ASYMMETRIES, BREAKS, AND LONG-RANGE DEPENDENCE: AN ESTIMATION FRAMEWORK FOR 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 autoregressive moving average (ARMA) time series models and apply our modeling framework to series of daily realized volatility. Asymptotic theory for the quasi-maximum likelihood estimator is developed and a sequence of model specification tests is described. Our framework allows for general nonlinear functions, including smoothly changing intercepts. The theoretical results in the paper can be applied to any series with long memory and nonlinearity. We apply the methodology to realized volatility of individual stocks of the Dow Jones Industrial Average during the period 1995 to 2005. We find strong evidence of nonlinear effects and explore different specifications of the model framework. A forecasting exercise demonstrates that allowing for nonlinearities in long memory models yields significant performance gains.

KEYWORDS: Realized volatility, structural breaks, smooth transitions, nonlinear models, long memory, persistence.

ACKNOWLEDGMENTS: We gratefully acknowledge comments from Richard Baillie, Efrem Castelnuovo, Christian Dahl, Marcelo Fernandes, Jean-Pierre Fouque, Ana Galvão, Liudas Giraitis, Peter Reinhard Hansen, Jim Hilliard, Hagen Kim, Asger Lunde, Tae-Hwy Lee, Michael McAleer, Morten Nielsen, Joon Park, Timo Teräsvirta, participants at the conference "Breaks and Persistence in Econometrics" at the Cass Business School London, at the Midwest Econometrics Meetings in St. Louis, and at seminars at LSU, Texas A&M, Temple University, CREATES, Queen Mary University of London, University of Padua, Getulio Vargas Foudation, and the University of Cincinnati. 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 and Pierre Xu 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."

Date: November 27, 2009.

1. Introduction

1.1. **Overview and Main Results.**

In this paper we propose a framework to model and forecast time series that display long-range dependence and nonlinear behavior. The methodology is applied to series of daily realized volatilities of 23 stocks of the Dow Jones Industrial Average during the period 3-Jan-1995 to 31-Dec-2005. Realized volatility can be seen as a proxy for the conditional volatility of daily returns.

Our modeling framework disentangles the confounding effects of long memory and nonlinearities, such as change points and threshold effects. We study the asymptotic behavior of the quasi-maximum likelihood estimator and propose a sequence of Lagrange multiplier (LM) tests for nonlinearity 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. Therefore, the results in this paper are not restricted to financial volatility. Furthermore, we make no explicit assumption about the distribution of the random term in the model.

Recently, Baillie and Kapetanios (2007, 2008) have considered a similar problem. Baillie and Kapetanios (2007) construct tests for the presence of nonlinearity of unknown form in addition to a fractionally integrated, long memory component in a time series process. The tests are based on artificial neural network approximations and do not restrict the parametric form of the nonlinearity. Baillie and Kapetanios (2008) consider joint maximum likelihood estimation of long memory and Exponential Smooth Transition Autoregressive (ESTAR) models. We extend their results in different ways. First, Baillie and Kapetanios (2007) consider linearity tests only, whereas a full modeling cycle, partly based on van Dijk, Franses, and Paap (2002), Medeiros, Teräsvirta, and Rech (2006), and McAleer and Medeiros (2008a), is developed here. Second, our sequence of tests is robust to non-Gaussian disturbances. Finally, contrary to the results in Baillie and Kapetanios (2008), which are derived under specific nonlinear dynamics, we formally derive the asymptotic properties of the QMLE under general nonlinearity without assuming that the error distribution is correctly specified. In a multiple-regime model, if time is the transition variable, asymptotic theory of the QMLE cannot be achieved in the standard way, because as the sample size T goes to infinity, the proportion of finite sub-samples goes to zero. Our solution to this problem is to scale the transition variable t so that the location of the transition is a certain fraction of the total sample rather than a fixed sample point. This modification allows asymptotic theory of the QMLE; see Andrews and McDermott (1995) and Saikkonen and Choi (2004) for similar transformations.

The joint modeling of long memory and structural breaks and/or nonlinearities in realized volatility has been considered in a couple of papers. For example, Morana and Beltratti (2004) test for the existence of long memory and structural breaks in the realized volatility series of Deutsche Mark/US Dollar and Japanese Yen/US Dollar exchange rates. They conclude that realized volatility series display long memory even when structural changes are accounted for. In addition, they find that neglecting breaks is not important for very short-term forecasting once a long-memory component is included in the model, but that superior forecasts can be obtained at longer horizons by modeling both long memory and structural change. However, their model and testing procedures are different from the ones considered here. Since that paper was written, estimators of realized volatility that are more robust to microstructure noise have been developed; see the discussion in Section 1.2. McAleer and Medeiros (2008a) put forward a nonlinear heterogeneous autoregressive (HAR) model that is able to describe both long range dependence and nonlinear dynamics, such as leverage effects. Long memory is approximated by aggregation and is not explicitly modeled by fractional differencing. Scharth and Medeiros (2009) consider a multiple-regime model based on regression trees to describe high persistence in daily realized volatility series. Similarly to McAleer and Medeiros (2008a), the authors do not consider long memory directly.

Simulation results show that our modeling strategy is successful in correctly determining the structure of models in a variety of situations where long-memory and nonlinearity such as breaks or asymmetry coexist. Applying our model and testing framework to 23 stocks of the Dow Jones Industrial Average, we find evidence of structural breaks in the individual realized volatility time series. In particular, we detect transitions from high to low volatility around 2003. Dependence of volatility on the level of lagged returns is a robust finding across all stocks and in different model specifications, indicating asymmetry effects. We conclude that both long memory and non-linear effects coexist in realized volatility data. Accounting for non-linear terms in the volatility model specification yields forecast gains, as we show in a prediction experiment.

1.2. **Volatility, Long-Memory, Breaks, and Nonlinearity: A Brief Overview of the Literature.** Andersen, Bollerslev, Christoffersen, and Diebold (2007) provide a recent overview of the literature on the key role of financial volatility in risk management. However, there is an inherent problem in using models where the volatility measure plays a central role. Conditional variance is not directly observable and must, in some sense, be specified as a latent variable. Common examples of such models are the (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, model of Engle (1982) and Bollerslev (1986), various stochastic volatility models (see, for example, Taylor (1986)), and the exponentially weighted moving averages (EWMA) approach, as advocated by the Riskmetrics methodology (J. P.

Morgan 1996). McAleer (2005) gives a recent exposition of a wide range of univariate and multivariate, conditional and stochastic models of volatility, and Asai, McAleer, and Yu (2006) provide a review of the rapidly growing literature on multivariate stochastic volatility models. However, as observed by Bollerslev (1987), Malmsten and Teräsvirta (2004), and Carnero, Peña, and Ruiz (2004), among others, most of the latent volatility models have been unable to simultaneously capture several important empirical features of financial time series. For example, standard latent volatility models fail to describe adequately the slowly decreasing autocorrelation in squared returns that is associated with the high kurtosis of returns.

