ASYMMETRIES, BREAKS, AND LONG-RANGE DEPENDENCE: AN ESTIMATION FRAMEWORK FOR TIME SERIES OF DAILY REALIZED VOLATILITY

ERIC HILLEBRAND AND MARCELO C. MEDEIROS

ABSTRACT. We study the simultaneous occurrence of long memory and nonlinear effects such as structural breaks and thresholds in conditional volatility. We propose a model framework for returns and conditional volatility and specify a Lagrange-multiplier test for nonlinear terms in the volatility equation in the presence of long memory. The system allows for general nonlinear functions in the volatility equation. Asymptotic theory for the quasi-maximum likelihood estimator of the system is provided using a triangular array setup. The theoretical results in the paper can be applied to any series with long memory and nonlinearity. The methodology is applied to individual stocks of the Dow Jones Industrial Average during the period 1995 to 2005. As a proxy for conditional volatility we consider a kernel-based realized volatility measure. We find strong evidence of nonlinear effects and explore different specifications of the model framework to study changes in the mean of realized volatility and leverage effects. A forecasting exercise demonstrates that allowing for nonlinearities in long memory models yields performance gains.
KEYWORDS: Realized volatility, structural breaks, smooth transitions, nonlinear models, long memory, persistence, triangular array asymptotics.

1. INTRODUCTION

In this paper we propose a system of equations framework to model the conditional mean and variance of daily asset returns using realized volatility time series data. The modeling framework disentangles the confounding effects of long memory and non-linearities such as change points and threshold effects in volatility. We study the asymptotic behavior of the maximum likelihood estimator and propose a Lagrange multiplier test for the null hypothesis of linearity of the volatility equation against the alternative of unspecified non-linearity in the presence of long memory. The test and estimation procedure can be applied to any time series that is suspected to have long memory and nonlinear effects, such that the results in the paper are not restricted to financial volatility. We apply our modeling and test procedure to 28 stocks of the Dow Jones Industrial Average during the period 3-Jan-1995 to 31-Dec-2005.

Financial volatility plays a central role in risk-management. Andersen, Bollerslev, Christoffersen, and Diebold (2007) give a recent overview. Earlier classes of volatility models such as (Generalized) Autoregressive Conditional Heteroskedasticity proposed by Engle (1982) and Bollerslev (1986), stochastic volatility models of Taylor (1986), or exponentially weighted moving averages (J. P. Morgan 1996) used squared daily returns as measure for volatility (see McAleer (2005) for an exposition). Since this measure is very noisy, volatility was specified as unobservable, latent conditional standard

Date: February 14, 2008.

deviation. However, as noted by Bollerslev (1987), Malmsten and Teräsvirta (2004), and Carnero, Peña, and Ruiz (2004) among others, most of the latent volatility models fail to capture salient features of financial time series. For example, standard latent volatility models fail to describe adequately the slowly decreasing autocorrelation in the squared returns that is associated with the high kurtosis of returns.

High frequency intra-day data have been used to construct estimates of volatility that are less contaminated by noise (Andersen and Bollerslev 1998, Andersen, Bollerslev, Diebold, and Ebens 2001, Andersen, Bollerslev, Diebold, and Labys 2001b, Andersen, Bollerslev, Diebold, and Labys 2001a, Andersen, Bollerslev, Diebold, and Labys 2003). Merton (1980) noted that the variance of a semi-martingale over a fixed interval can be estimated as the sum of squared realizations within that interval, provided the sampling mesh is sufficiently small. Since semi-martingales are a common model for asset prices, this idea could in principle be applied to intra-day asset price data if these are sampled at sufficiently high frequency. Then, the sum of intra-day squared returns is called realized variance, and its square root is called realized volatility. Andersen and Bollerslev (1998) showed that daily foreign exchange volatility can be measured by aggregating squared five-minute returns. The five-minute frequency is a trade-off between accuracy and microstructure noise that can arise through bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors (Madhavan 2000, Biais, Glosten, and Spatt 2005).

Realized volatility reduces the noise in the volatility estimate considerably compared to squared or absolute daily returns. This enables researchers to specify explicit models for volatility and obviates the latent variable approach. Realized volatility can also be used as a benchmark for the forecasting performance of latent variable models (Andersen and Bollerslev 1998, Hansen and Lunde 2005, Patton 2005). Measurement error still remains an issue and is studied, for example, in Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold, and Labys (2003), and Meddahi (2002). There are now a number of consistent estimators of realized volatility for one day in the presence of microstructure noise: the two-time scales realized volatility estimator proposed by Zhang, Mykland, and Aït-Sahalia (2005), the realized kernel estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007), and the modified MA filter of Hansen, Large, and Lunde (2007); see McAleer and Medeiros (2006) for a recent review.

The day-to-day dynamics of realized volatility exhibit long memory or high persistence, just as the dynamics of squared or absolute daily returns (for example Ding, Granger, and Engle 1993). Andersen, Bollerslev, Diebold, and Labys (2003) use an ARFIMA specification to model this long-range dependence. An alternative to ARFIMA are models that approximate long memory by aggregation. Volatility is modeled as a sum of different processes, each with low persistence. The aggregation induces long memory; see, for example, Granger (1980), LeBaron (2001), Fouque, Papanicolaou, Sircar, and Sølna (2003), Davidson and Sibbertsen (2005), or Hyung, Poon, and Granger (2005). This phenomenon is utilized in Corsi's (2004) widely used HAR-RV model (Heterogeneous Autoregressive Model for Realized Volatility), which builds on the HARCH specification proposed by Müller, Dacorogna, Dave, Olsen, Pictet, and von Weizsäcker (1997).

The literature has also documented nonlinear effects in volatility, such as leverage and feedback effects or multiple regimes (Black 1976, Nelson 1991, Glosten, Jagannathan, and Runkle 1993, Campbell and Hentschel 1992). Regime changes can take the form of switches in the model parameters, for instance governed by a Markov chain as in Hamilton and Susmel (1994), Cai (1994) and Gray (1996), hard thresholds as discussed in Rabemananjara and Zakoian (1993), Li and Li (1996), and Liu, Li, and Li (1997), or smooth transitions as in Hagerud (1997), Gonzalez-Rivera (1998), or Medeiros and Veiga (2007). Commonly found are two regimes of low persistence for large positive and large negative returns and one regime of high persistence for medium-range returns; see for instance, Longin (1997) and Medeiros and Veiga (2007).

The statistical consequences of neglecting or misspecifying nonlinearities have been discussed in the context of structural breaks in the GARCH literature (Diebold 1986, Lamoureux and Lastrapes 1990, Mikosch and Starica 2004, Hillebrand 2005) and in the literature on long memory models (Lobato and Savin 1998, Granger and Hyung 2004, Diebold and Inoue 2001, Granger and Teräsvirta 2001, Smith 2005). Neglected changes in levels or persistence induce estimated high persistence. This has often been called "spurious" high persistence; see also Hillebrand and Medeiros (2006).

In the reverse direction, it is also possible to mistake data-generating high persistence (in the form of long memory or unit roots) for nonlinearity. Spuriously estimated structural breaks were reported for unit root processes (Nunes, Kuan, and Newbold 1995, Bai 1998) and extended to long memory processes (Hsu 2001). In summary, it has been found over a wide array of studies that nonlinearity (such as breaks) and long memory (or high persistence) are confounding factors.

Given these findings, it is desirable to have a modeling framework that is able to capture nonlinearities in the presence of long memory. Such a framework is a step towards disentangling the confounding factors. Recently, a number of papers have studied long memory, jumps, leverage, and volatility-in-mean effects (Martens, van Dijk, and de Pooter 2004, Christensen and Nielsen 2007, Christensen, Nielsen, and Zhu 2007, Andersen, Bollerslev, Federiksen, and Nielsen 2007, Andersen, Bollerslev, and Huang 2007, Bollerslev, Kretschmer, Pigorsch, and Tauchen 2007). The studies most closely related to ours are Baillie and Kapetanios (2007a, 2007b), who also propose a (different) test for nonlinearity in the presence of long memory and develop a smooth transition autoregression which is embedded within a long memory process. Following a different approach, McAleer and Medeiros (2007) put forward a nonlinear HAR-RV model that is able to describe both long range dependence and nonlinear dynamics, such as leverage effects.

Our model adds to this literature in two ways. (1) We propose a system of equations for returns and realized volatility of an asset that allows for long memory and general non-linearities in volatility. We derive the asymptotic behavior of the quasi-maximum-likelihood estimator of the parameter vector. The challenge is to allow for changing regimes in time. Then, as the number of observations approaches infinity, the parameters of regimes of finite length become unidentified. We employ triangular array asymptotics to solve this problem (Saikkonen and Choi 2004, Andrews and McDermott 1995). (2) We develop a Lagrange multiplier test for the null hypothesis of linearity of the volatility equation against general non-linearity *in the presence of long memory*. Our test statistic is based on the results of van Dijk, Franses, and Paap (2002) and Medeiros, Teräsvirta, and Rech (2006). The test employs a Taylor series approximation of the unknown nonlinear structure. Common test statistics for general non-linearity are proposed by Teräsvirta (1994) and Hansen (1996). For the test of Teräsvirta (1994), Andersson, Eklund, and Lyhagen (1999) provide simulation evidence that a long memory process appears to be nonlinear if the lagged time series is used as threshold variable, indicating a size problem due to the confusion of long memory and non-linearity. We show in size discrepancy simulations that our test is able to address this problem.

Applying our model and testing framework to 28 stocks of the Dow Jones Industrial Average, we find evidence of structural breaks in the individual realized volatility time series, in particular a transition from high to low volatility around the year 2003. Dependence of volatility on the level of lagged returns is a robust finding across all stocks and in different model specifications, indicating leverage and asymmetry effects. Both, long memory and non-linear effects like change-points and leverage effects coexist in the realized volatility data. Accounting for non-linear terms in the volatility model specification yields forecast gains as we show in a prediction experiment.

The paper is organized as follows. Section 2 presents the model and develops the asymptotic theory of the quasi maximum likelihood estimation. The linearity test is introduced in Section 3. Monte Carlo evidence for its adequacy is reported and we describe how the test can be used in the model selection process. Empirical results are shown in Section 4. Section 5 concludes.

2. Long Memory and Nonlinearity in Realized Volatility

2.1. **Model Specification.** We specify a system of non-linear equations for returns and conditional volatility:

$$r_t = \boldsymbol{\beta}'\mathbf{x}_t + \lambda v_t + \sigma_t e_t \tag{1}$$

$$v_t := (1-L)^d \log(\sigma_t) = g(\mathbf{z}_t;\boldsymbol{\xi}) + \Theta(L)u_t, \tag{2}$$

where $\mathbf{x}_t = (1, \widetilde{\mathbf{x}}_t')' \in \mathbb{R}^{k_x+1}$ and $\widetilde{\mathbf{x}}_t$ is a vector of k_x explanatory variables for the conditional mean of returns, which may include lagged values of the returns, days-of-the-week dummies and announcement dates. $\boldsymbol{\beta}$ is a (k_x+1)-vector of parameters, $\lambda \in \mathbb{R}$ is a volatility-in-mean coefficient. From equations (1) and (2) it is clear that $\mathbb{E}\left(r_t|\mathbf{x}_t,\sigma_t,\mathcal{F}_{t-1}\right) = \boldsymbol{\beta}'\mathbf{x}_t + \lambda v_t$ and $\mathrm{Var}\left(r_t|\mathbf{x}_t,\sigma_t,\mathcal{F}_{t-1}\right) = \sigma_t^2$. $\mathcal{F}_{t-1}$ is the filtration given by all information up to time $t-1$. Conditional volatility enters the return equation twice: as fractionally differenced log conditional volatility v_t in the volatility-in-mean term and as standard deviation σ_t in the error term. Note that fractionally differencing the volatility-in-mean term ensures that the long memory property of volatility does not carry over to the returns.

The volatility equation (2) specifies the model for σ_t. Here, $d \in (-1/2, 1/2)$ is the fractional differencing parameter, σ_t is conditional volatility, the function $g(\mathbf{z}_t; \boldsymbol{\xi})$ is some nonlinear function to be specified, which is indexed by the vector of parameters $\boldsymbol{\xi} \in \mathbb{R}^{k_\xi}$, and $\mathbf{z}_t \in \mathbb{R}^{k_z}$ is a vector of explanatory variables for the conditional variance possibly including lagged values of v_t. $\mathbf{x}_t$ and $\mathbf{z}_t$ may have common elements. The returns r_t and the volatility process σ_t are observable but exhibit errors e_t and u_t, which are independent martingale difference sequences. $\Theta(L) = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)$ is a moving average lag polynomial of order q. We assume that all dependencies between returns and volatility are explicitly captured in the volatility-in-mean parameter λ and the nonlinear term $g(\cdot)$. The model is indexed by the vector of parameters $\boldsymbol{\psi} = \left(\boldsymbol{\beta}', \lambda, d, \boldsymbol{\xi}', \boldsymbol{\theta}', \sigma_u^2\right)' \in \mathbb{R}^{k_\psi}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)' \in \mathbb{R}^q$.

In the asymptotic derivations in Section 2.3 we assume that the conditional volatility process is $\mathcal{F}_t-$measurable. We then replace σ_t by an unbiased and consistent estimator of daily integrated volatility, the kernel-based realized volatility estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007). Appendix A briefly introduces the theoretical foundations of realized volatility and provides a discussion of this assumption.

2.2. **Interpretation.** The choice of the function $g(\cdot)$ is very flexible and allows for different specifications. The following examples list some possibilities.

EXAMPLE 1 (Linear ARFIMA). *Set $\mathbf{z}_t = (v_{t-1}, \ldots, v_{t-p})'$ and consider the following choice for the function $g(\cdot)$:*

$$g(\mathbf{z}_t; \boldsymbol{\xi}) = \phi_0 + \phi_1 v_{t-1} + \cdots + \phi_p v_{t-p}.$$

In that case, equation (2) may be written as

$$\Phi(L)(1-L)^d \left[\log(\sigma_t) - \mu\right] = \Theta(L) u_t,$$

where $\Phi(L) = \left(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p\right)$ and $\mu = \Phi^{-1}(1)\phi_0$. In this case volatility follows an ARFIMA(p,d,q) model. If $d = 0$, the volatility process is short-memory. This type of specification of the volatility equation was advocated in Andersen, Bollerslev, Diebold, and Labys (2003).

EXAMPLE 2 (ARFIMA with smoothly changing parameters). *Define $\mathbf{w}_t = (v_{t-1}, \ldots, v_{t-p})'$ and set $\mathbf{z}_t = (\mathbf{w}_t', t)'$. v_t is defined as above. Consider the following choice for the function $g(\cdot)$:*

$$g(\mathbf{z}_t; \boldsymbol{\xi}) = \phi_0 + \boldsymbol{\phi}' \mathbf{w}_t + \left(\widetilde{\phi}_0 + \widetilde{\boldsymbol{\phi}}' \mathbf{w}_t\right) f[\gamma(t-c)],$$

where $f(y) = (1 + e^{-y})^{-1}$ is the logistic function. Equation (2) becomes

$$\begin{aligned}(1-L)^d \log(\sigma_t) =& \phi_0 + \sum_{i=1}^{p} \phi_i (1-L)^d \log(\sigma_{t-i}) + \\ & \left\{\widetilde{\phi}_0 + \sum_{i=1}^{p} \widetilde{\phi}_i (1-L)^d \log(\sigma_{t-i})\right\} f[\gamma(t-c)] + \Theta(L) u_t.\end{aligned}$$

The parameter γ controls the smoothness of the transition. In the limit $\gamma \longrightarrow \infty$, the model becomes an ARFIMA model with a structural break at $t = c$. In the regression framework this type of specification has been considered in Lin and Teräsvirta (1994).

A possible generalization of the model is to follow the ideas in Medeiros and Veiga (2003) and consider the specification below:

$$\begin{aligned}(1-L)^d \log(\sigma_t) = \phi_0 &+ \sum_{i=1}^{p} \phi_i (1-L)^d \log(\sigma_{t-i}) \\ &+ \sum_{m=1}^{M} \left\{ \widetilde{\phi}_{0,m} + \sum_{i=1}^{p} \widetilde{\phi}_{i,m} (1-L)^d \log(\sigma_{t-i}) \right\} f[\gamma_m (t - c_m)] \\ &+ \Theta(L) u_t.\end{aligned}$$

EXAMPLE 3 (ARFIMA with asymmetry). *Now let $\mathbf{z}_t = (\mathbf{w}_t', e_{t-1})'$ with $\mathbf{w}_t$ as in the example above. One possibility to accommodate leverage effects is to choose $g(\cdot)$ as*

$$g(\mathbf{z}_t; \boldsymbol{\xi}) = \phi_0 + \boldsymbol{\phi}'\mathbf{w}_t + \left(\widetilde{\phi}_0 + \widetilde{\boldsymbol{\phi}}'\mathbf{w}_t \right) f(\gamma e_{t-1}),$$

with $f(\cdot)$ being again the logistic function. In the case $\gamma \longrightarrow \infty$ the logistic function becomes a step function and the model has the same flavor as the GJR-GARCH specification of Glosten, Jagannathan, and Runkle (1993). See van Dijk, Franses, and Paap (2002) for a related specification for macroeconomic time series and Hagerud (1997), Gonzalez-Rivera (1998), Lundbergh and Teräsvirta (1998) for similar ideas in latent volatility models.

Another possible generalization is to consider multiple regimes as in Medeiros and Veiga (2007):

$$g(\mathbf{z}_t; \boldsymbol{\xi}) = \phi_0 + \boldsymbol{\phi}'\mathbf{w}_t + \sum_{m=1}^{M} \left(\widetilde{\phi}_{0,m} + \widetilde{\boldsymbol{\phi}}_m'\mathbf{w}_t \right) f\left[\gamma_m \left(e_{t-1} - c_m\right)\right].$$

The number of regimes is defined by the parameter M. For example, suppose that $M = 2$, c_1 is highly negative, and c_2 is large and positive. Then the resulting model model will have three regimes that can be interpreted as responding to very low negative shocks, tranquil periods, and highly positive shocks, respectively.

