The log transform is a special case of the Box-Cox model and is used in EGARCH and many of the SV models. We did some preliminary analysis using log transforms of realized volatility. In summary, our main results do not change. The NL test, reported in table 3 for the ARMAX parameterizations of volatility levels, also strongly rejects an ARMAX(1,1) model of log volatility. We find evidence of regime switching, and the forecasting ranking of the models is preserved. Obviously, the dynamics (such as persistence in the transition matrix) change considerably. However, the main point is that the nonlinear features we uncover in FX volatility are not eliminated by the log transformation. Future work will investigate some of these issues.

## VII. Conclusion

This paper uses a nonparametric measure of daily volatility (realized volatility) that allows us to investigate the time series features of latent volatility and sidestep the choice of the innovation distribution driving returns. It is important to note that we explicitly incorporate the fact that realized volatility will not be an error-free measure of ex post latent volatility. Specifying a functional relationship between our direct estimate of volatility and the latent DGP for daily volatility allows the latter to be parameterized in terms of realized volatility, other variables in the information set, and an error term. In this framework, standard latent volatility models become variants of an ARMAX time series model. Due to our maintained relationship between realized volatility and latent volatility, any time series characteristics found using the former will also be true of the latter.

We begin by analyzing an ARMAX model and find evidence of nonlinear structure for which we propose a doubly stochastic process with duration-dependent mixing. Our nonlinear model uncovers several important features of FX volatility. For example, we find evidence of time-varying persistence in volatility. We also find that the conditional variance of volatility is time varying. FX returns exhibit normal periods of low and relatively stable volatility

---

[24] Equation (12) can be rearranged to give $\tilde{\sigma}_t^2 = \mu(1 - \beta) + (\alpha + \beta)\tilde{\sigma}_{t-1}^2 + \alpha(r_{t-1}^2 - \tilde{\sigma}_{t-1}^2)$.

[25] Latent volatility estimates using innovations to daily returns (instead of realized volatility) give $\hat{\alpha} + \hat{\beta} = 0.964$ for this sample.

and other periods of high and fluctuating volatility. The latter state does not occur often but periodically can persist for a longer period of time. The flexibility of our parameterization, which captures both volatility clustering and time-varying persistence, allows the high-volatility state to capture short-lived realizations in the tail of the unconditional distribution of volatility (for example, due to announcements) as well as more-sustained periods of high variability in FX returns. These different episodes of volatility embody very different levels of confidence that can be attached to forecasts.

Our evidence of nonlinearity in realized volatility has important implications for empirical finance including the precision of forecasts, hedging, and pricing of derivatives. This applies both to the effect of large abrupt changes in the level of volatility on FX options and to the effect of the time-varying variance of volatility for pricing derivatives, for which the underlying is a volatility measure such as a volatility index.

## REFERENCES

Alizadeh, S., M. W. Brandt, and F. X. Diebold, "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance* 57 (2002, forthcoming).

Andersen, T. G., L. Benzoni, and J. Lund, "Estimating Jump-Diffusions for Equity Returns," Kellogg Graduate School of Management manuscript, Northwestern University (1999).

Andersen, T. G., and T. Bollerslev, "Answering the Skeptics: Yes, Standard Volatility Models DO Provide Accurate Forecasts," *International Economic Review* 39:4 (1998), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics* 61:1 (2001), 43–76.

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

Andreou, E., and E. Ghysels, "Rolling-sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business and Economic Statistics* (2001, forthcoming).

Bai, X., J. R. Russell, and G. C. Tiao, "Beyond Merton's Utopia: Effects of Non-Normality and Dependence on the Precision of Variance Estimates using High-Frequency Financial Data," University of Chicago manuscript (1999).

Barndorff-Nielsen, O. E., and N. Shephard, "Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society,* Series B, 63 (2001), 167–241.

——, "Econometric Analysis of Realised Volatility and Its Uses in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society,* Series B, 64 (2002).

Bates, D. S., "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics* 94 (2000), 181–238.

Billio, M., A. Monfort, and C. Robert, "Bayesian Estimation of Switching ARMA Models," *Journal of Econometrics* 93:2 (1999), 229–255.

Bollerslev, T., and I. Domowitz, "Trading Patterns and Prices in the Interbank Foreign Exchange Market," *Journal of Finance* 48 (1993), 1421–1443.