The search for an adequate framework for estimation and prediction of the conditional variance of financial asset returns has led to the analysis of high frequency intraday data. Merton (1980) noted that asset return variance over a fixed interval can be estimated to any degree of accuracy as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency. More recently, Andersen and Bollerslev (1998) showed that ex-post daily foreign exchange volatility is best measured by aggregating squared five-minute returns. The five-minute frequency is a trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and microstructure noise, which can arise through bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among other factors (see Madhavan (2000) and Biais, Glosten, and Spatt (2005) for recent reviews). Andersen and Bollerslev (1998), Hansen and Lunde (2005), and Patton (2005) use realized volatility to evaluate the out-of-sample forecasting performance of several latent volatility models. See also Martens, van Dijk, and de Pooter (2009), Scharth and Medeiros (2009) and McAleer and Medeiros (2008a), among others. Realized volatility can be used as a benchmark for the forecasting performance of latent variable models (Andersen and Bollerslev 1998, Hansen and Lunde 2005, Patton 2005).

Based on the results of Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002), several recent studies have documented the statistical properties of realized volatility that is constructed from high frequency data. Measurement error still remains an issue. 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 (2008), and the modified MA filter of Hansen, Large, and Lunde (2008); see McAleer and Medeiros (2008b) 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. Here 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 used in Corsi's (2009) 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 (2009). Commonly found are three regimes associated with the size and sign of past returns; see for instance, Longin (1997) and Medeiros and Veiga (2009).

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 (2008). 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.

The rest of the paper is organized as follows. Section 2 presents the model and develops the asymptotic theory of the quasi-maximum-likelihood estimation. The sequence of nonlinearity tests is introduced in Section 3. Monte Carlo evidence for its adequacy is reported in Section 4 and we describe how the test can be used in the model selection process. Empirical results are shown in Section 5. Section 6 gives some concluding remarks. All proofs are relegated to an appendix.

2. Long Memory and Nonlinearity in Realized Volatility

2.1. **Model Specification.** Set $y_t := \log(RV_t) - \mu$, where RV_t is any consistent estimator of daily integrated volatility and $\mu = \mathbb{E}[\log(RV_t)] < \infty$.[1] Consider the following long memory model with time-varying coefficients:

$$\begin{aligned} v_t &\equiv (1-L)^d y_t, \\ v_t &= \phi_0(\boldsymbol{s}_t;\boldsymbol{\xi}_0) + \phi_1(\boldsymbol{s}_t;\boldsymbol{\xi}_1)v_{t-1} + \ldots + \phi_p(\boldsymbol{s}_t;\boldsymbol{\xi}_p)v_{t-p} + \Theta(L)u_t, \end{aligned} \tag{1}$$

where $d \in (-1/2, 1/2)$ is the fractional differencing parameter, such that

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)},$$

with $\Gamma(\cdot)$ denoting the Gamma function; $\phi_i(\boldsymbol{s}_t;\boldsymbol{\xi}_i)$, $i = 0, \ldots, p$, is some nonlinear function to be specified. It is indexed by the vector of parameters $\boldsymbol{\xi}_i \in \mathbb{R}^{k_{\xi_i}}$, and $\boldsymbol{s}_t \in \mathbb{R}^{k_s}$ is a vector of state variables. The error process u_t is a martingale difference sequence. $\Theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)$ is a moving average lag polynomial of order q. Set $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_0', \boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_p'\right)' \in \mathbb{R}^{k_\xi}$. The model is indexed by the vector of parameters $\boldsymbol{\psi} = (d, \boldsymbol{\xi}', \boldsymbol{\theta}', \sigma_u^2)' \in \mathbb{R}^{k_\psi}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)' \in \mathbb{R}^q$.

2.2. **Interpretation.** The choice of the function $\phi_i(\cdot)$, $i = 1, \ldots, p$, is very flexible and allows for different specifications. The following examples list some possibilities.

Example 1 (Linear ARFIMA). *Set $\phi_i(\boldsymbol{s}_t;\boldsymbol{\xi}_i) = \phi_i$, $i = 1, \ldots, p$ and $\phi_0(\boldsymbol{s}_t;\boldsymbol{\xi}_0) = 0$. In this case, equation (1) may be written as*

$$\Phi(L)(1-L)^d y_t = \Theta(L)u_t,$$

where $\Phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$, such that y_t follows an ARFIMA(p,d,q) model. If $d = 0$, y_t is short memory. This type of specification was advocated in Andersen, Bollerslev, Diebold, and Labys (2003) to model daily realized volatility.

[1] We employ the kernel-based realized volatility estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Note that our model is specified for realized volatility (observed) and not for integrated or conditional volatility (unobserved).

EXAMPLE 2 (ARFIMA with smoothly changing parameters). *Set $\boldsymbol{s}_t = t$. Consider the following choice for the function $\phi_i(\cdot)$, $i = 0, \ldots, p$:*

$$\phi_i(\boldsymbol{s}_t; \boldsymbol{\xi}_i) = \phi_{i0} + \phi_{i1} f[\gamma(t - c)],$$

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

$$v_t = \phi_{00} + \phi_{01} f[\gamma(t-c)] + \sum_{i=1}^{p} \phi_{i0} v_{t-i} + \sum_{i=1}^{p} \phi_{i1} f[\gamma(t-c)] v_{t-i} + \Theta(L) u_t.$$

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). The model can be generalized to M transitions following Medeiros and Veiga (2003):

$$v_t = \phi_{00} + \sum_{m=1}^{M} \phi_{0m} f[\gamma_m(t - c_m)] + \sum_{i=1}^{p} \phi_{i0} v_{t-i} + \sum_{i=1}^{p} \left[\sum_{m=1}^{M} \phi_{im} f[\gamma_m(t - c_m)] \right] v_{t-i} + \Theta(L) u_t.$$

EXAMPLE 3 (ARFIMA with asymmetry). *Now let $\boldsymbol{s}_t = r_{t-1}$, where r_{t-1} is a pre-determined variable, such as past daily returns. One possibility to incorporate asymmetric effects in the model is to choose $\phi_i(\cdot)$, $i = 0, \ldots, p$, as*

$$\phi_i(\boldsymbol{s}_t; \boldsymbol{\xi}_i) = \phi_{i0} + \phi_{i1} f[\gamma(r_{t-1} - c)],$$

with $f(\cdot)$ being again the logistic function. In the case $\gamma \longrightarrow \infty$ the logistic function becomes a step function and the resulting model is related to 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), and Lundbergh and Teräsvirta (1998) for similar ideas in latent volatility models. Another possible generalization is to consider multiple regimes as in Medeiros and Veiga (2009):

$$\phi_i(\boldsymbol{s}_t; \boldsymbol{\xi}_i) = \phi_{i0} + \sum_{m=1}^{M} \phi_{im} f\left[\gamma_m \left(r_{t-1} - c_m\right)\right].$$

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 will have three regimes that can be interpreted as responding to highly negative shocks, tranquil periods, and highly positive shocks, respectively.

EXAMPLE 4 (General Nonlinear ARFIMA). *Another alternative is to leave the type of nonlinearity partially unspecified. This can be done by specifying the function $\phi_i(\cdot)$, $i = 0, \ldots, p$,* *as a single hidden layer neural network (NN) of the following form*

$$\phi_i(\boldsymbol{s}_t;\boldsymbol{\xi}_i) = \phi_{i0} + \sum_{m=1}^{M} \phi_{im} f\left[\gamma_m\left(\boldsymbol{\omega}_m'\boldsymbol{s}_t - \eta_m\right)\right], \tag{2}$$

where $f(\cdot)$ *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 and Veiga (2005).