EXAMPLE 4 (General Nonlinear ARFIMA). *An interesting alternative is to leave the type of nonlinearity partially unspecified. This can be done by specifying the function $g(\cdot)$ as a single hidden layer neural network (NN) of the following form*

$$g(\mathbf{z}_t; \boldsymbol{\xi}) = \nu_0 + \boldsymbol{\nu}'\mathbf{z}_t + \sum_{m=1}^{M} \nu_m f\left[\gamma_m \left(\boldsymbol{\omega}_m'\mathbf{z}_t - \eta_m\right)\right], \tag{3}$$

where again $f(\cdot)$ is the logistic function, $\gamma_m > 0$, and $\|\boldsymbol{\omega}_m\| = 1$ with

$$\omega_{m1} = \sqrt{1 - \sum_{j=2}^{q} \omega_{mj}^2}, \quad m = 1, \ldots, M.$$

This is a long-memory version of the model discussed in Medeiros, Teräsvirta, and Rech (2006).

In a related paper, Martens, van Dijk, and de Pooter (2004) put forward a model to jointly describe long-range dependence, nonlinearity, structural breaks and effects of days of the week. The model considered in their paper is nested in specification (1), (2). We also propose a linearity test and asymptotic theory for the maximum likelihood estimator.

The return equation (1) is flexible enough to allow for permanent and temporary components specifications (Fama and French 1988) with or without exogenous predictors (Harvey 2001). The volatility-in-mean component is included to consistently capture the correlation between contemporaneous returns and volatility. This correlation is usually found to be negative (Brandt and Kiang 2004) and is attributed to the leverage effect (Black 1976) and to the feedback effect (Campbell and Hentschel 1992). Our model does not resolve the difference between the two effects; our research focus is on volatility dynamics. The volatility-in-returns term in equation (1) captures this direction of the correlation while we focus on returns-in-volatility effects like asymmetry and leverage in equation (2).

Model (1) and (2) is a model for time series of daily realized volatility and daily returns. We do not advocate it as data-generating process for *intraday* returns and volatility. It is in general possible, however, to interpret the model as data-generating process for intraday data. Appendix B provides some details. It is also possible to specify continuous time diffusions as models of intraday data that imply many of the statistical features considered in (1) and (2). Such a diffusion may feature stochastic volatility that is driven by the sum of several Ornstein-Uhlenbeck processes of different decorrelation lengths (LeBaron 2001, Fouque, Papanicolaou, and Sircar 2000). This aggregation induces long memory similar to the argument in Granger (1980). Asymmetry effects can be captured by negative contemporaneous correlation between the Brownian motions in the return and in the volatility equation.

2.3. **Parameter Estimation.**

2.3.1. *Triangular Arrays.* In this paper, we will use triangular array asymptotics to analyze model (1), (2) (Saikkonen and Choi 2004, Andrews and McDermott 1995). Let T_0 be the actual sample size. Then, model (1), (2) is embedded in a sequence of models

$$r_{tT} = \boldsymbol{\beta}'\mathbf{x}_{tT} + \lambda v_{tT} + \sigma_{tT} e_t, \tag{4}$$

$$v_{tT} := (1-L)^d \log(\sigma_{tT}) = g(\mathbf{z}_{tT};\boldsymbol{\xi}) + \Theta(L)u_t, \tag{5}$$

where $y_{tT} := (T_0/T)y_t$ for any sequence $\{y_t\}$, $t = 1,\ldots,T$.

To illustrate the point of writing the model in triangular arrays, consider Example 2, where $g(\mathbf{z}_t;\boldsymbol{\xi})$ contains a single logistic transition in an otherwise linear model:

$$g(\mathbf{z}_t;\boldsymbol{\xi}) = \phi_0 + \boldsymbol{\phi}'\mathbf{w}_t + \left(\widetilde{\phi}_0 + \widetilde{\boldsymbol{\phi}}'\mathbf{w}_t\right) f[\gamma(t-c)].$$

Since, for T large,

$$f[\gamma(t-c)] = f[T\gamma(T^{-1}t - T^{-1}c)] \approx \mathbb{1}_{\{T^{-1}t>0\}},$$

the parameters ϕ_0 and $\boldsymbol{\phi}$ that govern the first regime as well as the transition parameters γ and c vanish from the model and become unidentified. Triangular array asymptotics consider suitably scaled variables, here:

$$f\left[\gamma\left(\frac{T_0}{T}t - c\right)\right] = f\left[T^{-1}\gamma\left(T_0 t - Tc\right)\right].$$

Here, the slope of the logistic function is decreasing with T while the locus of the transition is increasing with T, whereas the scaling of the time counter, T_0, remains constant. Thus, the proportions of observations in the first regime, during the transition, and in the last regime remain the same. The parameters in these groups of observations remain identified.

2.3.2. *Assumptions.* We denote the data-generating parameter vector as

$$\boldsymbol{\psi}_0 = \left(\boldsymbol{\beta}_0', \lambda_0, d_0, \boldsymbol{\xi}_0', \boldsymbol{\theta}_0', \sigma_{u,0}^2\right)'.$$

We write $e_t(\boldsymbol{\psi})$ and $u_t(\boldsymbol{\psi})$ such that the notation can be used for both, the residuals from the estimation and the data-generating errors:

$$e_t(\boldsymbol{\psi}) = \sigma_{tT}^{-1}(r_{tT} - \boldsymbol{\beta}'\mathbf{x}_{tT} - \lambda v_{tT}),$$
$$u_t(\boldsymbol{\psi}) = \Theta^{-1}(L)\left[(1-L)^d \log \sigma_{tT} - g(\mathbf{z}_t; \boldsymbol{\xi})\right],$$

and we use the shorthand notation $e_{t,0} := e_t(\boldsymbol{\psi}_0)$, $u_{t,0} := u_t(\boldsymbol{\psi}_0)$ for the data-generating errors and e_t, u_t for $e_t(\boldsymbol{\psi})$ and $u_t(\boldsymbol{\psi})$. Note that the fractional integration parameter d is an element of $\boldsymbol{\psi}$ and estimated jointly with the other parameters. Maximum likelihood estimation of d is addressed in Sowell (1992) and Chung and Baillie (1993).

ASSUMPTION 1 (Parameter Space). *The parameter vector $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi} \subset \mathbb{R}^{k_x+k_\xi+q+4}$ is an interior point of $\boldsymbol{\Psi}$, a compact parameter space.*

ASSUMPTION 2 (Measurability of σ_{tT}). *The conditional variance process σ_{tT} is $\mathcal{F}_t$-measurable.*

This assumption is discussed in Appendix A.

ASSUMPTION 3 (Errors).

(1) *$e_{t,0}$ is a martingale difference sequence with mean zero and unit variance.*
(2) *$u_{t,0}$ is a martingale difference sequence with mean zero and constant variance $\sigma_{u,0}^2 > 0$.*
(3) $\mathbb{E}|u_{t,0}|^q < \infty$ *for* $q = 1, 2, 4$.
(4) $\mathbb{E}\exp(u_{t,0})^q < \infty$ *for* $q = 1, 2, 4$.
(5) *$e_{t,0}$, $u_{t,0}$ are independent.*

ASSUMPTION 4 (Stationarity and Moments).

(1) *The model defined by equations (4) and (5) is stationary and ergodic.*
(2) $\mathbb{E}|\mathbf{x}_{tT}|^q < \infty$, $q = 1, 2, 4$.
(3) $\mathbb{E}|\mathbf{z}_{tT}|^q < \infty$, $q = 1, 2, 4$.

Assumption 4 implies that $\mathbf{x}_{tT}$ and $\mathbf{z}_{tT}$ are stationary and ergodic, $d_0 \in (-1/2, 1/2)$, and $\Theta(L)$ is invertible.

ASSUMPTION 5 (Nonlinear Function).

(1) $g(\mathbf{z}_{tT};\, \boldsymbol{\xi})$ *is continuous in* $\boldsymbol{\xi}$ *and measurable in* $\mathbf{z}_{tT}$.
(2) $g(\mathbf{z}_{tT};\, \boldsymbol{\xi})$ *is parameterized such that the parameters are well defined.*
(3) $g(\mathbf{z}_{tT};\, \boldsymbol{\xi})$ *and* $u_{t,0}$ *are independent.*
(4) $\mathbb{E}|g(\mathbf{z}_{tT};\, \boldsymbol{\xi})|^q < \infty$, $q = 1, 2, 4$.
(5) $\mathbb{E}\left\{\exp\left[g\left(\mathbf{z}_{tT};\, \boldsymbol{\xi}\right)\right]^q\right\} < \infty$, $q = 1, 2, 4$.
(6) $\mathbb{E}\left|\frac{\partial}{\partial \boldsymbol{\xi}} g(\mathbf{z}_{tT};\, \boldsymbol{\xi})\right|^q < \infty$, $q = 1, 2, 4$. *Note: We allow for* $\boldsymbol{\xi}_i = d$ *for some* i.
(7) $\mathbb{E}\left|\frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} g(\mathbf{z}_{tT};\, \boldsymbol{\xi})\right|^q < \infty$, $q = 1, 2$.

EXAMPLE 5 (for Assumption 5 (2): Logistic Transition). *If there are* $H+1$ *different regimes of volatility depending on a state variable* s_t *(for example past excess returns* e_{t-1} *or time* t*) with transitions governed by logistic functions, then the transition parameters* c_i *and* γ_i, $i = 1, \ldots, H$ *are such that*

(1) $-\infty < -M < c_1 < \ldots < c_H < M < \infty$.
(2) $\gamma_i > 0$ *for all* i.
(3) $f[\gamma_1(s_t - c_1)] \geq f[\gamma_2(s_t - c_2)] \geq \ldots \geq f[\gamma_H(s_t - c_H)]$.

2.3.3. *Quasi Maximum Likelihood Estimator.* The assumption on the error vector $\boldsymbol{\epsilon}_{t,0} := (e_{t,0}, u_{t.0})'$ implies that $\mathbb{E}(\boldsymbol{\epsilon}_{t,0}\boldsymbol{\epsilon}_{t,0}') = \text{diag}(1, \sigma_u^2)$ and the quasi log-likelihood function is given by

$$\mathcal{L}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\psi}),$$

where

$$\ell_t(\boldsymbol{\psi}) = -\frac{1}{2}\left(\log 2\pi + \log \sigma_u^2 + e_t^2 + u_t^2 \sigma_u^{-2}\right).$$

The parameter vector is estimated by quasi maximum likelihood as

$$\widehat{\boldsymbol{\psi}}_T = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\text{argmax}}\, \mathcal{L}_T(\boldsymbol{\psi}),$$

where $\boldsymbol{\Psi}$ is the parameter space.

THEOREM 1 (Consistency). *Under Assumptions 1 through 5, the quasi maximum likelihood estimator* $\widehat{\boldsymbol{\psi}}_T$ *is consistent:*

$$\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_0.$$

The proof is provided in the Appendix.

THEOREM 2 (Asymptotic Normality). *Under Assumptions 1 through 5, the quasi maximum likelihood estimator* $\widehat{\boldsymbol{\psi}}_T$ *is asymptotically normally distributed:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0\right) \xrightarrow{d} \mathcal{N}\left(0,\, \mathbf{A}(\boldsymbol{\psi}_0)^{-1}\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{A}(\boldsymbol{\psi}_0)^{-1}\right),$$

where

$$\mathbf{A}(\boldsymbol{\psi}_0) = -\mathbb{E}\left.\frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0},$$

$$\mathbf{B}(\boldsymbol{\psi}_0) = \mathbb{E}\left(\left.\frac{\partial \ell_t}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} \left.\frac{\partial \ell_t}{\partial \boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right).$$

The proof is provided in the Appendix.

PROPOSITION 1 (Covariance Matrix Estimation). *Under Assumptions 1 through 5,*

$$\mathbf{A}_T \xrightarrow{p} \mathbf{A}, \quad \mathbf{B}_T \xrightarrow{p} \mathbf{B},$$

where

$$\mathbf{A}_T(\boldsymbol{\psi}) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'},$$

and

$$\mathbf{B}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}'},$$

and $\mathbf{A}$, $\mathbf{B}$ *as defined in Theorem 2.*

The proof is provided in the Appendix.

3. LINEARITY TESTING AGAINST GENERAL NONLINEARITY

3.1. **Test Statistic.** In this section we will describe a linearity test against a flexible nonlinear form. We advocate the use of the neural network (NN) specification as in Example 4. This type of model has found applications in a number of fields, including economics and finance. The use of the NN model in applied work is generally motivated by a mathematical result stating that under mild regularity conditions, a relatively simple NN model is capable of approximating any Borel-measurable function to any given degree of accuracy. See, for example, Fine (1999) and the references therein.

The testing procedure will be partially based on the results of Teräsvirta, Lin, and Granger (1993) and Medeiros, Teräsvirta, and Rech (2006). Consider a similar specification as in Example 4 with $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{s}_t')' \in \mathbb{R}^{k_w + k_s}$. To simplify the exposition we consider the case where there is no moving average term ($q = 0$). However, it is not difficult to extend our results to the case with $q > 0$. The volatility equation in model (1) and (2) using (3) can be rewritten as

$$(1-L)^d \log(\sigma_t) = \nu_0 + \boldsymbol{\nu}'\mathbf{z}_t + \sum_{m=1}^{M} \nu_m f\left[\gamma_m\left(\boldsymbol{\omega}_m'\mathbf{w}_t - \eta_m\right)\right] + u_t. \qquad (6)$$

Consider the case where we want to test $M = 0$ against $M > 0$. The appropriate null hypothesis is

$$\mathbb{H}_0: \gamma_1 = \gamma_2 = \cdots = \gamma_M = 0. \qquad (7)$$

Under (7), the additional hidden unit is identically equal to a constant and merges with the intercept in the linear unit.

Under the null of linearity the parameters of (6) can be estimated consistently. Model (6) is only identified under the alternative, which means that the standard asymptotic inference is not available. This problem is circumvented as in Medeiros, Teräsvirta, and Rech (2006) by expanding $f\left[\gamma_m\left(\boldsymbol{\omega}_m'\mathbf{w}_t - \eta_m\right)\right]$, $m = 1, \ldots, M$, into a Taylor series around the null hypothesis (7). The order of the expansion is a compromise between a small approximation error (high order) and availability of data (short time series necessarily imply a relatively low order). Using a third-order Taylor expansion, rearranging and merging terms results in the following model

$$\begin{aligned}(1-L)^d \log(\sigma_t) = \pi_0 + \boldsymbol{\pi}'\mathbf{w}_t + \boldsymbol{\rho}'\mathbf{s}_t &+ \sum_{i=1}^{k_w}\sum_{j=i}^{k_w} \rho_{ij} w_{i,t} w_{j,t} \\ &+ \sum_{i=1}^{k_w}\sum_{j=i}^{k_w}\sum_{k=j}^{k_w} \rho_{ijk} w_{i,t} w_{j,t} w_{k,t} + u_t^*,\end{aligned} \tag{8}$$

where $u_t^* = u_t + R_3(\mathbf{z}_t;\boldsymbol{\xi})$ and $R_3(\mathbf{z}_t;\boldsymbol{\xi})$ is the remainder of the Taylor expansion.

The null hypothesis (7) is then approximated by

$$\mathbb{H}_0:\ \rho = 0,\ \rho_{ij} = 0,\ \rho_{ijk} = 0.$$

Under the null, $R_3(\mathbf{z}_t;\boldsymbol{\xi}) = 0$. We can use (8) to test linearity. The local approximation for observation t takes the form

$$\begin{aligned}\ell_t(\boldsymbol{\psi}) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}\left\{\frac{r_t - \boldsymbol{\beta}'\mathbf{x}_t - \lambda\left[(1-L)^d\log(\sigma_t)\right]}{\sigma_t}\right\}^2 \\ &-\frac{1}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\times\Bigg\{(1-L)^d\log(\sigma_t) - \pi_0 - \boldsymbol{\pi}'\mathbf{w}_t - \boldsymbol{\rho}'\mathbf{s}_t \\ &-\sum_{i=1}^{k_w}\sum_{j=i}^{k_w}\rho_{ij}w_{i,t}w_{j,t} - \sum_{i=1}^{k_w}\sum_{j=i}^{k_w}\sum_{k=j}^{k_w}\rho_{ijk}w_{i,t}w_{j,t}w_{k,t}\Bigg\}^2.\end{aligned} \tag{9}$$

Because the information matrix is block diagonal, the error variance σ_u^2 can be treated as fixed. The partial derivatives of (9) evaluated under the null hypothesis are:

$$
\begin{aligned}
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}}\right|_{\mathbb{H}_0} &= \widehat{e}_t \frac{\mathbf{x}_t}{\sigma_t}; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \lambda}\right|_{\mathbb{H}_0} &= \widehat{e}_t \frac{\widehat{v}_t}{\sigma_t}; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial d}\right|_{\mathbb{H}_0} &= -\frac{u_t}{\sigma_u^2}\left[\Phi(L)\left(\frac{\partial}{\partial d}(1-L)^d\right)\log h_t - \alpha'\left(\frac{\partial}{\partial d}(1-L)^d\right)x_t\right] \\
&\quad + \lambda \frac{e_t}{h_t}\left[\frac{\partial}{\partial d}(1-L)^d\right]\log h_t; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \pi_0}\right|_{\mathbb{H}_0} &= \frac{1}{\widehat{\sigma}_u^2}\widehat{u}_t; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\pi}}\right|_{\mathbb{H}_0} &= \frac{1}{\widehat{\sigma}_u^2}\widehat{u}_t \mathbf{w}_t; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\rho}}\right|_{\mathbb{H}_0} &= \frac{1}{\widehat{\sigma}_u^2}\widehat{u}_t \mathbf{s}_t; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \rho_{ij}}\right|_{\mathbb{H}_0} &= \frac{1}{\widehat{\sigma}_u^2}\widehat{u}_t w_{i,t} w_{j,t}; \\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \rho_{ijk}}\right|_{\mathbb{H}_0} &= \frac{1}{\widehat{\sigma}_u^2}\widehat{u}_t w_{i,t} w_{j,t} w_{k,t},
\end{aligned}
$$

where $\widehat{e}_t = (r_t - \widehat{\boldsymbol{\beta}}'\mathbf{x}_t - \widehat{\lambda}\widehat{v}_t)/\sigma_t$, $\widehat{u}_t = \widehat{v}_t - \widehat{\pi}_0 - \widehat{\boldsymbol{\pi}}'\mathbf{w}_t - \widehat{\boldsymbol{\rho}}'\mathbf{s}_t$ is the residual estimated under the null, $\widehat{v}_t = (1-L)^{\widehat{d}}\log(\sigma_t)$, $\widehat{\sigma}_u^2 = \sum_{t=1}^{T}\widehat{u}_t^2/T$, and

$$
\frac{\partial}{\partial d}(1-L)^d = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)\prod_{i=0}^{j-1}(d-i)L^j.
$$

Under the information matrix equality, the Lagrange Multiplier (LM) statistic is given by

$$
LM = \sum_{t=1}^{T}\widehat{\mathbf{q}}_t'\left\{\sum_{t=1}^{T}\widehat{\mathbf{q}}_t\widehat{\mathbf{q}}_t'\right\}^{-1}\sum_{t=1}^{T}\widehat{\mathbf{q}}_t, \tag{10}
$$

where $\widehat{\mathbf{q}}_t = (\widehat{\mathbf{q}}_{0,t}, \mathbf{q}_{a,t})'$ with

$$
\widehat{\mathbf{q}}_{0,t} = \left(1, \mathbf{w}_t', \left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial d}\right|_{\mathbb{H}_0}\right)'
$$

and

$$
\mathbf{q}_{a,t} = \left(w_{1,t}^2, w_{1,t}w_{2,t}, \ldots, w_{i,t}w_{j,t}, \ldots, w_{1,t}^3, \ldots, w_{i,t}w_{j,t}w_{k,t}, \ldots, w_{k_w,t}^3\right)'
$$

is the gradient of the log-likelihood function evaluated under the null as above.