Cai, J., "A Markov Model of Switching-Regime ARCH," *Journal of Business & Economic Statistics* 12:3 (1994), 309–316.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen, "A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation," CIRANO working paper (1999).

Drost, F., and T. Nijman, "Temporal Aggregation of GARCH Processes," *Econometrica* 61 (1993), 909–927.

Duan, J.-C., "Augmented GARCH(p,q) Process and its Diffusion Limit," *Journal of Econometrics* 79 (1997), 97–127.

Durland, J. M., and T. H. McCurdy, "Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth," *Journal of Business & Economic Statistics* 12:3 (1994), 279–288.

Engle, R. F., "Autoregressive Conditional Heteroskedasticity with Estimates of the UK Inflation," *Econometrica* 50 (1982), 987–1008.

Filardo, A. J., and S. F. Gordon, "Business Cycle Durations," *Journal of Econometrics* 85 (1998), 99–123.

French, K. R., G. W. Schwert, and R. F. Stambaugh, "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19 (1987), 3–29.

Gottschling, A., C. Haefke, and H. White, "Closed Form Integration of Artificial Neural Networks with Some Applications to Finance," University of California–San Diego manuscript (1999).

Gourieroux, C., and A. Monfort, "Qualitative Threshold ARCH Models," *Journal of Econometrics* 52 (1992), 159–200.

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

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

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

Higgins, M. L., and A. K. Bera, "A Class of Nonlinear ARCH Models," *International Economic Review* 33 (1992), 137–158.

Howard, R. A., *Semi-Markov and Decision Processes,* vol. II of *Dynamic Probabilistic Systems* (New York: John Wiley & Sons, 1971).

Hsieh, D., "Chaos and Nonlinear Dynamics: Applications to Financial Markets," *Journal of Finance* 46 (1991), 1839–1877.

——, "Implications of Nonlinear Dynamics for Financial Risk Management," *Journal of Financial and Quantitative Analysis* 28:1 (1993), 41–64.

Jorion, P., "On Jump Processes in the Foreign Exchange and Stock Markets," *Review of Financial Studies* 1:4 (1988), 427–445.

Kim, C. J., and C. R. Nelson, "Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching," this REVIEW, 80:2 (1998), 188–201.

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

Klaassen, F., "Improving GARCH Volatility Forecasts," Tilburg University manuscript (1998).

Lam, P., "A Markov Switching Model of GNP Growth with Duration Dependence," Federal Reserve Bank of Minneapolis discussion paper no. 124 (1997).

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

Maheu, J. M., and T. H. McCurdy, "Identifying Bull and Bear Markets in Stock Returns," *Journal of Business & Economic Statistics* 18:1 (2000a), 100–112.

——, "Volatility Dynamics under Duration-Dependent Mixing," *Journal of Empirical Finance* 7:3–4 (2000b), 345–372.

Meddahi, N., and E. Renault, "Temporal Aggregation of Volatility Models," CIRANO discussion paper no. 2000b-22 (2000).

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

Nelson, D. B., "ARCH Models as Diffusion Approximations," *Journal of Econometrics* 45 (1990), 7–39.

Pagan, A., and G. W. Schwert, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 45:1–2 (1990), 267–290.

Palm, F. C., "GARCH Models of Volatility" (pp. 209–240), in G. S.

Maddala and C. R. Rao (Eds.), *Handbook of Statistics,* vol. 14 (New York: Elsevier Science, 1996).

Poterba, J., and L. Summers, "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review* 76 (1986), 1142–1151.

Schwert, G., "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance* 44 (1989), 1115–1154.

———, "Stock Market Volatility," *Financial Analysts Journal* 46 (1990), 23–34.

Schwert, G. W., and P. J. Seguin, "Heteroskedasticit y in Stock Returns," *Journal of Finance* 45 (1990), 1129–1155.

So, M. K., K. Lam, and W. Li, "A Stochastic Volatility Model with Markov Switching," *Journal of Business & Economic Statistic s* 16 (1998), 244–253.

Taylor, S., and X. Xu, "The Incremental Information in One Million Foreign Exchange Quotations," *Journal of Empirical Finance* 4 (1997), 317–340.

Tsay, R. S., "Nonlinearit y Tests for Time Series," *Biometrika* 73:2 (1986), 461–466.

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