In a related paper, Martens, van Dijk, and de Pooter (2009) suggested a model to describe jointly long-range dependence, nonlinearity, structural breaks, and the effects of days of the week. The model considered in their paper is nested in specification (1).

2.3. **Parameter Estimation.**

2.3.1. *Time Transformation.* In this paper, we employ a simple time transformation to analyze model (1). Let T_0 be the size of a given data sample. Set $\boldsymbol{V}_{t-1} = (v_{t-1}, \ldots, v_{t-p})'$ and, for any sequence $\{x_t\}$, $t = 1, \ldots, T$, define $x_{tT} := (T_0/T)x_t$. Then, model (1) is embedded in a sequence of models:

$$\Phi_{tT}(L)v_{tT} = \phi_{0,tT} + \Theta(L)u_t, \tag{3}$$

where $\phi_{0,tT} \equiv \phi_0(\boldsymbol{s}_{tT}, \boldsymbol{\xi}_0)$ and

$$\Phi_{tT}(L) = 1 - \phi_1(\boldsymbol{s}_{tT};\boldsymbol{\xi}_1)L - \ldots - \phi_p(\boldsymbol{s}_{tT};\boldsymbol{\xi}_p)L^p.$$

Without this transformation, parameter regimes of finite length become unidentified as $T \to \infty$. The transformation allows for a proper scaling of the logistic function such that all regimes remain identified. Consider the logistic function under the transformation:

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

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. In this sense, the time transformation is the smooth equivalent of the assumption of constant break fractions in the change-point literature, e.g. Perron (1989). Similar transformations are used in Saikkonen and Choi (2004) and Andrews and McDermott (1995).

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

$$\boldsymbol{\psi}_* = \left(d_*, \boldsymbol{\xi}_*', \boldsymbol{\theta}_*', \sigma_{u,*}^2\right)'.$$

Define $u_t(\boldsymbol{\psi}) = \Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT} - \phi_{0,tT}\right]$. We use the shorthand notation $u_{t,*} := u_t(\boldsymbol{\psi}_*)$ for the data-generating errors and u_t for $u_t(\boldsymbol{\psi})$. Note that the fractional integration parameter, d, is an element of $\boldsymbol{\psi}$ and is 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 $\boldsymbol{\psi}_* \in \mathbb{R}^{k_\psi}$ is an interior point of $\boldsymbol{\Psi} \subset \mathbb{R}^{k_\psi}$, a compact parameter space.*

ASSUMPTION 2 (Errors).

(1) *$u_{t,*}$ is a martingale difference sequence with mean zero and constant unconditional variance $\sigma_{u,*}^2 > 0$.*
(2) $\mathbb{E}|u_{t,*}|^q < \infty$ *for* $q = 1, \ldots, 4$.
(3) $\mathbb{E}\left[\exp(u_{t,*})^q\right] < \infty$ *for* $q = 1, \ldots, 4$.

ASSUMPTION 3 (Stationarity and Moments).

(1) $\mathbb{E}|\boldsymbol{z}_{tT}|^q < \infty$, $q = 1, \ldots, 4$, *where* $\boldsymbol{z}_{tT} = (v_{tT}, \boldsymbol{s}_{tT}')'$.
(2) $d_* \in (-1/2, 1/2)$.
(3) $\Theta(L)$ *is invertible.*

ASSUMPTION 4 (Autoregressive Transition Function).

(1) *The transition functions are parameterized such that they are well defined.*
(2) $\mathbb{E}|\phi_{0,tT}|^q < \infty$, $q = 1, \ldots, 4$.
(3) *For all $\boldsymbol{s}_t$, $\boldsymbol{\xi}$, the roots of $\Phi_{tT,*}(\boldsymbol{s}_t; \boldsymbol{\xi})$ are outside the unit circle and the inverse lag polynomial $\Phi_{tT,*}^{-1}(\boldsymbol{s}_t; \boldsymbol{\xi})$ exists.*
(4) $\mathbb{E}|\Phi_{tT,*}^{-1}(L)\Theta_*(L)u_{t,*}|^q < \infty$, $q = 1, \ldots, 4$.
(5) $\mathbb{E}\left|\frac{\partial}{\partial \boldsymbol{\xi}}\Phi_{tT}(L)v_{tT}\right|^q < \infty$, $q = 1, \ldots, 4$.
(6) $\mathbb{E}\left|\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'}\Phi_{tT}(L)v_{tT}\right|^q < \infty$, $q = 1, \ldots, 4$.
(7) $\mathbb{E}\left|\frac{\partial}{\partial \boldsymbol{\xi}}\phi_{0,tT}\right|^q < \infty$, $q = 1, \ldots, 4$.
(8) $\mathbb{E}\left|\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'}\phi_{0,tT}\right|^q < \infty$, $q = 1, \ldots, 4$.

EXAMPLE 5 (for Assumption 4 (1): Logistic Transition). *If there are $M + 1$ different regimes of volatility depending on a state variable s_t (for example past returns r_{t-1} or time t), with* *transitions governed by logistic functions, then the transition parameters* c_m *and* γ_m, $m = 1, \ldots, M$, *are such that*

(1) $-\infty < -C < c_1 < \ldots < c_M < C < \infty$.
(2) $\gamma_m > 0$ *for all* m.
(3) $f[\gamma_1(s_t - c_1)] \geq f[\gamma_2(s_t - c_2)] \geq \ldots \geq f[\gamma_M(s_t - c_M)]$.

2.3.3. *Quasi-Maximum Likelihood Estimator.* The martingale difference sequence assumption on the errors $u_{t,*}$ implies that 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

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

The parameter vector is estimated by quasi-maximum likelihood as

$$\widehat{\boldsymbol{\psi}}_T = \operatorname*{argmax}_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathcal{L}_T(\boldsymbol{\psi}),$$

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

THEOREM 1 (Consistency). *Under Assumptions 1 through 4, the quasi-maximum likelihood estimator* $\widehat{\boldsymbol{\psi}}_T$ *is consistent:*

$$\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_*.$$

The proof is provided in the Appendix.

THEOREM 2 (Asymptotic Normality). *Under Assumptions 1 through 4, the quasi-maximum likelihood estimator* $\widehat{\boldsymbol{\psi}}_T$ *is asymptotically normally distributed:*

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

where

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

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

The proof is provided in the Appendix.

PROPOSITION 1 (Covariance Matrix Estimation). *Under Assumptions 1 through 4,*

$$\boldsymbol{A}_T\left(\widehat{\boldsymbol{\psi}}_T\right) \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\psi}_*), \quad \boldsymbol{B}_T\left(\widehat{\boldsymbol{\psi}}_T\right) \xrightarrow{p} \boldsymbol{B}(\boldsymbol{\psi}_*),$$

where

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

and

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

The proof is provided in the Appendix.