Under standard regularity conditions and the additional assumption $\mathbb{E}|w_{i,t}|^{\delta} < \infty$, $i = 1, \ldots, k_w$, for some $\delta > 6$, (10) has an asymptotic χ^2 distribution with $m = k_w(k_w+1)/2 + k_w(k_w+1)(k_w+2)/6$ degrees of freedom. Defining $\boldsymbol{\iota} = (1, 1, \ldots, 1)' \in \mathbb{R}^T$ and

$$\widehat{\mathbf{Q}} = \begin{pmatrix} \widehat{q}_1' \\ \widehat{q}_2' \\ \vdots \\ \widehat{q}_T' \end{pmatrix},$$

the LM statistic can be written as

$$LM = \boldsymbol{\iota}'\widehat{\mathbf{Q}} \left(\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}}\right)^{-1} \widehat{\mathbf{Q}}'\boldsymbol{\iota}$$

and the test can be carried out in stages as follows:

(1) Estimate the parameters under the null and compute the residuals $\widehat{u}_t$ and $\widehat{e}_t$. If the sample size is small, usually the fractional difference parameter d is difficult to estimate such that the first order condition

$$\left.\frac{\delta \mathcal{L}(\boldsymbol{\psi})}{\delta \boldsymbol{\psi}}\right|_{\mathbb{H}_0} = 0$$

is not met. This has an adverse effect on the empirical size of the test. To circumvent this problem, we regress the residuals $\widehat{u}_t$ and $\widehat{e}_t$ on the derivative of the gradient with respect to $\widehat{u}_t$ and $\widehat{e}_t$. Finally, we compute a new sequence of residuals $\widetilde{u}_t$ and $\widetilde{e}_t$ from these regressions.

(2) Regress $\boldsymbol{\iota}$ on $\mathbf{Q}$ and compute the sum of squared residuals (SSR) from this regression.

(3) Compute the χ^2 statistic

$$LM_{\chi} = T - SSR.$$

3.2. **Model Selection.** The modeling cycle consists of several steps. First, it is necessary to select the variables $\mathbf{x}_t$ and $\mathbf{z}_t$. Possible choices are lags of fractionally differenced volatility, lagged values of returns and squared returns, a time index as in Example 2 or lagged values of e_t to capture asymmetries. Second, if linearity is rejected, we should choose the nonlinear function $g(\mathbf{z}_t; \boldsymbol{\xi})$. Linearity tests against specific forms of $g(\mathbf{z}_t; \boldsymbol{\xi})$, such as in Example 2 and 3, can also be developed and used to discriminate among different nonlinear alternatives. After estimating the models, diagnostic tests must be used in order to check model adequacy.

In this paper we propose the following steps for model building:

(1) Start setting $g(\mathbf{z}_t; \boldsymbol{\xi}) = \xi_0 + \boldsymbol{\xi}'\mathbf{z}_t$.

(2) The elements of $\mathbf{x}_t$ and $\mathbf{z}_t$ are selected using a given choice of information criterion, such as the AIC or SBIC.

(3) Linearity is tested for different choices of $\mathbf{w}_t$ in (6) and the one that minimizes the p-value of the test is selected as the final choice.

(4) Based on the results of the linearity tests, different nonlinear alternatives may be estimated, including a neural network.
(5) The estimated models are evaluated by a sequence of diagnostic tests and also by their forecasting performance.

3.3. **Monte-Carlo Evidence.** The purpose of this section is to evaluate the small sample performance of the linearity test described in the previous section. Set $v_t = (1-L)^{0.4}\log(\sigma_t)$ and $u_t \sim \mathsf{NID}(0, 0.25)$. To check if the test is well-sized, we simulate 1000 replications with 1000 and 200 observations of the following model:

$$r_t = \lambda v_t + \sigma_t e_t$$
$$v_t = 0.2 + 0.8 v_{t-1} + u_t,$$

We consider two cases: $\lambda = 0$ and with $\lambda = 0.5$. We run the linearity test described in Section 3 with three choices of variable specification. In the first one $\mathbf{z}_t = \mathbf{w}_t = v_{t-1}$. The second choice is $\mathbf{w}_t = v_{t-1}$ and the variable in the nonlinear specification is t. Finally, we run the test with e_{t-1} as the nonlinear variable. The results are described in Figure 1, which shows the size discrepancy plots.

FIGURE 1 ABOUT HERE

As can be observed, the size distortions are small, especially for the nominal values typically used in practical applications. We also simulated a series of nonlinear models to check the power of the test. As expected, the power converges to one very rapidly and the results are omitted for brevity.

4. Empirical Application

4.1. **Data.** We use high frequency tick-by-tick quotes on twenty eight Dow Jones Industrial Average Index stocks as listed in Table 1: Alcoa (aa), American International Group (aig), American Express (axp), Boeing (ba), Citigroup (c), Caterpillar (cat), Du Pont (dd), Walt Disney (dis), General Electric (ge), General Motors (gm), Home Depot (hd), Honeywell (hon), Hewlett Packard (hpq), International Business Machines (ibm), Johnson and Johnson (jnj), JP Morgan Chase (jpm), Coca Cola (ko), McDonald's (mcd), 3M Company (mmm), Altria Group (mo), Merck (mrk), Pfizer Inc. (pfe), Procter and Gamble (pg), AT&T (t), United Tech (utx), Verizon Communications (vz), Wal-Mart Stores (wmt) and Exxon Mobil (xom). The data were obtained from the NYSE TAQ database and they cover the period January 3, 1995 up to December 31,2005. (Exceptions are xom, vz, and hpq, for which only data from 1999 to 2005, 2000 to 2005, and 2002 to 2005 are available, respectively.)

In calculating the daily realized volatility we employ the realized kernel estimator with modified generalized Tukey-Hanning weights of order two according to Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007). We clean the data for outliers. We discard transactions outside trading hours, considering transactions between 9.30am through 4.00pm. Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007) we use a 60-second activity fixed tick time sampling scheme such that we obtain the same number of observations each day.

4.2. **Model Specification and Estimation.** In this section, we estimate model (1) and (2) employing four different specifications for the nonlinear function g. Our intention is not to be authoritative about any of these specifications being the "right" one but we intend to illustrate how the model framework can be adapted for different purposes of study. Prior to estimating the model, we estimate the long memory parameter d on the realized volatility time series using Geweke and Porter-Hudak's (1983) estimator (GPH) and using the Whittle estimator with different AR filters. Table 1 presents the results. For the GPH estimator with $\alpha = 0.65$ and for the Whittle estimator with lag order $p = 5$, roughly one-half of the time series are estimated in the non-stationary range. For the GPH estimator with the commonly used $\alpha = 0.5$, the majority of stocks is estimated to be non-stationary. For the Whittle estimate with lag structures $p = 3, 4$, the majority of stocks is estimated to be stationary with d close to 0.40.

First, we consider g to consist only of an autoregressive term, so that equation (2) is linear. This specification can capture HAR-RV models (in the case of $d = 0$) that conform with AR(p) structures as in Bollerslev, Kretschmer, Pigorsch, and Tauchen (2007). Here, we choose the autoregressive lag order p by minimizing the Akaike Information Criterion (AIC). Table 2 reports the results. Even though a linear structure is chosen for the volatility equation, the ML-estimates of d are already all in the stationary region. Caterpillar (cat) is an outlier in that it shows an estimated negative d (corresponding to anti-persistence, or negative serial correlation) consistently across all specifications. The estimates of the volatility-in-mean parameter λ are negative for the majority of the stocks. This corresponds to the leverage and feedback effect concepts. For seven stocks, however, the estimates are significantly positive, and in one case (Citigroup) the estimate of λ is not significantly different from zero. It is beyond the scope of this paper to analyze these effects more deeply, see Brandt and Kiang (2004) and the work cited therein for reference.

The next step in the model building procedure is to test for nonlinearities in the presence of long memory. We apply the test statistic proposed in equation (10) to the time series with the following candidates for transition variables: (1) returns at lag one, (2) cumulative returns over five days, (3) cumulative returns over 22 days, (4) time. Table 3 reports the p-values for the test statistic across the different candidate transition variables. All test statistics are significant at the 1% significance level at least, with the exception of time in the case of eight stocks. This strongly suggests non-linear effects that correspond to movements in the candidate transition variables.

FIGURE 2 ABOUT HERE

Motivated by the results of the test, we contrast the linear specification of equation (2) with three different nonlinear specifications of g.

(1) We consider time as transition variable so that we can analyze the realized volatility time series for structural breaks (i.e. changes in the mean of the volatility time series). The number of transition functions is chosen by minimizing AIC. Table 4 reports the estimates and Figure 2 shows the actual and fitted realized volatility series for four selected stocks together with the sum of the transition functions. Almost all estimations pick up a change in the volatility level from high to low around the year 2003, many also detect a transition from low to high volatility around the year 1998. For most time series, one or two transitions are detected. There is one instance of three and one instance of four estimated transitions. Again, the estimates of the d-parameter are all within the stationary realm and cluster around 0.40.

(2) We consider returns at lag one as transition variable. Changing volatility levels as functions of past returns correspond to the idea of leverage and/or feedback effects. The number of transitions is chosen by AIC. Table 5 reports the estimates and Figure 3 shows the actual and fitted realized volatility series for four selected stocks. For the large majority of stocks, the transition characteristics resemble the one depicted for Johnson and Johnson in panel (a) of Figure 3: Volatility is relatively low for an intermediate range of relatively small negative and positive returns and high in two different regimes of highly positive and highly negative returns. Similar findings have been reported before, for example in Medeiros and Veiga (2007). Sometimes there is no intermediate regime but the transition is smeared across two distinct regimes for large positive and large negative returns, as shown in panel (b) for General Electric. Rarely the smooth transition is interrupted by a hard one for very large positive returns (panel (c), Coca-Cola), and the case of four regimes with hard transitions has only one instance (panel (d), Procter and Gamble). The estimation findings for the long memory parameter d and the volatility-in-mean parameter λ are qualitatively similar to the case of time as transition function.

(3) We consider linear combinations of lagged returns as transition variable. The order of the linear combination as well as the number of transitions are chosen according to AIC. Each transition can be driven by a different linear combination of returns. The economic interpretation of these combinations is not immediate but Table 6 shows in the last column that the R^2 of this specification is higher, indicating a better in-sample fit. This specification is closer in spirit to the fully general nonlinear specification proposed in equation (3).

FIGURE 3 ABOUT HERE

4.3. **Forecast Exercise.** We construct a forecasting experiment using the last 600 data points as forecast sample, i.e., we exclude the last 600 points from the estimation. We consider the linear specification of model (1) and (2) as in Example 1 and in Table 2, the specification with returns at lag one as in Example 3 and in Table 5, and the general nonlinear specification as in Table 6. We compare model (1) and (2) with the HAR-RV specification of Corsi (2004) and Andersen, Bollerslev, and Diebold (2007).

We employ the realized volatility version of the HAR-RV model (as opposed to the logarithmic realized volatility version) with lags 1, 5, and 22. The R^2 from Mincer-Zarnowitz regressions is reported in Table 7 for the four specifications. The numbers in parentheses report the p-value for the Superior Predictive Ability test of Hansen (2005). A low p-value indicates that the model was significantly outperformed by at least one other model in the set. At a ten percent significance level, the benchmark HAR-RV model could not be rejected for 12 out of 28 stocks. The linear specification of (1) and (2) could not be rejected for 26 out of 28 stocks. This may seem surprising; note that the linear specification still features fractional integration. The specification with past returns at lag one could not be rejected in 19 of 28 cases; the general nonlinear specification with AIC-selected linear combinations of lags of past returns could not be rejected in 20 out of 28 cases. We can conclude that accounting for nonlinear terms in the presence of long memory yields forecast advantages in our sample.

5. Conclusion

In financial volatility, nonlinearities such as structural breaks are difficult to tell apart from long memory. In this paper, we propose an estimation framework for nonlinear effects such as structural breaks and leverage in the presence of long memory. The framework specifies a nonlinear system of equations for returns and realized volatility that accommodates volatility-in-mean effects, long memory, and a general non-linear function that may include transitions between parameter regimes and leverage effects.

From the model specification, we derive a test statistic that allows to test for nonlinear terms in the volatility equation in the presence of long memory. The test evaluates the significance of second and higher order terms in the Taylor expansion of the nonlinear function in the volatility equation.

Once the type of nonlinearity and the relevant variables are identified, a process for which we propose a model selection cycle, the full specification is estimated using the stocks in the Dow Jones Industrial Average. We find strong evidence for nonlinear effects driven by time and past returns in all stocks. The results indicate that long memory, changes in the mean, and leverage effects in a wide sense, i.e. dependence on linear combinations of past returns, coexist in realized volatility data. In accordance with earlier findings, the long memory parameter estimates are reduced once the nonlinear effects are accounted for.

Acknowledgments

We gratefully acknowledge comments from Jean-Pierre Fouque, Peter Reinhard Hansen, Jim Hilliard, Hagen Kim, Tae-Hwy Lee, Morten Nielsen, Joon Park, participants at the conference "Breaks and Persistence in Econometrics" at the Cass Business School London in December 2006, at the Midwest Econometrics Meetings in St. Louis in October 2007, and at seminars at LSU and Texas A&M. Parts of the research for this paper were done while the first author was visiting the Department of Economics at the Pontifical Catholic University and the Ibmec Business School in Rio de Janeiro, Brazil. Their hospitality is gratefully appreciated. The research of the second author is partially supported by the CNPq/Brazil. Mihaela Craioveanu provided excellent research assistance. An earlier draft of the paper was circulated under the title "Asymmetries, Breaks, and Long-Range Dependence in Realized Volatility: A Simultaneous Equations Approach."

TABLE 1. FRACTIONAL DIFFERENCE ESTIMATION.

Estimation results for parameter d. The Whittle estimator is considered under different autoregressive orders (p). The number of ordinates in the GPH estimator is set as $l = T^{\alpha}$, where T is the number of observations.

Series	GPH ($\alpha = 0.5$)	GPH ($\alpha = 0.65$)	Whittle ($p = 3$)	Whittle ($p = 4$)	Whittle ($p = 5$)
aa	0.5416	0.5183	0.3870	0.4475	0.4887
aig	0.6054	0.5683	0.4292	0.4928	0.5226
axp	0.6079	0.5498	0.4353	0.4807	0.5281
ba	0.5302	0.4630	0.3697	0.4060	0.4511
c	0.5815	0.5269	0.4326	0.4916	0.5155
cat	0.5770	0.3910	0.3383	0.3746	0.4374
dd	0.5432	0.4907	0.3966	0.4518	0.4923
dis	0.6139	0.4867	0.4118	0.4567	0.4938
ge	0.5371	0.4606	0.4277	0.4771	0.5109
gm	0.5911	0.5733	0.3717	0.4430	0.4838
hd	0.4112	0.4660	0.3976	0.4575	0.4780
hon	0.4234	0.5137	0.3693	0.4274	0.4644
hpq	0.7505	0.5291	0.4853	0.5326	0.5480
ibm	0.6990	0.4805	0.4206	0.4800	0.5270
jnj	0.6342	0.5240	0.4019	0.4603	0.4838
jpm	0.5909	0.5133	0.4319	0.4767	0.5292
ko	0.5070	0.5152	0.4105	0.4735	0.4913
mcd	0.5749	0.4472	0.3498	0.4032	0.4471
mmm	0.4838	0.4905	0.3980	0.4465	0.4704
mo	0.5081	0.4486	0.3610	0.4018	0.4242
mrk	0.4349	0.4423	0.3609	0.4089	0.4669
pfe	0.4961	0.4362	0.3435	0.3967	0.4464
pg	0.4459	0.5105	0.4244	0.4732	0.5238
t	0.4411	0.4173	0.3987	0.4388	0.4849
utx	0.5446	0.4490	0.4171	0.4725	0.5172
vz	0.8258	0.6441	0.4768	0.5442	0.5818
wmt	0.5711	0.4806	0.4116	0.4592	0.4819
xom	0.6627	0.6805	0.5041	0.5684	0.6224

TABLE 2. LINEAR ESTIMATION RESULTS.