3. MODEL SPECIFICATION

3.1. **Test Statistic.** In this section, we propose a sequence of tests for misspecification in the sense of remaining nonlinearity. We advocate the use of a multiple regime specification, as in Section 2 Examples 2–4. The testing procedure will be partially based on the results of Medeiros and Veiga (2005) and Baillie and Kapetanios (2007). 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 $q > 0$. Consider the following model.

$$\begin{aligned}
v_t &= \boldsymbol{\phi}_0'\boldsymbol{V}_{t-1} + \sum_{m=1}^{M}\boldsymbol{\phi}_m'\boldsymbol{V}_{t-1}f\left[\gamma_m\left(s_t - c_m\right)\right] + u_t \\
&= \boldsymbol{\phi}_0'\boldsymbol{V}_{t-1} + \sum_{m=1}^{M^*}\boldsymbol{\phi}_m'\boldsymbol{V}_{t-1}f\left[\gamma_m\left(s_t - c_m\right)\right] \\
&\quad + \sum_{m=M^*+1}^{M}\boldsymbol{\phi}_m'\boldsymbol{V}_{t-1}f\left[\gamma_m\left(s_t - c_m\right)\right] + u_t,
\end{aligned} \tag{4}$$

where $\boldsymbol{V}_{t-1} = (1, v_{t-1}, \ldots, v_{t-p})$.

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

$$\mathbb{H}_0 : \gamma_{M^*+1} = \gamma_{M^*+2} = \cdots = \gamma_M = 0. \tag{5}$$

Model (4) is only identified under the alternative, which means that standard asymptotic inference is not available. This problem is circumvented, as in Teräsvirta (1994), by expanding $f\left[\gamma_m\left(s_t - c_m\right)\right]$, $m = M^*+1, \ldots, M$, into a Taylor series around the null hypothesis (5). The order of the expansion is a compromise between a small approximation error (high order) and availability of data (as short time series necessarily imply a relatively low order). Using a third-order Taylor expansion and rearranging terms results in the following model:

$$v_t = \widetilde{\boldsymbol{\phi}}_0' \boldsymbol{V}_{t-1} + \sum_{m=1}^{M^*} \widetilde{\boldsymbol{\phi}}_m' \boldsymbol{V}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right] + \boldsymbol{\rho}_1' \boldsymbol{V}_{t-1} s_t + \boldsymbol{\rho}_2' \boldsymbol{V}_{t-1} s_t^2 + \boldsymbol{\rho}_3' \boldsymbol{V}_{t-1} s_t^3 + u_t^*, \quad (6)$$

where $u_t^* = u_t + R_3$ and R_3 is the remainder in the Taylor expansion.

The null hypothesis (5) is then approximated by

$$\mathbb{H}_0: \boldsymbol{\rho}_1 = \boldsymbol{\rho}_2 = \boldsymbol{\rho}_3 = \mathbf{0}.$$

Under the null, $R_3(\boldsymbol{z}_t; \boldsymbol{\xi}) = 0$. We can use (6) to test for absence of remaining nonlinearity. The local approximation of the log-density for observation t takes the form

$$\begin{aligned}\ell_t(\boldsymbol{\psi}) = & -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_u^2) \\ & -\frac{1}{2\sigma_u^2} \times \left\{ v_t - \widetilde{\boldsymbol{\phi}}_0' \boldsymbol{V}_{t-1} - \sum_{m=1}^{M^*} \boldsymbol{\phi}_m' \boldsymbol{V}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right] \right. \\ & \left. - \boldsymbol{\rho}_1' \boldsymbol{V}_{t-1} s_t - \boldsymbol{\rho}_2' \boldsymbol{V}_{t-1} s_t^2 - \boldsymbol{\rho}_3' \boldsymbol{V}_{t-1} s_t^3 \right\}^2 .\end{aligned} \quad (7)$$

Since the information matrix is block diagonal, the error variance σ_u^2 can be treated as fixed. The partial derivatives of (7) evaluated at the estimated parameter vector under the null hypothesis are:

$$
\begin{aligned}
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial d}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= -\frac{\widehat{u}_t}{\sigma_u^2}\frac{\partial}{\partial d}\left\{\widehat{v}_t - \widehat{\boldsymbol{\phi}}_0'\widehat{\boldsymbol{V}}_{t-1} - \sum_{m=1}^{M^*}\widehat{\boldsymbol{\phi}}_m'\widehat{\boldsymbol{V}}_{t-1} f\left[\gamma_m\left(s_t - c_m\right)\right]\right\}\\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \widetilde{\boldsymbol{\phi}}_0}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= -\frac{\widehat{u}_t}{\widehat{\sigma}_u^2}\widehat{\boldsymbol{V}}_{t-1};\\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\phi}_m}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= -\frac{\widehat{u}_t}{\widehat{\sigma}_u^2}\widehat{\boldsymbol{V}}_{t-1} f\left[\widehat{\gamma}_m\left(s_t - \widehat{c}_m\right)\right];\\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \gamma_m}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= -\frac{\widehat{u}_t}{\widehat{\sigma}_u^2}\widehat{\boldsymbol{\phi}}_m'\widehat{\boldsymbol{V}}_{t-1} f\left[\widehat{\gamma}_m\left(s_t - \widehat{c}_m\right)\right]\left\{1 - f\left[\widehat{\gamma}_m\left(s_t - \widehat{c}_m\right)\right]\right\}\left(s_t - \widehat{c}_m\right);\\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial c_m}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= \frac{\widehat{u}_t}{\widehat{\sigma}_u^2}\widehat{\boldsymbol{\phi}}_m'\widehat{\boldsymbol{V}}_{t-1} f\left[\widehat{\gamma}_m\left(s_t - \widehat{c}_m\right)\right]\left\{1 - f\left[\widehat{\gamma}_m\left(s_t - \widehat{c}_m\right)\right]\right\};\\
\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\rho}_j}\right|_{\mathbb{H}_0,\widehat{\boldsymbol{\psi}}} &= -\frac{\widehat{u}_t}{\widehat{\sigma}_u^2}\widehat{\boldsymbol{V}}_{t-1}s_t^j,\ j = 1, 2, 3;
\end{aligned}
$$

where $\widehat{u}_t$ is the residual estimated under the null, $\widehat{\boldsymbol{V}}_{t-1} = (1, \widehat{v}_{t-1}, \ldots, \widehat{v}_{t-p})'$, $\widehat{v}_{t-i} = (1 - L)^{\widehat{d}} y_{t-i}$, $i = 1, \ldots, p$, 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

$$
LM = \sum_{t=1}^{T}\widehat{\boldsymbol{q}}_t'\left(\sum_{t=1}^{T}\widehat{\boldsymbol{q}}_t\widehat{\boldsymbol{q}}_t'\right)^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{q}}_t, \tag{8}
$$

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

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

and

$$
\widehat{\boldsymbol{q}}_{a,t} = \left[\widehat{\boldsymbol{V}}_{t-1}'s_t, \widehat{\boldsymbol{V}}_{t-1}'s_t^2, \widehat{\boldsymbol{V}}_{t-1}'s_t^3\right]'.
$$

Under standard regularity conditions and the additional assumption $\mathbb{E}|s_t|^{\delta} < \infty$, for some $\delta > 6$, (8) has an asymptotic χ^2 distribution with $m = 3(p + 1)$ degrees of freedom. Defining

$\boldsymbol{\iota} = (1, 1, \ldots, 1)' \in \mathbb{R}^T$ and

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

the LM statistic can be written as

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

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

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

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

is not met. This has an adverse effect on the empirical size of the test. To circumvent this problem, we regress the residuals $\widehat{u}_t$ on $\widehat{\boldsymbol{q}}_{0,t}$ (Eitrheim and Teräsvirta 1996). Finally, we compute a new sequence of residuals $\widetilde{u}_t$ from this regression.