Estimates of the optimal autoregressive order of the volatility equation selected by the AIC, parameter estimates (standard errors in parentheses), and R^2.

Series	p	β	λ	d	R^2
aa	4	−0.0623 (0.0230)	−0.1254 (0.0139)	0.4873 (0.0244)	0.7644
aig	3	−0.0096 (0.0150)	−0.1342 (0.0079)	0.4368 (0.0262)	0.6736
axp	4	−0.0463 (0.0100)	−0.2305 (0.0042)	0.4722 (0.0245)	0.6629
ba	4	−0.0014 (0.0168)	−0.1373 (0.0086)	0.4042 (0.0271)	0.7043
c	3	0.0817 (0.0120)	0.0038 (0.0063)	0.4842 (0.0260)	0.6763
cat	4	−0.2394 (0.0079)	0.4541 (0.0019)	−0.3097 (0.0013)	0.6729
dd	4	0.1096 (0.0186)	0.1212 (0.0130)	0.4200 (0.0238)	0.7296
dis	4	0.1043 (0.0182)	0.1051 (0.0085)	0.4578 (0.0270)	0.7153
ge	5	0.0342 (0.0126)	−0.1732 (0.0075)	0.4208 (0.0274)	0.6906
gm	3	−0.0651 (0.0188)	−0.1381 (0.0118)	0.4344 (0.0262)	0.6602
hd	5	0.0796 (0.0207)	−0.1156 (0.0069)	0.4151 (0.0293)	0.6965
hon	5	0.0049 (0.0183)	−0.0786 (0.0090)	0.4321 (0.0270)	0.5992
hpq	3	0.1052 (0.0564)	0.1786 (0.0433)	0.4257 (0.0368)	0.7888
ibm	3	0.0377 (0.0146)	−0.0071 (0.0104)	0.4345 (0.0262)	0.7104
jnj	5	0.0021 (0.0114)	−0.0112 (0.0120)	0.4294 (0.0328)	0.7360
jpm	5	−0.1445 (0.0087)	−0.2381 (0.0022)	0.4723 (0.0091)	0.6840
ko	5	0.0584 (0.0111)	−0.0346 (0.0056)	0.4903 (0.0331)	0.7411
mcd	4	0.0869 (0.0204)	−0.0455 (0.0125)	0.4305 (0.0291)	0.6861
mmm	2	0.0631 (0.0145)	0.0806 (0.0119)	0.4050 (0.0220)	0.7118
mo	5	0.0917 (0.0138)	−0.0710 (0.0096)	0.4561 (0.0343)	0.5801
mrk	5	0.0064 (0.0066)	0.0440 (0.0022)	0.4234 (0.0340)	0.6314
pfe	5	−0.1015 (0.0140)	−0.3205 (0.0038)	0.4813 (0.0200)	0.6877
pg	4	0.0931 (0.0101)	−0.0300 (0.0033)	0.4554 (0.0273)	0.7022
t	4	0.0463 (0.0147)	0.0542 (0.0101)	0.4291 (0.0281)	0.6597
utx	4	0.1073 (0.0038)	−0.0101 (0.0014)	0.4724 (0.0250)	0.7194
vz	3	−0.0751 (0.0306)	−0.0795 (0.0339)	0.4345 (0.0282)	0.7544
wmt	5	−0.0019 (0.0123)	−0.0312 (0.0058)	0.4881 (0.0254)	0.7106
xom	2	0.1337 (0.0196)	−0.2062 (0.0215)	0.4040 (0.0263)	0.7383

TABLE 3. LINEARITY TEST.

Results of the linearity test: p-value for different choices of transition variables.

Series	past return	past 5-day returns	past 22-day returns	time
aa	$7.4614e-011$	$7.7716e-016$	$1.7386e-013$	$5.7236e-002$
aig	$4.4409e-016$	0	$2.7756e-014$	$8.9596e-003$
axp	$2.0306e-013$	0	0	$6.3512e-006$
ba	0	0	0	$6.2752e-007$
c	$1.6029e-011$	0	0	$2.3226e-007$
cat	0	0	0	$6.9072e-002$
dd	$2.7756e-015$	0	$2.2204e-016$	$3.9567e-007$
dis	$1.9916e-011$	$2.5757e-014$	$3.1040e-012$	$7.2862e-006$
ge	0	0	0	$7.8109e-009$
gm	$7.2018e-012$	$9.4036e-014$	$1.3334e-013$	$8.5783e-001$
hd	0	0	0	$1.7102e-005$
hon	$2.6860e-011$	0	0	$6.5725e-004$
hpq	$8.4052e-009$	$1.3436e-007$	$4.2223e-006$	$5.8328e-002$
ibm	$1.1102e-016$	0	$5.0671e-013$	$5.8471e-007$
jnj	0	0	$2.6079e-013$	$1.3064e-005$
jpm	$1.1102e-016$	0	0	$1.4548e-006$
ko	$1.4166e-013$	0	$5.5753e-011$	$8.5256e-005$
mcd	$1.3819e-010$	$6.7724e-014$	$4.9084e-012$	$1.6919e-002$
mmm	$3.7597e-011$	$1.3460e-010$	$5.5536e-010$	$1.2813e-005$
mo	$1.5848e-007$	$3.4539e-013$	$3.3347e-006$	$1.9979e-002$
mrk	$1.2318e-009$	0	$3.2601e-009$	$1.2510e-001$
pfe	$2.0837e-007$	$7.3617e-009$	$7.5223e-008$	$2.8729e-002$
pg	$7.9030e-010$	0	$1.4207e-010$	$1.7706e-003$
t	$3.3307e-016$	$9.9920e-016$	$1.6742e-013$	$5.8954e-007$
utx	$2.8315e-010$	$3.4417e-015$	$3.4149e-012$	$4.3828e-005$
vz	$2.3395e-007$	$8.8299e-010$	$8.4433e-010$	$4.5696e-005$
wmt	$3.2361e-011$	0	$8.4714e-011$	$4.1408e-005$
xom	$2.4691e-010$	$3.5527e-015$	$5.5971e-011$	$5.7165e-003$

TABLE 4. NONLINEAR ESTIMATION RESULTS. TRANSITION VARIABLE: TIME.

Parameters estimates (standard errors in parentheses). The number M of transitions is chosen by AIC.

Series	β	λ	d	M	R^2
aa	-0.0635 (0.0316)	-0.1257 (0.0245)	0.4665 (0.0347)	2	0.7349
aig	-0.0100 (0.0518)	-0.1345 (0.0368)	0.4246 (0.0462)	2	0.6903
axp	-0.0450 (0.0179)	-0.2306 (0.0080)	0.4856 (0.0355)	2	0.6725
ba	-0.0014 (0.0248)	-0.1373 (0.0177)	0.4029 (0.0387)	2	0.7058
c	0.0818 (0.0155)	0.0039 (0.0099)	0.4728 (0.0342)	1	0.6836
cat	-0.2284 (0.0218)	0.4568 (0.0082)	-0.3071 (0.0024)	1	0.6705
dd	0.1105 (0.0232)	0.1219 (0.0250)	0.4044 (0.0285)	2	0.7371
dis	0.1062 (0.0263)	0.1064 (0.0268)	0.3960 (0.0360)	3	0.6992
ge	0.0348 (0.0336)	-0.1704 (0.1813)	0.4040 (0.0342)	2	0.6919
gm	-0.0631 (0.0627)	-0.1349 (0.1419)	0.4109 (0.0642)	1	0.6590
hd	0.0815 (0.0675)	-0.1145 (0.0374)	0.4742 (0.0354)	1	0.7057
hon	0.0044 (0.0252)	-0.0776 (0.0098)	0.3772 (0.0350)	2	0.6070
hpq	0.1074 (0.0556)	0.1776 (0.0518)	0.1547 (0.0943)	1	0.8044
ibm	0.0374 (0.0205)	-0.0079 (0.0351)	0.4498 (0.3602)	2	0.7184
jnj	0.0028 (0.0260)	-0.0098 (0.0806)	0.4038 (0.0323)	1	0.7235
jpm	-0.1491 (0.0808)	-0.2392 (0.0191)	0.4097 (0.0253)	1	0.6858
ko	0.0581 (0.0169)	-0.0342 (0.0071)	0.4220 (0.0263)	2	0.7483
mcd	0.0873 (0.0250)	-0.0452 (0.0180)	0.4401 (0.0301)	2	0.6813
mmm	0.0605 (0.0205)	0.0780 (0.0228)	0.4469 (0.0232)	1	0.7207
mo	0.0919 (0.0210)	-0.0714 (0.0212)	0.4785 (0.0349)	1	0.5585
mrk	0.0912 (0.0189)	0.0777 (0.0067)	-0.4046 (0.0195)	2	0.6022
pfe	-0.1085 (0.0221)	-0.3231 (0.0060)	0.3950 (0.0363)	2	0.6910
pg	0.0934 (0.0177)	-0.0296 (0.0109)	0.4221 (0.0452)	1	0.7028
t	0.0462 (0.0244)	0.0542 (0.0259)	0.4341 (0.0300)	2	0.6713
utx	0.1032 (0.0167)	-0.0121 (0.0078)	0.3980 (0.0127)	4	0.7250
vz	-0.0744 (0.0529)	-0.0764 (0.1452)	0.4058 (0.0243)	1	0.7743
wmt	0.0074 (0.1419)	-0.0214 (0.1496)	0.4206 (0.0435)	2	0.7189
xom	0.1359 (0.1000)	-0.2054 (0.0827)	0.4346 (0.0332)	1	0.7506

TABLE 5. NONLINEAR ESTIMATION RESULTS. TRANSITION VARIABLE: PAST RETURNS (LEVERAGE).

Parameters estimates (standard errors in parentheses). The number M of transitions is chosen by AIC.

Series	β	λ	d	M	R^2
aa	-0.0651 (0.0524)	-0.1260 (0.0281)	0.4346 (0.0453)	2	0.7615
aig	-0.0084 (0.0204)	-0.1332 (0.0299)	0.4719 (0.0534)	3	0.6286
axp	-0.0533 (0.1139)	-0.2302 (0.0110)	0.4056 (0.0364)	1	0.6663
ba	-0.0008 (0.0436)	-0.1385 (0.0179)	0.4742 (0.0463)	2	0.7158
c	0.0819 (0.0222)	0.0040 (0.0205)	0.4650 (0.3632)	1	0.6714
cat	-0.2394 (0.0362)	0.4542 (0.0111)	-0.3097 (0.0069)	1	0.6712
dd	0.1093 (0.0779)	0.1210 (0.0700)	0.4246 (0.0364)	1	0.7393
dis	0.1053 (0.0340)	0.1059 (0.0352)	0.4262 (0.2508)	2	0.7137
ge	0.0338 (0.0442)	-0.1753 (0.0520)	0.4346 (0.0464)	1	0.7314
gm	-0.0687 (0.0622)	-0.1441 (0.0824)	0.4841 (0.0374)	2	0.6640
hd	0.0810 (0.0775)	-0.1148 (0.0434)	0.4598 (0.0453)	2	0.7285
hon	0.0053 (0.0298)	-0.0790 (0.0163)	0.4583 (0.2172)	1	0.6118
hpq	0.1019 (0.0720)	0.1768 (0.0573)	0.4631 (0.0365)	2	0.8053
ibm	0.0384 (0.0867)	-0.0054 (0.0304)	0.4058 (0.0464)	2	0.7105
jnj	0.0021 (0.0418)	-0.0112 (0.0588)	0.4299 (0.0453)	2	0.7439
jpm	-0.1437 (0.0781)	-0.2379 (0.0226)	0.4829 (0.0325)	2	0.7017
ko	0.0581 (0.0169)	-0.0341 (0.0055)	0.4018 (0.0855)	2	0.7589
mcd	0.0883 (0.0598)	-0.0443 (0.0480)	0.4648 (0.0235)	2	0.6959
mmm	0.0601 (0.0771)	0.0776 (0.0624)	0.4529 (0.0354)	2	0.7246
mo	0.0918 (0.0213)	-0.0711 (0.0436)	0.4620 (0.0364)	2	0.4739
mrk	0.0113 (0.0202)	0.0467 (0.0102)	0.3968 (0.0405)	2	0.6346
pfe	-0.1026 (0.0873)	-0.3209 (0.0315)	0.4688 (0.0865)	2	0.6864
pg	0.0929 (0.0182)	-0.0302 (0.0141)	0.4710 (0.0346)	3	0.6811
t	0.0463 (0.0340)	0.0542 (0.0612)	0.4279 (0.0342)	1	0.6873
utx	0.1078 (0.0901)	-0.0098 (0.0449)	0.4892 (0.0353)	2	0.7285
vz	-0.0756 (0.0340)	-0.0827 (0.0616)	0.4687 (0.1630)	2	0.7593
wmt	0.0027 (0.1486)	-0.0263 (0.1522)	0.4529 (0.0342)	1	0.7185
xom	0.1347 (0.0281)	-0.2059 (0.0465)	0.4175 (0.0319)	1	0.7711

TABLE 6. NONLINEAR ESTIMATION RESULTS. TRANSITION VARIABLE: AIC-CHOSEN LINEAR COMBINATIONS OF LAGGED RETURNS.

Parameters estimates (standard errors in parentheses). The number M of transitions is chosen by AIC.

Series	β	λ	d	M	R^2
aa	-0.0651 (0.0316)	-0.1260 (0.0245)	0.4356 (0.0268)	6	0.7865
aig	-0.0106 (0.0206)	-0.1350 (0.0099)	0.4050 (0.0280)	1	0.7149
axp	-0.0475 (0.0174)	-0.2305 (0.0081)	0.4607 (0.0240)	3	0.7167
ba	-0.0009 (0.0163)	-0.1384 (0.0116)	0.4657 (0.0234)	6	0.7484
c	0.0821 (0.0198)	0.0041 (0.0106)	0.4521 (0.0214)	7	0.7144
cat	-0.2407 (0.0361)	0.4538 (0.0112)	-0.3100 (0.0068)	4	0.6655
dd	0.1056 (0.0247)	0.1176 (0.0257)	0.4889 (0.0255)	2	0.7418
dis	0.1033 (0.0258)	0.1043 (0.0265)	0.4868 (0.0273)	1	0.7323
ge	0.0331 (0.0201)	-0.1828 (0.0239)	0.4957 (0.0290)	2	0.7219
gm	-0.0633 (0.0275)	-0.1351 (0.0291)	0.4123 (0.0365)	3	0.6957
hd	0.0797 (0.0275)	-0.1156 (0.0129)	0.4182 (0.0275)	3	0.7490
hon	0.0044 (0.0253)	-0.0777 (0.0096)	0.3830 (0.0259)	5	0.6267
hpq	0.1071 (0.0575)	0.1794 (0.0596)	0.4020 (0.0449)	2	0.7971
ibm	0.0373 (0.0224)	-0.0080 (0.0319)	0.4519 (0.0245)	2	0.6631
jnj	0.0017 (0.0163)	-0.0121 (0.0034)	0.4511 (0.0335)	3	0.7629
jpm	-0.1472 (0.0166)	-0.2388 (0.0048)	0.4376 (0.0249)	2	0.6789
ko	0.0581 (0.0165)	-0.0341 (0.0031)	0.4115 (0.0295)	1	0.7569
mcd	0.0873 (0.0779)	-0.0452 (0.0587)	0.4415 (0.9295)	3	0.6930
mmm	0.0619 (0.0210)	0.0794 (0.0226)	0.4236 (0.0215)	2	0.7286
mo	0.0916 (0.0214)	-0.0697 (0.0214)	0.4151 (0.0295)	6	0.6237
mrk	0.0798 (0.0184)	0.0723 (0.0063)	-0.4230 (0.0187)	4	0.5957
pfe	-0.1046 (0.0234)	-0.3217 (0.0063)	0.4455 (0.0330)	2	0.6933
pg	0.0934 (0.0191)	-0.0296 (0.0146)	0.4206 (0.0352)	1	0.7205
t	0.0462 (0.0239)	0.0542 (0.0269)	0.4299 (0.0260)	2	0.6733
utx	0.1055 (0.0166)	-0.0110 (0.0078)	0.4331 (0.0267)	3	0.7491
vz	-0.0746 (0.0332)	-0.0774 (0.0518)	0.4144 (0.0321)	1	0.7878
wmt	0.0079 (0.0224)	-0.0210 (0.0165)	0.4176 (0.0288)	4	0.7224
xom	0.1372 (0.0799)	-0.2048 (0.0623)	0.4549 (0.0433)	2	0.7873

TABLE 7. FORECASTING RESULTS.

Numbers in parenthesis are p-values of the Superior Predictive Ability (SPA) test proposed by Hansen (2005).

Series	linear	leverage	fully nonlinear	HAR
aa	0.8193 (0.0835)	0.8217 (0.0705)	0.8339 (0.5885)	0.8182 (0.0450)
aig	0.4786 (0.1040)	0.4878 (0.1195)	0.5552 (0.5450)	0.4789 (0.1325)
axp	0.6476 (0.7330)	0.6480 (0.8190)	0.6217 (0.0060)	0.6439 (0.1345)
ba	0.8269 (0.8570)	0.5740 (0)	0.8267 (0.7355)	0.8249 (0.0685)
c	0.8435 (0.9190)	0.8221 (0.0305)	0.8242 (0.0030)	0.8399 (0.0120)
cat	0.7053 (0.8775)	0.7047 (0.4790)	0.6934 (0.0345)	0.7031 (0.4590)
dd	0.8133 (0.7545)	0.8128 (0.6030)	0.8139 (0.6515)	0.8129 (0.6965)
dis	0.7924 (0.0720)	0.7985 (0.6775)	0.7918 (0.0625)	0.7889 (0.0170)
ge	0.8244 (0.1245)	0.7757 (0)	0.8302 (0.7405)	0.8220 (0.0710)
gm	0.6130 (0.1965)	0.6172 (0.2055)	0.6889 (0.5680)	0.6184 (0.1730)
hd	0.7866 (0.5820)	0.7670 (0)	0.7908 (0.6860)	0.7842 (0.1375)
hon	0.5986 (0.8845)	0.5968 (0.5350)	0.5966 (0.5485)	0.5934 (0.0950)
hpq	0.7668 (0.3220)	0.7675 (0.1430)	0.7718 (0.7140)	0.7599 (0.0245)
ibm	0.8264 (0.4985)	0.8306 (0.6950)	0.8045 (0.0065)	0.8228 (0.0885)
jnj	0.6617 (0.1840)	0.6708 (0.7055)	0.6614 (0.1075)	0.6575 (0.0830)
jpm	0.7785 (0.7130)	0.7805 (0.6635)	0.7790 (0.5740)	0.7748 (0.0390)
ko	0.8092 (0.8405)	0.8004 (0.0200)	0.8032 (0.1550)	0.8065 (0.0420)
mcd	0.7264 (0.2580)	0.7302 (0.8695)	0.6713 (0)	0.7259 (0.1895)
mmm	0.7673 (0.3585)	0.7691 (0.2945)	0.7725 (0.6685)	0.7621 (0.0005)
mo	0.4155 (0.7575)	0.3698 (0.0140)	0.4233 (0.6775)	0.4151 (0.6370)
mrk	0.3162 (0.5620)	0.3144 (0.3610)	0.3057 (0.0245)	0.3177 (0.6765)
pfe	0.5800 (0.9235)	0.5756 (0.1650)	0.5776 (0.3745)	0.5801 (0.6650)
pg	0.7768 (0.7105)	0.7439 (0.0005)	0.7780 (0.7200)	0.7751 (0.4185)
t	0.7274 (0.2040)	0.7230 (0.1790)	0.7344 (0.7705)	0.7247 (0.0885)
utx	0.8032 (0.7145)	0.8045 (0.6645)	0.7954 (0.0500)	0.8007 (0.0830)
vz	0.7651 (0.9215)	0.7647 (0.6105)	0.7568 (0.1610)	0.7595 (0.0270)
wmt	0.8412 (0.8740)	0.8351 (0.0700)	0.8403 (0.5220)	0.8382 (0.0995)
xom	0.7951 (0.1370)	0.7973 (0.3605)	0.8033 (0.8055)	0.8006 (0.6210)

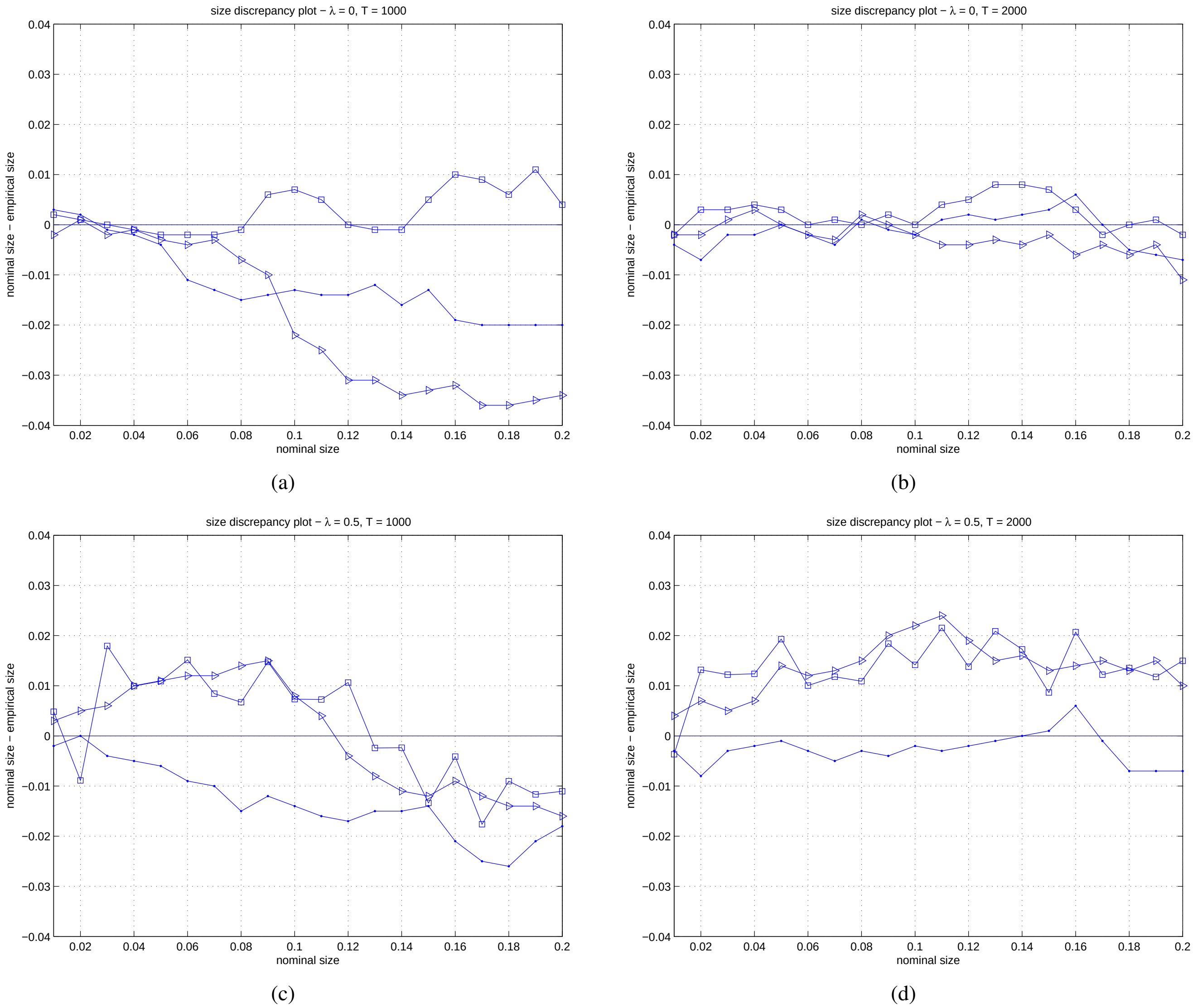
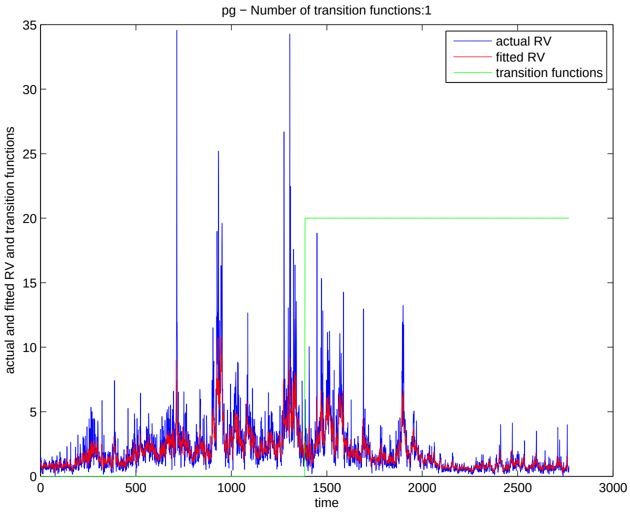
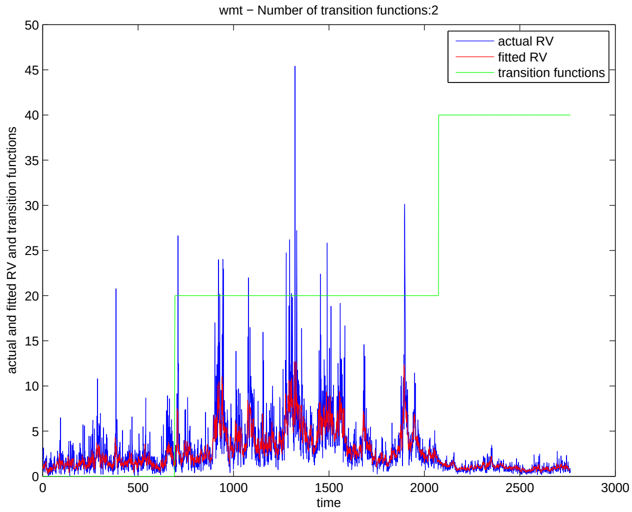


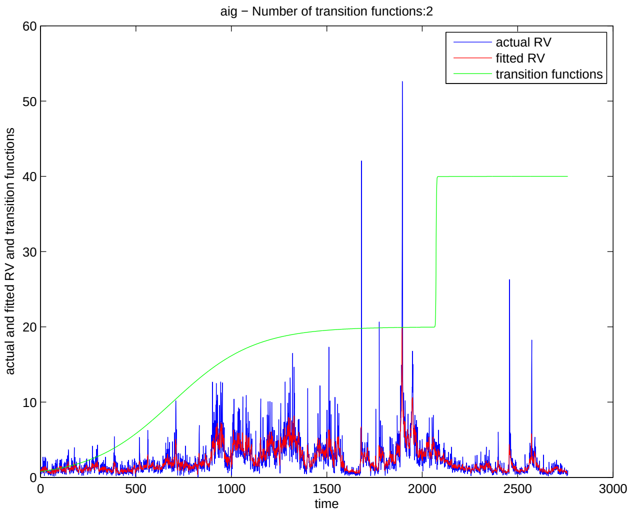
FIGURE 1. Size discrepancy curves of the linearity test. The dots refer to v_{t-1} as nonlinear variable, triangles refer to t as nonlinear variable, and squares refer to e_{t-1} as nonlinear variable. Panel (a): $\lambda = 0$ and $T = 1000$. Panel (b): $\lambda = 0$ and $T = 2000$. Panel (c): $\lambda = 0.5$ and $T = 1000$. Panel (d): $\lambda = 0.5$ and $T = 2000$.



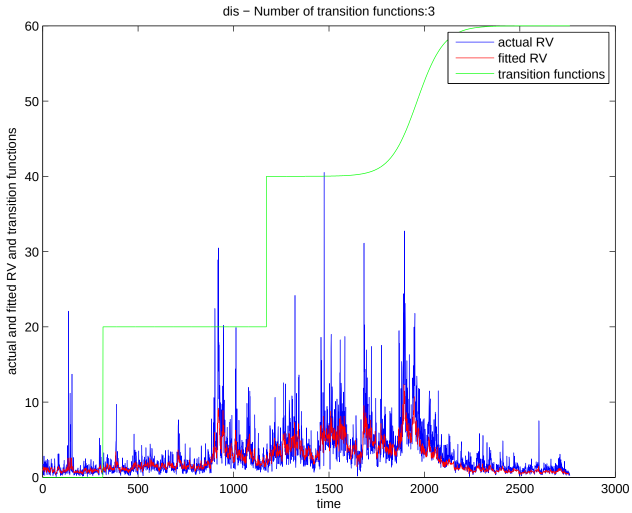
(a)



(b)

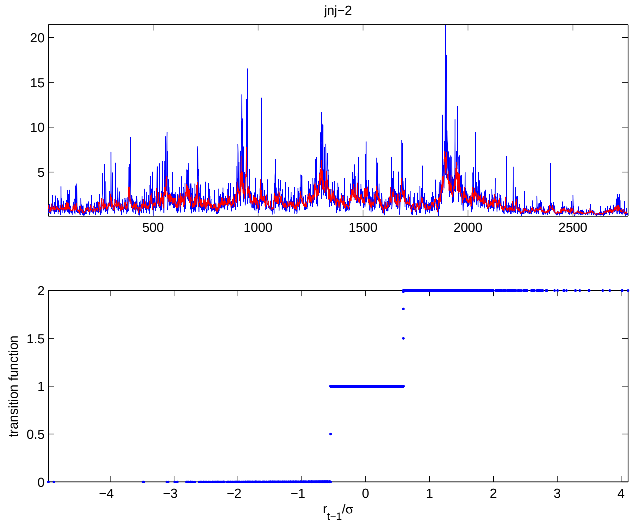


(c)

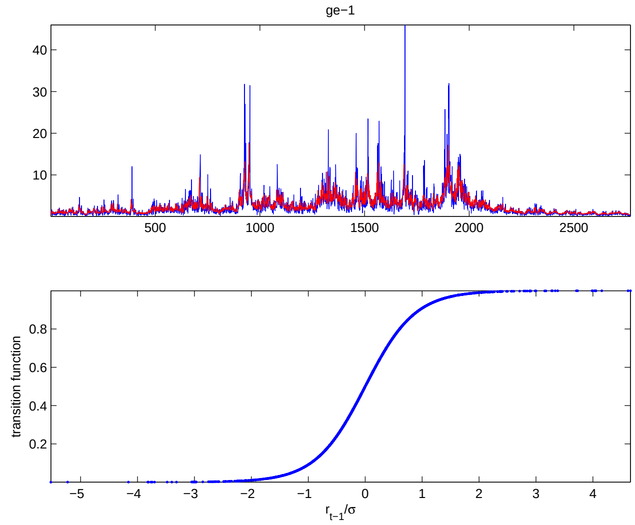


(d)

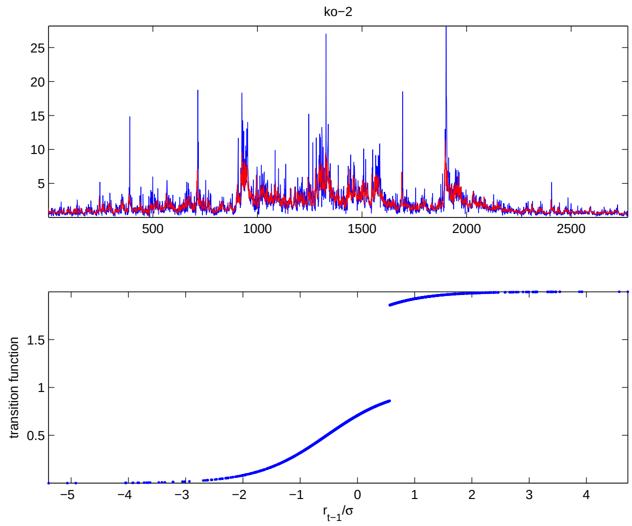
FIGURE 2. Nonlinear specifications with time as transition variable. The sum of the transition functions (scaled up for clarity) is plotted as solid line. Panel (a): Procter and Gamble, single hard transition from high to low volatility regime. Panel (b): Walmart, two hard transitions from low to high and back to low volatility regime. Panel (c): American International Group, one smooth and one hard transition from low to high and back to low volatility. Panel (d): Walt Disney, two hard transitions from low to high to higher, then one smooth transition to low volatility.



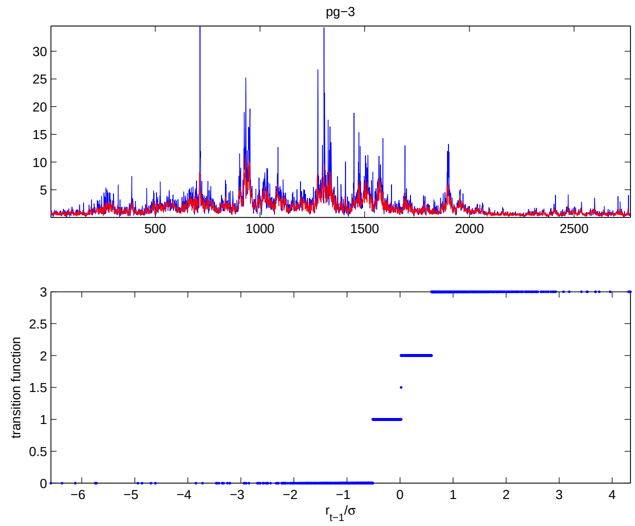
(a)



(b)



(c)



(d)

FIGURE 3. Nonlinear specifications with returns at lag one as transition variable. The sum of the transition functions is plotted as solid line in the bottom panels. Panel (a): Johnson and Johnson, three volatility regimes according to large negative, small, and large positive returns with two hard transitions. Panel (b): General Electric, two regimes for negative and positive returns with one smooth transition. Panel (c): Coca-Cola, one smooth and one hard transition. The smooth transition captures the leverage effect: the lower the return, the higher the volatility. However, for very large positive returns, volatility makes an upward jump. Panel (d): Procter and Gamble, three hard transitions. Here the intermediate range of returns is split into two sub-regimes with higher volatility for small negative returns than for small positive returns.

APPENDIX A. REALIZED VOLATILITY

Suppose that at day t the logarithmic price p of a given asset at time $t+\tau$ follows a continuous time diffusion:

$$dp(t+\tau) = \mu(t+\tau) + \sigma(t+\tau)dW(t+\tau),\ 0 \leq \tau \leq 1,\ t = 1, 2, 3, \ldots,$$

where $\mu(t+\tau)$ is the drift component, $\sigma(t+\tau)$ is the instantaneous volatility (or standard deviation), and $W(t+\tau)$ is standard Brownian motion.

Andersen, Bollerslev, Diebold, and Labys (2001a), Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002), among others, consider daily compound returns $r_t = p(t) - p(t-1)$ conditioned on $\mathcal{F}_t = \sigma(p(s),\ s \leq t)$, the σ-algebra (information set) generated by the sample paths of p. They define

$$\int_0^1 \sigma^2(t-1+\tau)d\tau = \text{Var}(r_t|\mathcal{F}(\{\mu_\theta\}_{\theta\leq t},\ \{\sigma_\theta\}_{\theta\leq t})) + \eta_t, \tag{11}$$

or *integrated variance* as the object of interest in realized volatility theory. The set $\mathcal{F}(\{\mu_\theta\}_{\theta\leq t},\ \{\sigma_\theta\}_{\theta\leq t})$ is the sigma-algebra generated by the sample paths of the drift and diffusion processes μ_s and σ_s for s up to time t, but *not* by the Brownian motion W_t that constitutes the randomness in the return equation. Thus r_t conditioned on $\mathcal{F}(\{\mu_\theta\}_{\theta\leq t},\ \{\sigma_\theta\}_{\theta\leq t})$ remains a random variable. The error η_t has zero mean. Integrated variance is a measure of the day-t ex post volatility. Compare with equation (11a) in Andersen, Bollerslev, Diebold, and Labys (2001a).