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

(3) Compute the χ^2 statistic

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

We now combine the procedure above into a coherent modeling strategy that involves a sequence of LM tests. The idea is to test a linear model against an alternative model with one nonlinear term at a λ_1-level of significance. In the event that the null hypothesis is rejected, one logistic term is added, the nonlinear model is re-estimated, and then tested against an alternative with more than one nonlinear term. The procedure continues testing J logistic terms against alternative models with $J^* \geq J + 1$ terms at significance level $\lambda_J = \lambda_1 \, C^{J-1}$ for some arbitrary constant $0 < C < 1$. The testing sequence is terminated at the first non-rejection outcome. The number of nonlinear terms, M, is estimated by $\widehat{M} = \bar{J} - 1$, where $\bar{J}$ is the number of rejections prior to the first non-rejection. By reducing the significance level at each step of the sequence, it is possible to control the overall level of significance, and hence to avoid excessively large models. The Bonferroni procedure ensures that such a sequence of LM tests is consistent, and that $\sum_{J=1}^{\bar{J}} \lambda_J$ acts as an upper bound on the overall level of significance. As for the determination of the arbitrary constant C, it is good practice to perform the sequential testing procedure with different values of C to avoid selecting models that are too parsimonious. Finally, equation (6) combined with the use of some information criterium, such as the AIC or BIC, can be used to determine the lag structure of the model.

4. Monte-Carlo Evidence

The purpose of this section is to evaluate the performance of the modeling cycle strategy described in the previous section. We simulate 1000 replications of the models below with $T = 500$, 1000, and 5000 observations. We report both descriptive statistics for the parameter estimates under correct specification, assuming that M and p are known, as well as the frequency of correct specification in the case where both M and p are endogenously determined from the simulated data. The DGPs considered are as follows:

(1) Model 1: Short-memory linear model

$$y_t = 0.8y_{t-1} + u_t, \tag{9}$$

where $u_t \sim \mathsf{NID}(0, 0.5)$.

(2) Model 2: Short-memory nonlinear model I

$$y_t = 0.8y_{t-1} - 0.4y_{t-1}f[30(t/T - 0.35)] + 0.4y_{t-1}f[30(t/T - 0.65)] + u_t, \tag{10}$$

where $u_t \sim \mathsf{NID}(0, 0.5)$.

(3) Model 3: Short-memory nonlinear model II

$$\begin{aligned} y_t &= 0.8y_{t-1} - 0.4y_{t-1}f[5(r_{t-1} + 1)] + 0.4y_{t-1}f[5(r_{t-1} - 1)] + u_t, \\ r_t &= [\exp(y_t) + \nu_t]\,\varepsilon_t \end{aligned} \tag{11}$$

where $u_t \sim \mathsf{NID}(0, 0.1)$, $\varepsilon_t \sim \mathsf{NID}(0, 1)$, and $\nu_t \sim \mathsf{NID}(0, 0.0001)$.

(4) Model 4: Long-memory linear model

$$\begin{aligned} v_t &= 0.8v_{t-1} + u_t, \\ y_t &= (1 - L)^{-0.4}v_t, \end{aligned} \tag{12}$$

where $u_t \sim \mathsf{NID}(0, 0.5)$.

(5) Model 5: Long-memory nonlinear model I

$$\begin{aligned} v_t &= 0.8v_{t-1} - 0.4v_{t-1}f[30(t/T - 0.35)] + 0.4v_{t-1}f[30(t/T - 0.65)] + u_t, \\ y_t &= (1 - L)^{-0.4}v_t, \end{aligned} \tag{13}$$

where $u_t \sim \mathsf{NID}(0, 0.5)$.

(6) Model 6: Long-memory nonlinear model II

$$
\begin{aligned}
v_t &= 0.8v_{t-1} - 0.4v_{t-1}f[5(r_{t-1}+1)] + 0.4v_{t-1}f[5(r_{t-1}-1)] + u_t,\\
r_t &= [\exp(y_t) + \nu_t]\,\varepsilon_t,\\
y_t &= (1-L)^{-0.4}v_t,
\end{aligned} \tag{14}
$$

where $u_t \sim \mathsf{NID}(0, 0.1)$, $\varepsilon_t \sim \mathsf{NID}(0, 1)$, and $\nu_t \sim \mathsf{NID}(0, 0.0001)$.

Model 1 is a simple short-memory autoregressive model and is important to check if our procedure is able to detect very parsimonious specifications. Models 2 and 3 are both nonlinear short-memory processes, while Models 4–6 are all long-memory specifications. Models 5 and 6 are the two most general specifications, with both long-memory and nonlinearity. The specification and estimation results are reported in Tables 1 and 2. Table 1 shows the average bias and the mean squared error (MSE) of the parameter estimates under the assumption of correct model specification, i.e., correct number of regimes (M) and autoregressive order (p). Apart from the γ parameter, all estimates are reasonable. We only report results for $T = 500$ and $T = 1000$ in Table 1, since the results for $T = 5000$ are only marginally different from those for $T = 1000$. In order to evaluate the performance of the modeling strategy proposed in this paper, we also check the frequency of correct specification when the lag and regime structures are unknown. The number of regimes is determined by the sequence of LM tests described earlier, while the autoregressive order is determined by the BIC. The results are reported in Table 2. The proposed procedure always finds the correct number of lags and the performance concerning the choice of the number of regimes improves as the sample size increases.

5. Empirical Application

5.1. **Data.** We use high frequency tick-by-tick quotes on 23 Dow Jones Industrial Average Index stocks: Alcoa (aa), American International Group (aig), American Express (axp), Boeing (ba), Caterpillar (cat), Du Pont (dd), Walt Disney (dis), General Electric (ge), General Motors (gm), Home Depot (hd), Honeywell (hon), 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), United Tech (utx), and Wal-Mart Stores (wmt). The data were obtained from the NYSE TAQ database and they cover the period January 3, 1995 up to December 31,2005.

In calculating 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 (2008). 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 (2008) we use a 60-second activity fixed tick time sampling scheme such that we obtain the same number of observations each day.

5.2. **Model Specification and Estimation.** For each stock return time series, we use the first 2170 observations to estimate the model and leave the remaining 600 observations for out-of-sample forecasting. We consider the following alternative specifications:

(1) linear ARFIMA;
(2) nonlinear ARFIMA with time as transition variable;
(3) nonlinear ARFIMA with past daily return as transition variable;
(4) nonlinear ARFIMA with past cumulated return over five days as transition variable;
(5) nonlinear ARFIMA with past cumulated return over ten days as transition variable;
(6) nonlinear ARFIMA with past cumulated return over 22 days as transition variable;
(7) nonlinear ARFIMA with past cumulated return over 66 days as transition variable;
(8) nonlinear ARFIMA with past cumulated return over 252 days as transition variable;
(9) nonlinear ARFIMA with past average volatility over one day as transition variable;
(10) nonlinear ARFIMA with past average volatility over five days as transition variable;
(11) nonlinear ARFIMA with past average volatility over 22 days as transition variable;
(12) the heterogeneous autoregressive (HAR-RV) model of Corsi (2009).

Table 3 displays estimation results. The table reports the estimated fractional difference parameter d, the estimated autoregressive order p, and the number of regimes M determined by the sequence of LM tests. As an example, Figure 1 shows the estimated transition functions for three stocks (AIG, GM, and IBM), where lagged daily returns are used as transition variable. In each case, we observe a different pattern. For AIG, there are two regimes and, surprisingly, the transition is not around zero. For GM there are two sharp transitions. The first one occurs at past returns around -2 and the other one at small positive returns. On the other hand, the two transitions that we observe in the case of IBM are quite smooth, especially the one associated with positive returns.

5.3. **Forecasting Exercise.** Table 4 presents results for one-day-ahead forecasts for the last 600 observations of each time series of realized volatility. The table shows the ratio of the mean squared one-step-ahead forecast error (MSFE) from the different linear and nonlinear specifications (numerator) and the benchmark HAR-RV model (denominator). Numbers below one indicate that the HAR-RV model is outperformed by the competing specification.

The stars indicate that the considered model significantly improves over the HAR-RV model according to the Diebold-Mariano test (one star represents 10%, two stars 5%, and three stars 1% significance level). Since the linear and nonlinear ARFIMA models are not nested in the HAR-RV model, the Diebold-Mariano test is valid. From Table 4 it is clear that the nonlinear model (3) with past daily returns as transition variable is the one that systematically outperforms the benchmark HAR-RV specification.

6. 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 accommodates long memory and a general non-linear function that may include transitions between parameter regimes and asymmetry effects.

We show consistency and asymptotic normality of the quasi-maximum likelihood estimator. Asymptotic theory requires a time transformation that ensures that regimes of finite length remain identified as the sample size grows to infinity.

We propose 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 a Taylor expansion of the nonlinear function in the volatility equation.

Once the type of nonlinearity and the relevant variables are identified, the full specification is estimated using realized volatility of stocks in the Dow Jones Industrial Average. We find strong evidence for nonlinear effects driven by time and past returns in all time series. The results indicate that long memory and leverage effects in a wide sense, i.e. dependence on linear combinations of past returns, coexist in realized volatility data. A forecast horse race indicates that a specification with long memory and asymmetry can outperform the standard HAR-RV model.

APPENDIX A. PROOF OF CONSISTENCY

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

(1) $\boldsymbol{\Psi}$ is a compact parameter set.
(2) $\mathcal{L}_T(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi}$ and measurable in u_t.
(3) As $T \to \infty$, $\mathcal{L}_T(\boldsymbol{\psi})$ converges in probability to a deterministic function $\mathcal{L}(\boldsymbol{\psi})$ uniformly on $\boldsymbol{\Psi}$.
(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 quasi-likelihood function and the construction of u_t. Item (3) holds by the Ergodic Theorem if $\mathbb{E}\left[\sup|\ell_t(\boldsymbol{\psi})|\right] < \infty$. The latter holds by the Jensen´s inequality and $\mathbb{E}\left[\sup|g(\cdot,\boldsymbol{\psi})|\right] < \infty$, where g denotes the normal density function. The finiteness of the last expression follows from the definition of the normal density as long as $\sigma_u^2 > 0$.

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

$$\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \mathbb{E}\left[\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_*)\right].$$

Now,

$$\begin{aligned}
\mathbb{E}\left[\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_*)\right] &= \mathbb{E}\left\{\log\left[\frac{g(u_t,\boldsymbol{\psi})}{g(u_t,\boldsymbol{\psi}_*)}\right]\right\}, \\
&= \mathbb{E}\left[-\frac{1}{2}\log\frac{\sigma_u^2}{\sigma_{u,*}^2} - \frac{1}{2}\left(\frac{u_t^2}{\sigma_u^2} - \frac{u_{t,*}^2}{\sigma_{u,*}^2}\right)\right], \\
&= -\frac{1}{2}\log\frac{\sigma_u^2}{\sigma_{u,*}^2} - \frac{1}{2}\left[\mathbb{E}(u_t^2\sigma_u^{-2}) - 1\right].
\end{aligned}$$

Next, we show that$\mathbb{E}\left[u_t^2(\boldsymbol{\psi})\right] \geq \mathbb{E}\left(u_{t,*}^2\right) = \sigma_{u,*}^2$ and that the expressions attain their respective lower bounds at $\boldsymbol{\psi} = \boldsymbol{\psi}_*$ uniquely. Consider

$$\begin{aligned}
\mathbb{E}\left[u_t^2(\boldsymbol{\psi})\right] &= \mathbb{E}\left\{\Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT} - \phi_{0,tT}\right]\right\}^2, \\
&= \mathbb{E}\left(\Theta^{-1}(L)\left\{\Phi_{tT}(L)\Phi_{tT,*}^{-1}(L)\left[\phi_{0,tT,*} + \Theta_*(L)u_{t,*}\right] - \phi_{0,tT}\right\}\right)^2 \\
&\geq \mathbb{E}\left(u_{t,*}^2\right) = \sigma_{u,*}^2,
\end{aligned}$$

and therefore, $\mathbb{E}\left[u_t^2(\boldsymbol{\psi})\right]$ attains its minimum of $\sigma_{u,*}^2$ uniquely at $\boldsymbol{\psi} = \boldsymbol{\psi}_*$ under Assumption 2.

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

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

at $d = d_*$, $\Theta(L) = \Theta_*(L)$, $\boldsymbol{\xi} = \boldsymbol{\xi}_*$. 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 $\sigma_u^2 = \sigma_{u,*}^2$. This shows that $\mathbb{E}\left[\ell_t(\boldsymbol{\psi}) - \ell_t(\boldsymbol{\psi}_*)\right]$ is uniquely maximized at $\boldsymbol{\psi} = \boldsymbol{\psi}_*$. □

APPENDIX B. 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*

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

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,*

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

for some $K < \infty$.