In practical applications, prices are observed at discrete and irregularly spaced intervals and there are many ways to sample the data. Suppose that at a given day t, we partition the interval [0,1] in subintervals and define the grid of observation times $\mathcal{G} = \{\tau_1, \ldots, \tau_n\}$, $0 = \tau_0 < \tau_1 < \cdots, \tau_n = 1$. The length of the ith subinterval is given by $\delta_i = \tau_i - \tau_{i-1}$. The most widely used sampling scheme is calendar time sampling (CTS), where the intervals are equidistant in calendar time, that is $\delta_i = 1/n$. Set $S_{i,t}$, $t = 1, \ldots, n$, to be the ith price observation during day t, such that $r_{t,i} = p_{t,i} - p_{t,i-1}$ is the ith intra-period return of day t. Realized variance is defined as

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2. \tag{12}$$

Realized volatility is the square-root of RV_t.

Under additional regularity conditions including the assumption of uncorrelated intraday returns, realized variance is a consistent estimator of integrated variance, such that $RV_t \xrightarrow{p} IV_t$. When returns are correlated, however, realized volatility will be a biased estimator of daily volatility. Serial correlation may be the result of market microstructure (Campbell, Lo, and MacKinlay 1997, Chapter 3). In light of these results, our assumption 2 can be criticized as falling short of the true situation where σ_t is replaced by a consistent estimator. The asymptotic theory for the fully general setup with realized volatility time series $\sigma_t + \eta_t$ would be substantially more complex.

The effects of microstructure and optimal sampling of intraday returns have been discussed in, for example, Bandi and Russell (2005a, 2005b, 2006), Barndorff-Nielsen and Shephard (2002), Meddahi (2002), Oomen (2005), Zhang, Mykland, and Aït-Sahalia (2005), Hansen and Lunde (2006), among others. Consistent estimators of realized volatility for a single day in the presence of microstructure noise are developed in Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007), and Hansen, Large, and Lunde (2007).

Several salient features of realized volatility have been identified in the literature: The unconditional distribution of daily returns exhibits excess kurtosis. Daily returns are not autocorrelated (except for the first order in some cases). Daily returns standardized by the realized variance measure are almost Gaussian. The unconditional distribution of realized variance and volatility is distinctly non-normal and extremely right-skewed. On the other hand, the natural logarithm of volatility is close to normality. The log of realized volatility displays a high degree of (positive) autocorrelation, which dies out very slowly. Finally, realized volatility does not seem to have a unit root, but there is strong evidence of fractional integration; see McAleer and Medeiros (2006) for a recent survey.

APPENDIX B. TIME SERIES MODELS FOR REALIZED VOLATILITY AS DATA-GENERATING PROCESSES

The interpretation of time series models for realized volatility as data-generating processes may not appear straightforward. In this paper, we advocate model (1), (2) as an estimation model that captures features that can be produced by stochastic volatility diffusion models that generate the intraday returns. For example, long memory may be the result of stochastic volatility that is driven by the sum of several Ornstein-Uhlenbeck processes of different decorrelation lengths (LeBaron 2001, Fouque, Papanicolaou, and Sircar 2000). We do not advocate our model as data-generating process because the fractional difference operator would imply convergence to fractional Brownian motion in the volatility equation in the continuous time limit. We rather understand fractional integration as an abbreviation for aggregation over several Ornstein-Uhlenbeck processes.

It is possible, however, to interpret model (1), (2) as data-generating process. The only complication involved is the correct scaling of the parameters and of σ_t. Write the model as

$$r_{t+s} = x'_{t+s}\beta_M + \lambda_M \sigma_{t+s} + h(M)\sigma_{t+s}e_{t+s},$$
$$\sigma_{t+s} = f\left(\{\sigma_\theta\}_{\theta<t+s};\, \boldsymbol{\xi}_M\right), \quad s \in \left\{\frac{0}{M}, \frac{1}{M}, \ldots, \frac{M}{M}\right\}, t = 0, \ldots, T-1,$$

where a couple differences to the notation in (1), (2) have been introduced: M is the number of intraday observations, s denotes the intraday time indicator and M as subscript indicates the dependence of parameters on the intraday time scale, f is a general volatility function that depends on the history of σ_t, may contain stochastic elements and is parameterized by $\boldsymbol{\xi}_M$. The function $h(M)$ is a deterministic scaling factor that depends on the number M of intraday observations and on the scaling of σ_t. The numerical values of the parameters β_M, λ_M, and so forth depend on the intraday scale as well.

EXAMPLE 6. *Consider the following two cases:*

(1) *HAR-RV*

$$f\left(\{\sigma_\theta\}_{\theta<t+s};\, \boldsymbol{\xi}_M\right) = c + \alpha_1\sigma_{t-1+s} + \frac{\alpha_2}{5}\sum_{j=1}^{5}\sigma_{t-j+s} + \frac{\alpha_3}{21}\sum_{j=1}^{21}\sigma_{t-j+s} + u_{t+s},\ u_t \sim\ \mathit{WN}(0,\sigma_u^2).$$

Here, the parameters in $\boldsymbol{\xi}_M$ *are* $(c, \alpha_1, \alpha_2, \alpha_3, \sigma_u^2)$ *and can be chosen such that* σ_t *scales to any desired level of magnitude, say annualized volatility with a realistic numerical magnitude of* 10^{-1} *for stock market data.*

(2) *Model* (2)

$$f\left(\{\sigma_\theta\}_{\theta<t+s};\, \boldsymbol{\xi}_M\right) = \exp\left\{(1-L)^{-d}\left[g(\mathbf{z}_t;\boldsymbol{\xi}) + \Theta(L)u_t\right]\right\}$$

Again, the parameters can be chosen such that σ_t *scales to any desired level, say annualized volatility.*

In both examples we have assumed that the dimension of σ_t is annualized volatility. Then, the deterministic scaling function h is

$$h(M) = \frac{1}{\sqrt{250}}\frac{1}{\sqrt{M}}.$$

Similarly, if $\lambda \in \mathbb{R}$ is the annualized numerical parameter value for the volatility-in-mean parameter, then $\lambda_M = \lambda/(250M)$ in the example above. The web page of the first author provides an Excel spreadsheet file that illustrates intraday data generation by the HAR-RV model.

APPENDIX C. THE GRADIENT

The gradient will be used subsequently; this subsection lists its elements.

$$\frac{\partial \ell_t}{\partial d} = -e_t\frac{\partial e_t}{\partial d} - \frac{u_t}{\sigma_u^2}\frac{\partial u_t}{\partial d}$$

$$
\begin{aligned}
&= \frac{\lambda e_t}{\sigma_{tT}} \frac{\partial}{\partial d}(1-L)^d \log \sigma_{tT} - \frac{u_t}{\sigma_u^2} \Theta^{-1}(L) \left(\frac{\partial}{\partial d}(1-L)^d \log \sigma_{tT} - \frac{\partial}{\partial d} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right), \\
\frac{\partial \ell_t}{\partial \boldsymbol{\beta}} &= -e_t \frac{\partial e_t}{\partial \boldsymbol{\beta}} = e_t \frac{x_{tT}}{\sigma_{tT}}, \\
\frac{\partial \ell_t}{\partial \lambda} &= -e_t \frac{\partial e_t}{\partial \lambda} = \frac{e_t}{\sigma_{tT}} (1-L)^d \log \sigma_{tT}, \\
\frac{\partial \ell_t}{\partial \boldsymbol{\xi}} &= -\frac{u_t}{\sigma_u^2} \frac{\partial u_t}{\partial \boldsymbol{\xi}} = -\frac{u_t}{\sigma_u^2} \Theta^{-1}(L) \frac{\partial}{\partial \boldsymbol{\xi}} g(\mathbf{z}_{tT}; \boldsymbol{\xi}), \\
\frac{\partial \ell_t}{\partial \theta_i} &= -\frac{u_t}{\sigma_u^2} \frac{\partial u_t}{\partial \theta_i} = -\frac{u_t}{\sigma_u^2} \frac{\partial \Theta^{-1}(L)}{\partial \theta_i} ((1-L)^d \log \sigma_{tT} - g(\mathbf{z}_{tT}; \boldsymbol{\xi})), \\
\frac{\partial \ell_t}{\partial \sigma_u^2} &= -\frac{1}{2\sigma_u^2} + \frac{1}{2} \frac{u_t^2}{\sigma_u^4},
\end{aligned}
$$

where $\frac{\partial}{\partial \theta_i}\Theta^{-1}(L) = -c_i L/(1+\theta_i L)^2$ and c_i being the appropriate numerator constant from the partial fraction decomposition of $\Theta^{-1}(L)$.

Appendix D. Proof of Consistency

Proof of Theorem 1. Following Theorem 4.1.1 of Amemiya (1985), $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_0$ if the following conditions hold:

(1) $\boldsymbol{\Psi}$ is a compact parameter set.
(2) $\mathcal{L}_T(\boldsymbol{\psi}, \boldsymbol{\epsilon}_t)$ is continuous in $\boldsymbol{\psi}$ and measurable in $\boldsymbol{\epsilon}_t$.
(3) $\mathcal{L}_T(\boldsymbol{\psi})$ converges to a deterministic function $\mathcal{L}(\boldsymbol{\psi})$ in probability uniformly on $\boldsymbol{\Psi}$ as $T \to \infty$.
(4) $\mathcal{L}(\boldsymbol{\psi})$ attains a unique global maximum at $\boldsymbol{\psi}_0$.

Item (1) is given by assumption. Item (2) holds by definition of the normal density and construction of $\boldsymbol{\epsilon}_t$. Item (3) holds by the Ergodic Theorem if $\mathbb{E}\sup|\ell_t(\boldsymbol{\psi})| < \infty$. The latter holds by the Jensen inequality and $\mathbb{E}\sup|f(\boldsymbol{\epsilon}_t, \boldsymbol{\psi})| < \infty$, where f denotes the normal density function. The finiteness of the last expression follows from the definition of the normal density as long as $\sigma_u^2 > c > 0$ for some constant c, which is a reasonable assumption.

Consider Item (4). By the Ergodic Theorem, $\mathcal{L}(\boldsymbol{\psi}) = \mathbb{E}\ell_t(\boldsymbol{\psi})$. Rewrite the maximization problem as

$$
\max_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \mathbb{E}(\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_0)).
$$

Now,

$$
\begin{aligned}
\mathbb{E}(\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_0)) &= \mathbb{E}\log\left(\frac{f(\boldsymbol{\epsilon}_t, \boldsymbol{\psi})}{f(\boldsymbol{\epsilon}_t, \boldsymbol{\psi}_0)}\right), \\
&= \mathbb{E}\left[-\frac{1}{2}\log\frac{\sigma_u^2}{\sigma_{u,0}^2} - \frac{1}{2}\left(e_t^2 - e_{t,0}^2 + \frac{u_t^2}{\sigma_u^2} - \frac{u_{t,0}^2}{\sigma_{u,0}^2}\right)\right], \\
&= -\frac{1}{2}\log\frac{\sigma_u^2}{\sigma_{u,0}^2} - \frac{1}{2}\left(\mathbb{E}e_t^2 - 1 + \mathbb{E}(u_t^2\sigma_u^{-2}) - 1\right).
\end{aligned}
$$

Next, we show that $\mathbb{E}e_t^2(\boldsymbol{\psi}) \geq 1$ and $\mathbb{E}u_t^2(\boldsymbol{\psi}) \geq \mathbb{E}u_{t,0}^2 = \sigma_{u,0}^2$ and that the expressions attain their respective lower bounds at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ uniquely. Consider

$$
\begin{aligned}
\mathbb{E}e_t^2(\boldsymbol{\psi}) &= \mathbb{E}\left[\sigma_{tT}^{-1}(r_{tT} - \boldsymbol{\beta}'\mathbf{x}_{tT} - \lambda v_{tT})\right]^2, \\
&= \mathbb{E}\left[\sigma_{tT}^{-1}(\boldsymbol{\beta}_0'\mathbf{x}_{tT} + \lambda_0 v_{tT} + \sigma_{tT}e_{t,0} - \boldsymbol{\beta}'\mathbf{x}_{tT} - \lambda v_{tT})\right]^2, \\
&= \mathbb{E}\left[\sigma_{tT}^{-1}(\boldsymbol{\beta}_0' - \boldsymbol{\beta}')\mathbf{x}_{tT} + \sigma_{tT}^{-1}(\lambda_0 - \lambda)v_{tT} + e_{t,0}\right]^2, \\
&\geq \mathbb{E}e_{t,0}^2 = 1.
\end{aligned}
$$

The latter inequality holds since all cross terms involving $e_{t,0}$ are zero in expectation. The cross term

$$\mathbb{E}\left(2\sigma_{tT}^{-1}(\boldsymbol{\beta}_0' - \boldsymbol{\beta}')\mathbf{x}_{tT}(\lambda_0 - \lambda)v_{tT}\right) < \mathbb{E}\sigma_{tT}^{-2}(\boldsymbol{\beta}_0' - \boldsymbol{\beta}')\mathbf{x}_{tT}\mathbf{x}_{tT}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \mathbb{E}\sigma_{tT}^{-2}(\lambda_0 - \lambda)^2 v_{tT}^2,$$

and thus $\mathbb{E}e_t^2(\boldsymbol{\psi})$ takes its minimum of 1 at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ uniquely.

Consider

$$\begin{aligned}
\mathbb{E}u_t^2(\boldsymbol{\psi}) &= \mathbb{E}\left[\Theta^{-1}(L)\left((1-L)^d \log\sigma_{tT} - g(\mathbf{z}_t;\boldsymbol{\xi})\right)\right]^2, \\
&= \mathbb{E}\left[\Theta^{-1}(L)(1-L)^{d-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0) + \Theta^{-1}(L)(1-L)^{d-d_0}\Theta_0(L)u_{t,0} - \Theta^{-1}(L)g(\mathbf{z}_t;\boldsymbol{\xi})\right]^2, \\
&\geq \mathbb{E}\left[\Theta^{-1}(L)(1-L)^{d-d_0}\Theta_0(L)u_{t,0}\right]^2, \\
&\geq \mathbb{E}u_{t,0}^2 = \sigma_{u,0}^2,
\end{aligned}$$

and again, $\mathbb{E}u_t^2(\boldsymbol{\psi})$ attains its minimum of $\sigma_{u,0}^2$ uniquely at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ under Assumption 3.

So far, we have established that for any given σ_u^2, the objective function $\mathbb{E}(\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_0))$ attains its maximum of

$$-\frac{1}{2}\left[\log\frac{\sigma_u^2}{\sigma_{u,0}^2} + \frac{\sigma_{u,0}^2}{\sigma_u^2} - 1\right]$$

at $\boldsymbol{\beta} = \boldsymbol{\beta}_0, \lambda = \lambda_0, d = d_0, \Theta = \Theta_0, \boldsymbol{\xi} = \boldsymbol{\xi}_0$. Finding the value of σ_u^2 that maximizes the expression is tantamount to finding the minimum of $f(x) = \log x + 1/x$ at $x = 1$ and thus the optimal value is $\sigma_u^2 = \sigma_{u,0}^2$. This shows that $\mathbb{E}(\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_0))$ is uniquely maximized at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. □

Appendix E. Proof of Asymptotic Normality

REMARK 1.

(1) *In this section, terms will sometimes involve expectations of cross-products of the type $\mathbb{E}(XY)$, where X and Y are correlated random variables. Note that by the Cauchy-Schwarz inequality, we have*

$$\mathbb{E}XY \leq \left(\mathbb{E}X^2\right)^{\frac{1}{2}}\left(\mathbb{E}Y^2\right)^{\frac{1}{2}},$$

and thus in order to show that the cross-product has finite expectation, it suffices to show that both random variables have finite second moments.

(2) *By the same token, if both X and Y have finite second moments,*

$$\begin{aligned}
\mathbb{E}(X+Y)^2 &\leq \mathbb{E}X^2 + \mathbb{E}Y^2 + 2\left(\mathbb{E}X^2\right)^{\frac{1}{2}}\left(\mathbb{E}Y^2\right)^{\frac{1}{2}}, \\
&\leq K(\mathbb{E}X^2 + \mathbb{E}Y^2),
\end{aligned}$$

for some $K < \infty$.

LEMMA 1. *The sequence* $\left\{ \left.\frac{\partial \ell_t}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}, \mathcal{F}_t \right\}_{t=1,\ldots,T}$ *is a stationary martingale difference sequence.*

Proof. In this proof, all derivatives are evaluated at $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. The nought-subscript is suppressed to reduce notational clutter.

$$\begin{aligned}
&\mathbb{E}\left(\left.\frac{\partial \ell_t}{\partial d}\right| \mathcal{F}_{t-1}\right) \\
&= \mathbb{E}\left[\left.\frac{\lambda e_t}{\sigma_{tT}}\frac{\partial}{\partial d}(1-L)^d\log\sigma_{tT} - \frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\left(\frac{\partial}{\partial d}(1-L)^d\log\sigma_{tT} - \frac{\partial}{\partial d}g(\mathbf{z}_{tT};\boldsymbol{\xi})\right)\right|\mathcal{F}_{t-1}\right] \\
&= 0,
\end{aligned}$$

since e_t and u_t have mean zero, $\frac{\partial}{\partial d}(1-L)^d\log\sigma_{tT}$ is uncorrelated with e_t and σ_{tT}, and both $\frac{\partial}{\partial d}(1-L)^d\log\sigma_{tT}$ and $\frac{\partial}{\partial d}g(\mathbf{z}_{tT};\boldsymbol{\xi})$ do not contain u_t.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \boldsymbol{\beta}}\middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(e_t \frac{x_{tT}}{\sigma_{tT}}\middle| \mathcal{F}_{t-1}\right) = 0,$$

since e_t has mean zero, x_{tT} is predetermined, and σ_{tT} is independent of e_t.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \lambda}\middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(\frac{e_t}{\sigma_{tT}}(1-L)^d \log \sigma_{tT}\middle| \mathcal{F}_{t-1}\right) = 0,$$

since e_t has mean zero and is independent of both, σ_{tT} and $(1-L)^d \log \sigma_{tT}$.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(-\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\frac{\partial}{\partial \boldsymbol{\xi}} g(\mathbf{z}_{tT};\boldsymbol{\xi})\middle| \mathcal{F}_{t-1}\right) = 0,$$

since $g(\mathbf{z}_{tT};\boldsymbol{\xi})$ is independent of u_t.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \theta_i}\middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(-\frac{u_t}{\sigma_u^2}\frac{\partial \Theta^{-1}(L)}{\partial \theta_i}((1-L)^d \log \sigma_{tT} - g(\mathbf{z}_{tT};\boldsymbol{\xi}))\middle| \mathcal{F}_{t-1}\right) = 0,$$

since $\frac{\partial \Theta^{-1}(L)}{\partial \theta_i}(1-L)^d \log \sigma_{tT}$ does not contain u_t.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \sigma_u^2}\middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(-\frac{1}{2\sigma_u^2} + \frac{1}{2}\frac{u_t^2}{\sigma_u^4}\middle| \mathcal{F}_{t-1}\right) = 0,$$

since u_t has mean zero and variance σ_u^2. □

LEMMA 2.

(1)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial \ell_t}{\partial \boldsymbol{\psi}}\right| < \infty$$

(2)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial \ell_t}{\partial \boldsymbol{\psi}}\frac{\partial \ell_t}{\partial \boldsymbol{\psi}'}\right| < \infty$$

Proof. In this proof, the expressions are evaluated at any $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ if not otherwise stated. The data-generating parameters will be explicitly denoted by a 0-subscript. The processes r_{tT}, σ_{tT}, and $\log \sigma_{tT}$ are returns, realized volatility, and logarithmic realized volatility *data* and thus evaluated at $\boldsymbol{\psi}_0$ throughout.

We will consider the gradient vector element by element:

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial \ell_t}{\partial d}\right| =$$
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\lambda e_t}{\sigma_{tT}}\frac{\partial}{\partial d}(1-L)^d \log \sigma_{tT} - \frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\left(\frac{\partial}{\partial d}(1-L)^d \log \sigma_{tT} - \frac{\partial}{\partial d} g(\mathbf{z}_{tT};\boldsymbol{\xi})\right)\right|$$

Using the triangular and Cauchy-Schwarz inequalities, we need to find upper bounds for the following objects.

(1)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{e_t}{\sigma_{tT}}\right|^p,$$

(2)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|\frac{\partial}{\partial d}(1-L)^d \log \sigma_{tT}\right|^p,$$

(3)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|u_t(\boldsymbol{\psi})\right|^p,$$

(4)
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left|u_t(\boldsymbol{\psi})\frac{\partial}{\partial d} g(\mathbf{z}_{tT};\boldsymbol{\xi})\right|^p,$$

where $p = 1, 2$.

Re (1). Consider

$$\begin{aligned}
\mathbb{E}\left|\frac{e_t}{\sigma_{tT}}\right|^p &= \mathbb{E}\left|\frac{1}{\sigma_{tT}^2}(r_{tT} - \boldsymbol{\beta}'\mathbf{x}_{tT} - \lambda(1-L)^{-d}\log\sigma_{tT})\right|^p, \\
&= \mathbb{E}\left|e^{-2[(1-L)^{-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi})+(1-L)^{-d_0}\Theta_0(L)u_{t,0}]}\times\right. \\
&\quad \left.\left((\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{x}_{tT} + \left(\lambda_0(1-L)^{-d_0} - \lambda(1-L)^{-d}\right)\log\sigma_{tT} + \sigma_{tT}e_{t,0}\right)\right|^p, \\
&\leq \quad K\left[\left(\mathbb{E}\left|\sigma_{tT}^{-2}\right|^q\right)^{\frac{1}{q}}\left(\mathbb{E}\left|(\boldsymbol{\beta}_0-\boldsymbol{\beta})'\mathbf{x}_{tT}\right|^q\right)^{\frac{1}{q}}\right. \\
&\quad + \left(\mathbb{E}\left|\sigma_{tT}^{-2}\right|^q\right)^{\frac{1}{q}}\left(\mathbb{E}\left|\left(\lambda_0(1-L)^{-d_0} - \lambda(1-L)^{-d}\right)\log\sigma_{tT}\right|^q\right)^{\frac{1}{q}} \\
&\quad \left. + \left(\mathbb{E}\left|\sigma_{tT}^{-1}\right|^q\right)^{\frac{1}{q}}\left(\mathbb{E}\left|e_{t,0}\right|^q\right)^{\frac{1}{q}}\right],
\end{aligned}$$

where the latter inequality is obtained using the triangular and Hölder inequalities, and $K < \infty$. The exponent $q = 2p$, $p = 1, 2$ in the Hölder inequality is chosen identical for all terms for notational convenience; the more general statement of the inequality is not necessary for our purposes. Consider first

$$\begin{aligned}
\mathbb{E}\left|\sigma_{tT}^{-2}\right|^q &= \mathbb{E}\left|e^{-2[(1-L)^{-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0)+(1-L)^{-d_0}\Theta_0(L)u_{t,0}]}\right|^q \\
&= \mathbb{E}e^{-2q(1-L)^{-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0)}\mathbb{E}e^{-2q(1-L)^{-d_0}\Theta_0(L)u_{t,0}},
\end{aligned}$$

by Assumptions 5 (3) and (5) and 3 (2) and (4).

Consider next

$$\begin{aligned}
\mathbb{E}\left|(1-L)^d\log\sigma_{tT}\right|^q &= \mathbb{E}\left|(1-L)^d[(1-L)^{-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0) + (1-L)^{-d_0}\Theta_0(L)u_{t,0}]\right|^q, \\
&< K(\mathbb{E}\left|(1-L)^{d-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0)\right|^q + \mathbb{E}\left|(1-L)^{d-d_0}\Theta_0(L)u_{t,0}\right|^q) < \infty,
\end{aligned}$$

for some $K < \infty$ by Assumptions 5 (4), 3 (3), and 4.

Re (2),

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial}{\partial d}(1-L)^d\log\sigma_{tT}\right|^q \\
&= \mathbb{E}\left|\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)\prod_{i=0}^{j-1}(d-i)L^j\log\sigma_{tT}\right|^q, \\
&= \mathbb{E}\left|\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)\prod_{i=0}^{j-1}(d-i)L^j[(1-L)^{-d_0}g(\mathbf{z}_{tT};\boldsymbol{\xi}_0) + (1-L)^{-d_0}\Theta_0(L)u_{t,0}]\right|^q < \infty,
\end{aligned}$$

from the same set of assumptions as the last element of the gradient, recognizing that the derivative $\frac{\partial}{\partial d}(1-L)^d$ retains the stationarity of $(1-L)^d$ if $d \in (-1/2, 1/2)$.

Re (3),

$$\begin{aligned}
\mathbb{E}\left|u_t(\boldsymbol{\psi})\right|^p &= \mathbb{E}\left|\Theta^{-1}(L)\left[(1-L)^d\log\sigma_{tT} - g(\mathbf{z}_{tT};\boldsymbol{\xi})\right]\right|^p, \\
&\leq K(\mathbb{E}\left|\Theta^{-1}(L)(1-L)^d\log\sigma_{tT}\right|^p + \mathbb{E}\left|\Theta^{-1}(L)g(\mathbf{z}_{tT};\boldsymbol{\xi})\right|^p) < \infty.
\end{aligned}$$

The first term was shown to be finite above, recognizing that only parameters $\boldsymbol{\psi}$ are considered such that $\Theta(L)$ remains in the invertible region. Then, the second term is also finite by Assumption 5 (4).

Re (4),

$$
\begin{aligned}
&\mathbb{E}\left|u_t(\boldsymbol{\psi})\frac{\partial}{\partial d}g(\mathbf{z}_{tT};\boldsymbol{\xi})\right|^q \\
&= (\mathbb{E}\left|u_t(\boldsymbol{\psi})\right|^q)^{\frac{1}{q}}\left(\mathbb{E}\left|\frac{\partial}{\partial d}g(\mathbf{z}_{tT};\boldsymbol{\xi})\right|^q\right)^{\frac{1}{q}} < \infty.
\end{aligned}
$$

The first term is shown to be finite above. The second term is finite by Assumption 5 (6).

The next element in the gradient vector is

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \boldsymbol{\beta}}\right| &= \mathbb{E}\left|\mathbf{x}_{tT}\frac{e_t}{\sigma_{tT}}\right|, \\
&\leq (\mathbb{E}\left|\mathbf{x}_{tT}\right|^p)^{\frac{1}{p}}\left(\mathbb{E}\left|\frac{e_t}{\sigma_{tT}}\right|^p\right)^{\frac{1}{p}} < \infty.
\end{aligned}
$$

The finiteness of the second factor was shown above. The first factor is assumed to be finite in Assumption 4 (2).

All other elements of the gradient vector are bounded by the same arguments and assumptions:

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \lambda}\right| &= \mathbb{E}\left|\frac{e_t}{\sigma_{tT}}(1-L)^d \log \sigma_{tT}\right|, \\
&\leq \left(\mathbb{E}\left|\frac{e_t}{\sigma_{tT}}\right|^p\right)^{\frac{1}{p}}\left(\mathbb{E}\left|(1-L)^d \log \sigma_{tT}\right|^p\right)^{\frac{1}{p}} < \infty.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\right| &= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\frac{\partial}{\partial \boldsymbol{\xi}}g(\mathbf{z}_{tT};\boldsymbol{\xi})\right| \\
&\leq \left(\mathbb{E}\left|\frac{u_t}{\sigma_u^2}\right|^p\right)^{\frac{1}{p}}\left(\mathbb{E}\left|\Theta^{-1}(L)\frac{\partial}{\partial \boldsymbol{\xi}}g(\mathbf{z}_{tT};\boldsymbol{\xi})\right|^p\right)^{\frac{1}{p}} < \infty.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \theta_i}\right| &= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\frac{\partial u_t}{\partial \theta_i}\right|, \\
&= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\frac{\partial \Theta^{-1}(L)}{\partial \theta_i}\left[(1-L)^d \log \sigma_{tT} - g(\mathbf{z}_{tT};\boldsymbol{\xi})\right]\right|, \\
&= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\left(-\frac{c_i L}{(1+\theta_i L)^2}\right)\left[(1-L)^d \log \sigma_{tT} - g(\mathbf{z}_{tT};\boldsymbol{\xi})\right]\right|, \\
&\leq \left(\mathbb{E}\left|\frac{u_t}{\sigma_u^2}\right|^p\right)^{\frac{1}{p}}\left(\mathbb{E}\left|\frac{c_i L}{(1+\theta_i L)^2}\left[(1-L)^d \log \sigma_{tT} - g(\mathbf{z}_{tT};\boldsymbol{\xi})\right]\right|^p\right)^{\frac{1}{p}} < \infty.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \sigma_u^2}\right| &= \mathbb{E}\left|\frac{1}{2\sigma_u^2} + \frac{1}{2}\frac{u_t^2}{\sigma_u^4}\right|, \\
&\leq \frac{1}{2\sigma_u^2} + \frac{1}{2}\mathbb{E}\left|\frac{u_t^2}{\sigma_u^4}\right| < \infty.
\end{aligned}
$$

This shows statement (1) of Lemma 2. Statement (2) of Lemma 2 uses the same arguments with the only difference that for part (1), the exponents in the Hölder inequalities are at most equal to two, whereas for statement (2), we need $q = 4$. We omit the details of (2) for the sake of brevity; they can be obtained from the authors. □

LEMMA 3. *The function*

$$
g_t(\boldsymbol{\psi}) := -\frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \mathbf{A}(\boldsymbol{\psi})
$$

where

$$\mathbf{A}(\boldsymbol{\psi}) = -\mathbb{E}\frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}$$

is absolutely uniformly integrable:

$$\mathbb{E} \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g_t(\boldsymbol{\psi})| < \infty;$$

it is continuous in $\boldsymbol{\psi}$ *and has zero mean:* $\mathbb{E} g_t(\boldsymbol{\psi}) = 0$.

Proof. By the Ergodic Theorem, we have pointwise convergence of $-1/T \sum_{t=1}^{T} \partial^2 \ell_t / \partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'$ to $\mathbf{A}$. By the triangular inequality, showing the absolute uniform integrability reduces to showing that

$$\mathbb{E} \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \left| \frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right| < \infty.$$

We will show the statement for the second derivative of ℓ_t with respect to d, which requires most work and assumptions. There are 21 distinct second derivatives in $\mathbf{A}$; proving finiteness of the expected value of the supremum consists of repeated application of the Lebesgue Dominated Convergence Theorem (Shiryaev (1996, p. 187), Ling and McAleer (2003), Lemmata 5.3 and 5.4).

First, note that

$$\begin{aligned}\frac{\partial^2}{\partial d^2}(1-L)^d &= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left[\left(\sum_{i=0}^{j-1} \frac{1}{d-i} \right)^2 - \sum_{i=0}^{j-1} \left(\frac{1}{d-i} \right)^2 \right] \prod_{i=0}^{j-1} (d-i) L^j, \qquad (13)\\ &= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left[\sum_{\substack{i,k=0\\ i \neq k}}^{j-1} \frac{1}{(d-i)(d-k)} \right] \prod_{i=0}^{j-1} (d-i) L^j.\end{aligned}$$

Then we have

$$\begin{aligned}\frac{\partial^2 \ell_t}{\partial d^2} &= -\frac{\lambda^2}{\sigma_{tT}} \left(\frac{\partial}{\partial d} (1-L)^d \log \sigma_{tT} \right)^2 + \frac{\lambda e_t}{\sigma_{tT}} \frac{\partial^2}{\partial d^2} (1-L)^d \log \sigma_{tT}\\ &- \frac{1}{\sigma_u^2} \Theta^{-2}(L) \left(\frac{\partial}{\partial d} (1-L)^d \log \sigma_{tT} - \frac{\partial}{\partial d} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right)^2\\ &- \frac{u_t}{\sigma_u^2} \Theta^{-1}(L) \left(\frac{\partial^2}{\partial d^2} (1-L)^d \log \sigma_{tT} - \frac{\partial^2}{\partial d^2} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right)\\ &= R_1 + R_2 + R_3 + R_4.\end{aligned}$$

We proceed to show that $\mathbb{E} \sup |R_i| < \infty$ for $i = 1, \ldots, 4$.

$$\begin{aligned}|R_1| &= \left| \lambda e^{-2[(1-L)^{-d_0} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) + (1-L)^{-d_0} \Theta_0(L) u_{t,0}]} \left[\frac{\partial}{\partial d} \left((1-L)^{-d_0} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) + (1-L)^{-d_0} \Theta_0(L) u_{t,0} \right) \right]^2 \right|\\ &\leq K \left| \frac{\lambda^2}{\sigma_{tT}^2} \left(\frac{\partial}{\partial d} (1-L)^d (1-L)^{-d_0} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right)^2 \right| + K \left| \frac{\lambda^2}{\sigma_{tT}^2} \left(\frac{\partial}{\partial d} (1-L)^d (1-L)^{-d_0} \Theta_0(L) u_{t,0} \right)^2 \right|,\end{aligned}$$

for $K < \infty$. The expected value of the terms on the right-hand side is finite, as shown in the proof of Lemma 2. Therefore, the supremum of the left-hand side is dominated by the right-hand side and $\mathbb{E} \sup |R_1| < \infty$ by the Lebesgue Dominated Convergence Theorem.

$$\begin{aligned}|R_2| &= \frac{\lambda e_t}{\sigma_{tT}^2} \left(\frac{\partial^2}{\partial d^2} (1-L)^d (1-L)^{-d_0} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) + \frac{\partial^2}{\partial d^2} (1-L)^d (1-L)^{-d_0} \Theta_0(L) u_{t,0} \right),\\ &\leq \left| \frac{\lambda e_t}{\sigma_{tT}^2} \frac{\partial^2}{\partial d^2} (1-L)^d (1-L)^{-d_0} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right| + \left| \frac{\lambda e_t}{\sigma_{tT}^2} \frac{\partial^2}{\partial d^2} (1-L)^d (1-L)^{-d_0} \Theta_0(L) u_{t,0} \right|.\end{aligned}$$

The finiteness of the expected value of the terms on the right-hand side follows similar arguments as in the proof of Lemma 2, using representation (13). So again, the supremum of the left-hand side has finite expectation by the Lebesgue Dominated Convergence Theorem. The boundedness of R_3 and R_4 follow in the same fashion:

$$
\begin{aligned}
|R_3| &= \left| \frac{1}{\sigma_u^2} \Theta^{-2}(L) \left(\frac{\partial}{\partial d} \log \sigma_{tT} - \frac{\partial}{\partial d} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right)^2 \right|, \\
&\leq K \left[\left| \frac{1}{\sigma_u^2} \Theta^{-2}(L) \left(\frac{\partial}{\partial d} \log \sigma_{tT} \right)^2 \right| + \left| \frac{1}{\sigma_u^2} \Theta^{-2}(L) \left(\frac{\partial}{\partial d} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right)^2 \right| \right],
\end{aligned}
$$

$$
\begin{aligned}
|R_4| &= \left| \frac{u_t}{\sigma_u^2} \Theta^{-2}(L) \left(\frac{\partial^2}{\partial d^2} (1-L)^d \log \sigma_{tT} - \frac{\partial^2}{\partial d^2} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right) \right| \\
&\leq \left| \frac{u_t}{\sigma_u^2} \Theta^{-2}(L) \frac{\partial^2}{\partial d^2} (1-L)^d \log \sigma_{tT} \right| + \left| \frac{u_t}{\sigma_u^2} \Theta^{-2}(L) \frac{\partial^2}{\partial d^2} g(\mathbf{z}_{tT}; \boldsymbol{\xi}) \right|.
\end{aligned}
$$