LEMMA 1. *Under Assumptions 2-4, the sequence $\left\{\left.\frac{\partial \ell_t}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_*}, \mathcal{F}_t\right\}_{t=1,\ldots,T}$ is a stationary martingale difference sequence.*

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

$$\mathbb{E}\left(\left.\frac{\partial \ell_t}{\partial d}\right|\mathcal{F}_{t-1}\right) = \mathbb{E}\left[\left.-\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial}{\partial d}(1-L)^d y_{tT}\right|\mathcal{F}_{t-1}\right] = 0,$$

since u_t has mean zero, and $\frac{\partial}{\partial d}(1-L)^d y_{tT}$ does not contain u_t.

$$\mathbb{E}\left(\left.\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\right|\mathcal{F}_{t-1}\right) = \mathbb{E}\left\{\left.-\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\left[\frac{\partial}{\partial \boldsymbol{\xi}}\Phi_{tT}(L)v_{tT} + \frac{\partial}{\partial \boldsymbol{\xi}}\phi_{0,tT}\right]\right|\mathcal{F}_{t-1}\right\} = 0,$$

since $\Phi_{tT}(L)v_{tT}$ and $\phi_{0,tT}$ are uncorrelated with u_t.

$$\mathbb{E}\left(\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}\middle|\mathcal{F}_{t-1}\right) = \mathbb{E}\left\{-\frac{u_t}{\sigma_u^2}\frac{\partial}{\partial \boldsymbol{\theta}}\Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT}+\phi_{0,tT}\right]\middle|\mathcal{F}_{t-1}\right\} = 0,$$

since $\frac{\partial}{\partial \boldsymbol{\theta}}\Theta^{-1}(L)\left[\Phi_{tT}(L)v_{tT}+\phi_{0,tT}\right]$ 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. *Under Assumptions 2-4,*

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

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 subscribed by an asterisk.

We will consider the gradient vector element by element:

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\mathbb{E}\left|\frac{\partial \ell_t}{\partial d}\right| = \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\mathbb{E}\left|-\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial}{\partial d}(1-L)^d y_{tT}\right|.$$

Using the Cauchy-Schwarz inequality, we need to find upper bounds for the following objects $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\mathbb{E}\left|\frac{\partial}{\partial d}(1-L)^d y_{tT}\right|^p$ and $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\mathbb{E}\left|u_t(\boldsymbol{\psi})\right|^p$, $p=1,2$.

First, note that

$$\begin{aligned}\mathbb{E}\left|(1-L)^d y_{tT}\right|^q &= \mathbb{E}\left|(1-L)^d\left\{(1-L)^{-d_*}\Phi_{tT,*}^{-1}(L)\left[\phi_{0,tT,*}+\Theta_*(L)u_{t,*}\right]\right\}\right|^q \\ &= \mathbb{E}\left|(1-L)^{d-d_*}\Phi_{tT,*}^{-1}(L)\left[\phi_{0,tT,*}+\Theta_*(L)u_{t,*}\right]\right|^q < \infty,\end{aligned}$$

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

Then,

$$\begin{aligned}&\mathbb{E}\left|\frac{\partial}{\partial d}(1-L)^d y_{tT}\right|^q \\ &= \mathbb{E}\left|\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)\prod_{i=0}^{j-1}(d-i)L^j y_{tT}\right|^q, \\ &= \mathbb{E}\left|\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\sum_{i=0}^{j-1}\frac{1}{d-i}\right)\prod_{i=0}^{j-1}(d-i)L^j(1-L)^{-d_*}\Phi_{tT,*}^{-1}(L)\left[\phi_{0,tT,*}+\Theta_*(L)u_{t,*}\right]\right|^q \\ &< \infty,\end{aligned}$$

from the same set of assumptions and recognizing that $\frac{\partial}{\partial d}(1-L)^d y_{tT}$ is stationary if $d \in (-1/2, 1/2)$.

Now, note that

$$\begin{aligned}
\mathbb{E}\left|u_t(\boldsymbol{\psi})\right|^p &= \mathbb{E}\left|\Theta^{-1}(L)\left[\Phi_{tT}(L)(1-L)^d y_{tT} - \phi_{0,tT}\right]\right|^p, \\
&= \mathbb{E}\left|\Theta^{-1}(L)\left[\Phi_{tT}(L)(1-L)^{d-d_*}\Phi^{-1}_{tT,*}(L)\left[\Theta_*(L)u_{t,*} + \phi_{0,tT,*}\right] - \phi_{0,tT}\right]\right|^p < \infty
\end{aligned}$$

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

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

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \boldsymbol{\xi}}\right| &= \mathbb{E}\left|-\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\left[\frac{\partial}{\partial \boldsymbol{\xi}}\Phi_{tT}(L)v_{tT} + \frac{\partial}{\partial \boldsymbol{\xi}}\phi_{0,tT}\right]\right| \\
&\leq \left(\mathbb{E}\left|-\frac{u_t}{\sigma_u^2}\right|^p\right)^{\frac{1}{p}}\left\{\mathbb{E}\left|\Theta^{-1}(L)\left[\frac{\partial}{\partial \boldsymbol{\xi}}\Phi_{tT}(L)v_{tT} + \frac{\partial}{\partial \boldsymbol{\xi}}\phi_{0,tT}\right]\right|^p\right\}^{\frac{1}{p}} < \infty
\end{aligned}$$

by Assumption 4.

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial \ell_t}{\partial \theta_i}\right| &= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\frac{\partial \Theta^{-1}(L)}{\partial \theta_i}\left[\Phi_{tT}(L)v_{tT} - \phi_{0,tT}\right]\right|, \\
&= \mathbb{E}\left|\frac{u_t}{\sigma_u^2}\left[-\frac{L^i}{\Theta^2(L)}\right]\left[\Phi_{tT}(L)v_{tT} - \phi_{0,tT}\right]\right|, \\
&\leq \left(\mathbb{E}\left|\frac{u_t}{\sigma_u^2}\right|^p\right)^{\frac{1}{p}}\left\{\mathbb{E}\left|\left[-\frac{L^i}{\Theta^2(L)}\right]\left[\Phi_{tT}(L)v_{tT} - \phi_{0,tT}\right]\right|^p\right\}^{\frac{1}{p}} < \infty
\end{aligned}$$

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

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

This shows statement (1) of Lemma 2. Statement (2) of Lemma 2 follows the same arguments, except 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*

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

where

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

is absolutely uniformly integrable:

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

it is continuous in $\boldsymbol{\psi}$ and has zero mean: $\mathbb{E}\left[h_t(\boldsymbol{\psi})\right] = 0$.

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 $\boldsymbol{A}$. By the triangular inequality, showing absolute uniform integrability reduces to showing that

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

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 $\boldsymbol{A}(\cdot)$; 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

$$\begin{aligned}\frac{\partial^2}{\partial d^2}(1-L)^d &= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left[\left(\sum_{i=0}^{j-1} \frac{1}{d-i}\right)^2 - \sum_{i=0}^{j-1} \left(\frac{1}{d-i}\right)^2\right] \prod_{i=0}^{j-1} (d-i) L^j, \qquad (15) \\ &= \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \left[\sum_{\substack{i,k=0 \\ i \neq k}}^{j-1} \frac{1}{(d-i)(d-k)}\right] \prod_{i=0}^{j-1} (d-i) L^j.\end{aligned}$$

Then, we have

$$\begin{aligned}\frac{\partial^2 \ell_t}{\partial d^2} =& -\frac{1}{\sigma_u^2}\left[\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial}{\partial d}(1-L)^d y_{tT}\right]^2 \\ & -\frac{u_t}{\sigma_u^2}\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial^2}{\partial d^2}(1-L)^d y_{tT} \\ & =: R_1 + R_2.\end{aligned}$$