We use arguments from the proof of Lemma 2 and Assumption 5 (7) for the last term. Thus,

$$\mathbb{E} \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g_t(\boldsymbol{\psi})| < \infty.$$

Further, $g_t(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi}$ by the Continuous Mapping Theorem and has zero mean by construction. □

Proof of Theorem 2. The proof follows Theorem 4.1.3 of Amemiya (1985). First, we have to establish that $\widehat{\boldsymbol{\psi}}_T$ is consistent (Thm. 1). Then,

$$\mathbf{B}(\boldsymbol{\psi}_0)^{-\frac{1}{2}} \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \bigg|_{\boldsymbol{\psi}_0} \xrightarrow{d} W(r), \ r \in [0,1],$$

where $W(r)$ is $(k_x + k_\xi + q + 4)$-dimensional standard Brownian motion on the unit interval. This convergence follows from Theorem 18.3 in Billingsley (1999) if (a) $\left\{ \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \Big|_{\boldsymbol{\psi}_0}, \mathcal{F}_t \right\}$ is a stationary martingale difference sequence (Lemma 1), and (b) $\mathbf{B}(\boldsymbol{\psi}_0)$ exists (Lemma 2). Further, we have to show that

$$\mathbf{A}_T(\boldsymbol{\psi}_T^*) \xrightarrow{p} \mathbf{A}(\boldsymbol{\psi}_0)$$

for any sequence $\boldsymbol{\psi}_T^* \xrightarrow{p} \boldsymbol{\psi}_0$,

$$\mathbf{A}_T(\boldsymbol{\psi}_T^*) = -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \bigg|_{\boldsymbol{\psi}_T^*},$$

and

$$\mathbf{A}(\boldsymbol{\psi}_0) = -\mathbb{E} \left. \frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\boldsymbol{\psi}_0}$$

is non-singular. Conditions for this double stochastic convergence can be found, for example, in Theorem 21.6 of Davidson (1994). We need to have (a) consistency of $\widehat{\boldsymbol{\psi}}_T$ for $\boldsymbol{\psi}_0$ and (b) uniform convergence of $\mathbf{A}_T$ to $\mathbf{A}$ in probability, i.e.

$$\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |\mathbf{A}_T(\boldsymbol{\psi}) - \mathbf{A}(\boldsymbol{\psi})| \xrightarrow{p} 0.$$

To show uniform convergence, often a stochastic version of the Arzelà-Ascoli theorem (e.g. Theorem 21.9 in Davidson (1994)) is employed, which in a simple version shows the equivalence of uniform convergence and equicontinuity. By proving stochastic equicontinuity, for example by checking the conditions of Theorem 2 of Andrews (1992), which involves showing the finiteness of the third derivatives of the likelihood function, uniform convergence is established. In this proof, we follow Berkes, Horváth, and Kokoszka (2003) and Ling and McAleer (2003, Theorem 3.1) in particular, who employ the Ergodic Theorem to obtain uniform convergence directly by modifying Theorem 4.2.1 of Amemiya (1985). To employ

Theorem 3.1 of Ling and McAleer (2003), we have to show that

$$g_t(\boldsymbol{\psi}) = -\frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} - \mathbf{A}(\boldsymbol{\psi})$$

is continuous in $\boldsymbol{\psi}$, has expected value $\mathbb{E} g_t(\boldsymbol{\psi}) = 0$ and is absolutely uniformly integrable:

$$\mathbb{E} \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} |g_t(\boldsymbol{\psi})| < \infty$$

(Lemma 3). Thus, we have established all conditions for asymptotic normality according to Theorem 4.1.3 of Amemiya (1985). □

Proof of Proposition 1. We established uniform convergence in probability of $\mathbf{A}_T$ to $\mathbf{A}$ in Lemma 3 and Theorem 2. It remains to show uniform convergence of $\mathbf{B}_T$ to $\mathbf{B}$. We follow Theorem 3.1 of Ling and McAleer (2003) again. Define

$$h_t(\boldsymbol{\psi}) := \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}'} - \mathbf{B}(\boldsymbol{\psi}).$$

As we did for $\mathbf{A}$ in Lemma 3, we have to show that h_t is absolutely uniformly integrable, continuous in ψ, and has expected value $\mathbb{E} h_t(\boldsymbol{\psi}) = 0$. By the triangular inequality, showing absolute uniform integrability reduces to showing that

$$\mathbb{E} \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \frac{\partial \ell_t}{\partial \boldsymbol{\psi}'} < \infty.$$

This can be shown using Lebesgue dominated convergence arguments very similar to those employed in the proof of Lemma 3. We omit the details for brevity. The function h_t is continuous in $\boldsymbol{\psi}$ by the Continuous Mapping Theorem and has zero-mean by construction. □

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press, Cambridge.

ANDERSEN, T., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–906.

ANDERSEN, T., T. BOLLERSLEV, P. CHRISTOFFERSEN, AND F. DIEBOLD (2007): "Practical volatility and correlation modeling for financial market risk management," in *Risks of Financial Institutions*, ed. by M. Carey, and R. Stulz. University of Chicago Press for NBER.

ANDERSEN, T., T. BOLLERSLEV, AND F. DIEBOLD (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

ANDERSEN, T., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001a): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2001b): "Exchange rate returns standardized by realized volatility are (nearly) Gaussian," *Multinational Finance Journal*, forthcoming.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

ANDERSEN, T. G., T. BOLLERSLEV, P. H. FEDERIKSEN, AND M. Ø. NIELSEN (2007): "Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns," Working paper no. 21, CREATES, University of Aarhus.

ANDERSEN, T. G., T. BOLLERSLEV, AND X. HUANG (2007): "A Reduced Form Framework for Modeling Volatility of Speculative Prices Based on Realized Variation Measures," Working paper no. 14, CREATES, University of Aarhus.

ANDERSSON, M. K., B. EKLUND, AND J. LYHAGEN (1999): "A Simple Linear Time Series Model with Misleading Nonlinear Properties," *Economics Letters*, 65, 281–284.

ANDREWS, D. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

ANDREWS, D. W. K., AND C. J. MCDERMOTT (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies*, 62, 343–360.

BAI, J. (1998): "A Note on Spurious Break," *Econometric Theory*, 14, 663–669.

BAILLIE, R., AND G. KAPETANIOS (2007a): "Nonlinear Models for Strongly Dependent Processes with Financial Applications," *Journal of Econometrics*, forthcoming.

——— (2007b): "Testing for Neglected Nonlinearity in Long-Memory Models," *Journal of Business and Economic Statistics*, 25, 447–461.

BANDI, F., AND J. RUSSELL (2005a): "Market microstructure noise, integrated variance estimators, and the limitations of asymptotic approximations: a solution," Unpublished manuscript, Graduate School of Business, University of Chicago.

——— (2005b): "Microstructure noise, realized volatility, and optimal sampling," Unpublished manuscript, Graduate School of Business, University of Chicago.

——— (2006): "Separating market microstructure noise from volatility, Journal of Financial Economics," *Journal of Financial Economics*, 79, 655–692.

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2007): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," Unpublished manuscript, Stanford University.

BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2002): "Econometric analysis of realised volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society B*, 64, 253–280.

BERKES, I., L. HORVÁTH, AND P. KOKOSZKA (2003): "GARCH Processes: Structure and Estimation," *Bernoulli*, 9, 201–227.

BIAIS, B., L. GLOSTEN, AND C. SPATT (2005): "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications," *Journal of Financial Markets*, 8, 217–264.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. Wiley, New York, second edn.

BLACK, F. (1976): "Studies in Stock Price Volatility Changes," *Proceedings of The American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 21, 307–328.

——— (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *The Review of Economics and Statistics*, 69, 542–547.

BOLLERSLEV, T., U. KRETSCHMER, C. PIGORSCH, AND G. TAUCHEN (2007): "A Discrete-Time for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," Working paper no. 22, CREATES, University of Aarhus.

BRANDT, M. W., AND Q. KIANG (2004): "On the relationship between conditional mean and volatility of stock returns: A latent VAR approach," *Journal of Financial Economics*, 72, 217–257.

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

CAMPBELL, J., AND L. HENTSCHEL (1992): "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 281–318.

CAMPBELL, J., A. LO, AND A. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

CARNERO, M. A., D. PEÑA, AND E. RUIZ (2004): "Persistence and kurtosis in GARCH and stochastic volatility models," *Journal of Financial Econometrics*, 2, 319–342.

CHRISTENSEN, B. J., AND M. Ø. NIELSEN (2007): "The Effect of Long Memory in Volatility on Stock Market Fluctuations," *Review of Economics and Statistics*, forthcoming.

CHRISTENSEN, B. J., M. Ø. NIELSEN, AND J. ZHU (2007): "Long Memory in Stock Market Volatility and the Volatility-in-Mean Effect: The FIEGARCH-M Model," Working paper no. 10, CREATES, University of Aarhus.

CHUNG, C.-F., AND R. T. BAILLIE (1993): "Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models," *Empirical Economics*, 18, 791–806.

CORSI, F. (2004): "A Simple Long Memory Model of Realized Volatility," Manuscript, University of Southern Switzerland.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.

DAVIDSON, J., AND P. SIBBERTSEN (2005): "Generating schemes for long memory processes: regimes, aggregation and linearity," *Journal of Econometrics*, 128, 253–282.

DIEBOLD, F. (1986): "Modeling the persistence of conditional variances: A comment," *Econometric Reviews*, 5, 51–56.

DIEBOLD, F., AND A. INOUE (2001): "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflations," *Econometrica*, 50, 987–1007.

FAMA, E., AND K. FRENCH (1988): "Permanent and temporary components of stock prices," *Journal of Political Economy*, 96, 246–273.

FINE, T. L. (1999): *Feedforward Neural Network Methodology*. Springer, New York.

FOUQUE, J., G. PAPANICOLAOU, AND R. SIRCAR (2000): *Derivatives in financial markets with stochastic volatility*. Cambridge University Press, Cambridge.

FOUQUE, J. P., G. PAPANICOLAOU, R. SIRCAR, AND K. SØLNA (2003): "Short Time-Scale in S&P 500 Volatility," *Journal of Computational Finance*, 6, 1–23.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4(4), 221–238.

GLOSTEN, L., R. JAGANNATHAN, AND R. RUNKLE (1993): "On The Relationship Between The Expected Value and The Volatility of The Nominal Excess Returns on Stocks," *Journal of Finance*, 48, 1779–1801.

GONZALEZ-RIVERA, G. (1998): "Smooth Transition GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 3(2), 61–78.

GRANGER, C., AND N. HYUNG (2004): "Occasional Structural Breaks and Long Memory with an Application to the S&P500 Absolute Stock Returns," *Journal of Empirical Finance*, 3, 399–421.

GRANGER, C. W. J. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

GRANGER, C. W. J., AND T. TERÄSVIRTA (2001): "A Simple Nonlinear Time Series Model with Misleading Linear Properties," *Economics Letters*, 62, 161–165.

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

HAGERUD, G. E. (1997): "A New Non-Linear GARCH Model," Phd. thesis, Stockholm School of Economics.

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

HANSEN, B. E. (1996): "Inference when a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

HANSEN, P. (2005): "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics*, 23, 365–380.

HANSEN, P., J. LARGE, AND A. LUNDE (2007): "Moving average-based estimators of integrated variance," *Econometric Reviews*, forthcoming.

HANSEN, P. R., AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything beat a GARCH(1,1) Model?," *Journal of Applied Econometrics*, 20, 873–889.

——— (2006): "Realized variance and market microstructure noise (with discussion)," *Journal of Business and Economic Statistics*, 24, 127–218.

HARVEY, C. (2001): "The specification of conditional expectations," *Journal of Emprical Finance*, 8, 573–638.

HILLEBRAND, E. (2005): "Neglecting Parameter Changes in GARCH Models," *Journal of Econometrics*, 129, 121–138.

HILLEBRAND, E., AND M. MEDEIROS (2006): "Estimating and Forecasting GARCH Models in The Presence of Structural Breaks and Regime Switches," in *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, ed. by M. Wohar, and D. Rapach. Elsevier, forthcoming.

HSU, C.-C. (2001): "Change-Point Estimation in Regressions with I(d) Variables," *Economics Letters*, 70, 147–155.

HYUNG, N., S.-H. POON, AND C. GRANGER (2005): "The Source of Long Memory in Financial Market Volatility," in *Forecasting in The Presence of Structural Breaks and Model Uncertainty*, ed. by M. Wohar, and D. Rapach. Elsevier, forthcoming.

J. P. MORGAN (1996): *J. P. Morgan/Reuters Riskmetrics – Technical Document*. J. P. Morgan, New York.

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

LEBARON, B. (2001): "Stochastic Volatility as a Simple Generator of Apparent Financial Power Laws and Long Memory," *Quantitative Finance*, 1, 621–631.

LI, C. W., AND W. K. LI (1996): "On a Double Threshold Autoregressive Heteroscedastic Time Series Model," *Journal of Applied Econometrics*, 11, 253–274.

LIN, C. F. J., AND T. TERÄSVIRTA (1994): "Testing the Constancy of Regression Parameters Against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228.

LING, S., AND M. MCALEER (2003): "Asymptotic Theory for a Vector ARMA-GARCH Model," *Econometric Theory*, 19, 280–310.

LIU, J., W. K. LI, AND C. W. LI (1997): "On a Threshold Autoregression with Conditional Heteroscedastic Variances," *Journal of Statistical Planning and Inference*, 62(2), 279–300.

LOBATO, I. N., AND N. E. SAVIN (1998): "Real and Spurious Long-Memory Properties of Stock-Market Data," *Journal of Business and Economic Statistics*, 16, 261–268.

LONGIN, F. M. (1997): "The Threshold Effect in Expected Volatility: A Model Based on Asymmetric Information," *Review of Financial Studies*, 10, 837–869.

LUNDBERGH, S., AND T. TERÄSVIRTA (1998): "Modelling Economic High-Frequency Time Series with STAR-STGARCH Models," Working Paper Series in Economics and Finance 291, Stockholm School of Economics.

MADHAVAN, A. (2000): "Market Microstructure: A Survey," *Journal of Financial Markets*, 3, 205–258.

MALMSTEN, H., AND T. TERÄSVIRTA (2004): "Stylized Facts of Financial Time Series and Three Popular Models of Volatility," Working Paper Series in Economics and Finance 563, Stockholm School of Economics.

MARTENS, M., D. VAN DIJK, AND M. DE POOTER (2004): "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Discussion Paper 04-067/4, Tinbergen Institute.

MCALEER, M. (2005): "Automated Inference and Learning in Modeling Financial Volatility," *Econometric Theory*, 21, 232–261.

MCALEER, M., AND M. MEDEIROS (2006): "Realized Volatilty: A Review," *Econometric Reviews*, forthcoming.

——— (2007): "A Multiple Regime Smooth Transition Heterogenous Autoregressive Model for Long Memory and Asymmetries," *Journal of Econometrics*, forthcoming.

MEDDAHI, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17, 479–508.

MEDEIROS, M., AND A. VEIGA (2007): "Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH(1,1) Model," *Econometric Theory*, forthcoming.

MEDEIROS, M. C., T. TERÄSVIRTA, AND G. RECH (2006): "Building Neural Network Models for Time Series: A Statistical Approach," *Journal of Forecasting*, 25, 49–75.

MEDEIROS, M. C., AND A. VEIGA (2003): "Diagnostic Checking in a Flexible Nonlinear Time Series Model," *Journal of Time Series Analysis*, 24, 461–482.

MERTON, R. (1980): "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8, 323–361.

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

MÜLLER, U. A., M. M. DACOROGNA, R. D. DAVE, R. B. OLSEN, O. V. PICTET, AND J. E. VON WEIZSÄCKER (1997): "Volatilities of different time resolutions - Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–239.

NELSON, D. B. (1991): "Conditinal Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

NUNES, L. C., C.-M. KUAN, AND P. NEWBOLD (1995): "Spurious Break," *Econometric Theory*, 11, 736–749.

OOMEN, R. (2005): "Properties of bias-corrected realized variance under alternative sampling schemes," *Journal of Financial Econometrics*, 3, 555–577.

PATTON, A. (2005): "Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies," Unpublished manuscript, London School of Economics.

RABEMANANJARA, R., AND J. M. ZAKOIAN (1993): "Threshold ARCH Models and Asymmetries in Volatility," *Journal of Applied Econometrics*, 8(1), 31–49.

SAIKKONEN, P., AND I. CHOI (2004): "Cointegrating Smooth Transition Regressions," *Econometric Theory*, 20, 301–340.

SHIRYAEV, A. (1996): *Probability*. Springer, New York, second edn.

SMITH, A. (2005): "Level Shifts and the Illusion of Long Memory in Economic Time Series," *Journal mf Business and Economic Statistics*, 23, 321–335.

SOWELL, F. (1992): "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.

TAYLOR, S. J. (1986): *Modelling Financial Time Series*. John Wiley, Chichester, UK.

TERÄSVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

TERÄSVIRTA, T., C. F. LIN, AND C. W. J. GRANGER (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 309–323.

VAN DIJK, D., P. FRANSES, AND R. PAAP (2002): "A Nonlinear Long Memory Model, With an Application to US Unemployment," *Journal of Econometrics*, 110, 135–165.

ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

(E. Hillebrand) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.
E-mail address: erhil@lsu.edu

(M. C. Medeiros) DEPARTMENT OF ECONOMICS, PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO, RIO DE JANEIRO, RJ, BRAZIL.
E-mail address: mcm@econ.puc-rio.br