We first show that $\mathbb{E} \sup |R_i| < \infty$ for $i = 1, 2$.

$$|R_1| = \left|\frac{1}{\sigma_u^2}\left[\Theta^{-1}(L)\Phi_{tT}(L)\frac{\partial}{\partial d}(1-L)^d y_{tT}\right]^2\right|,$$

and

$$|R_2| = \left| \frac{u_t}{\sigma_u^2} \Theta^{-1}(L) \left[\Phi_{tT}(L) \frac{\partial^2}{\partial d^2} (1-L)^d y_{tT} \right] \right|.$$

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 $\mathbb{E} \sup |R_i| < \infty$, $i = 1, 2$, by the Lebesgue Dominated Convergence Theorem. Thus,

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

□

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 (Theorem 1). Then,

$$\boldsymbol{B}(\boldsymbol{\psi}_*)^{-\frac{1}{2}} \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} \left. \frac{\partial \ell_t}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}_*} \Rightarrow \boldsymbol{W}(r),\ r \in [0,1],$$

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

$$\boldsymbol{A}_T(\widehat{\boldsymbol{\psi}}_T) \xrightarrow{p} \boldsymbol{A}(\boldsymbol{\psi}_*)$$

for any sequence $\widehat{\boldsymbol{\psi}}_T \xrightarrow{p} \boldsymbol{\psi}_*$,

$$\boldsymbol{A}_T(\widehat{\boldsymbol{\psi}}_T) = -\frac{1}{T} \sum_{t=1}^{T} \left. \frac{\partial^2 \ell_t}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right|_{\widehat{\boldsymbol{\psi}}_T},$$

and

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

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 $\boldsymbol{\psi}_*$ and (b) uniform convergence of $\boldsymbol{A}_T$ to $\boldsymbol{A}$ in probability, i.e.

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

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

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

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

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

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

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

As we did for $\boldsymbol{A}$ in Lemma 3, we have to show that h_t is absolutely uniformly integrable, continuous in $\boldsymbol{\psi}$, and has expected value $\mathbb{E} m_t(\boldsymbol{\psi}) = 0$. 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 m_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, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

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.

ASAI, M., M. MCALEER, AND J. YU (2006): "Multivariate stochastic volatility: A review," *Econometric Reviews*, 25, 145–175.

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

BAILLIE, R., AND G. KAPETANIOS (2007): "Testing for Neglected Nonlinearity in Long-Memory Models," *Journal of Business and Economic Statistics*, 25, 447–461.

——— (2008): "Nonlinear Models for Strongly Dependent Processes with Financial Applications," *Journal of Econometrics*, 147, 60–71.

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

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.

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.

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.

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. (2009): "A Simple Long Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196.

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.

EITRHEIM, Ø., AND T. TERÄSVIRTA (1996): "Testing the Adequacy of Smooth Transition Autoregressive Models," *Journal of Econometrics*, 74, 59–75.

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

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.

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, P., J. LARGE, AND A. LUNDE (2008): "Moving average-based estimators of integrated variance," *Econometric Reviews*, 27, 79–111.

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.

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

HILLEBRAND, E., AND M. MEDEIROS (2008): "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, pp. 303–328. Elsevier.

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 (2009): "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," *International Journal of Forecasting*, 27, 282–303.

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

MCALEER, M., AND M. MEDEIROS (2008a): "A Multiple Regime Smooth Transition Heterogenous Autoregressive Model for Long Memory and Asymmetries," *Journal of Econometrics*, 147, 104–119.

——— (2008b): "Realized Volatilty: A Review," *Econometric Reviews*, 27, 10–45.

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

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

——— (2005): "A Flexible Coefficient Smooth Transition Time Series Model," *IEEE Transactions on Neural Networks*, 16, 97–113.

——— (2009): "Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH(1,1) Model," *Econometric Theory*, 25, 117–161.

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.

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.

MORANA, C., AND A. BELTRATTI (2004): "Structural change and long-range dependence in volatility of exchange rates: either, neither or both?," *Journal of Empirical Finance*, 11, 629–658.

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.

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

PERRON, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57(6), 1361–1401.

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.

SCHARTH, M., AND M. MEDEIROS (2009): "Asymmetric Effects and Long Memory in the Volatility of Dow Jones Stocks," *International Journal of Forecasting*, 27, 304–327.

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 of 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.

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.

TABLE 1. PARAMETER ESTIMATES UNDER CORRECT SPECIFICATION.

The table reports the average bias and the mean squared error (MSE) of the parameter estimates over 1000 replications. The estimation is carried out assuming correct model specification (M and p are assumed to be known).

	500 observations											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	-0.03	0.01	-0.04	0.10	0.01	0.02	0.01	0.00	-0.00	0.15	0.02	0.02
ϕ_{11}	–	–	0.20	8.13	0.74	71.34	–	–	0.24	10.01	0.18	49.44
ϕ_{12}	–	–	-0.22	8.19	-0.73	71.11	–	–	-0.22	10.01	-0.15	49.44
γ_1	–	–	-64.78	4304.77	26.39	904.27	–	–	5.99	133.29	26.26	900.00
γ_2	–	–	-64.51	4266.36	28.42	966.70	–	–	4.38	150.49	27.72	945.68
c_1	–	–	0.02	0.02	0.12	0.26	–	–	-0.02	0.02	0.15	0.31
c_2	–	–	-0.01	0.01	0.24	0.21	–	–	0.02	0.02	0.26	0.21
d	0.01	0.02	0.01	0.02	-0.04	0.02	-0.03	0.01	-0.06	0.01	-0.06	0.02
σ_u	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00
	1000 observations											
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
ϕ_{10}	-0.01	0.01	-0.01	0.06	0.00	0.01	0.01	0.00	0.01	0.05	0.02	0.01
ϕ_{11}	–	–	0.02	0.25	0.10	0.03	–	–	0.07	0.12	0.22	3.78
ϕ_{12}	–	–	-0.01	0.13	-0.09	0.02	–	–	-0.06	0.15	-0.20	3.74
γ_1	–	–	-62.81	4006.51	23.38	782.04	–	–	7.27	112.12	24.13	801.57
γ_2	–	–	-62.45	3957.86	24.32	802.35	–	–	6.27	123.78	25.94	862.33
c_1	–	–	0.01	0.01	-0.06	0.06	–	–	-0.02	0.01	-0.03	0.10
c_2	–	–	-0.01	0.01	0.22	0.11	–	–	0.02	0.01	0.24	0.14
d	0.00	0.01	0.00	0.01	-0.02	0.01	-0.02	0.00	-0.04	0.01	-0.04	0.01
σ_u	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00

TABLE 2. FREQUENCY OF CORRECT SPECIFICATION.

.

500 observations			
Model	Number of regimes (M)	Lag length (p)	Both (M and p)
1	0.97	1	0.97
2	0.41	1	0.41
3	0.23	1	0.23
4	0.97	1	0.97
5	0.26	1	0.26
6	0.19	1	0.19
1000 observations			
Model	Number of regimes (M)	Lag length (p)	Both (M and p)
1	0.98	1	0.98
2	0.94	1	0.94
3	0.59	1	0.59
4	0.95	1	0.95
5	0.73	1	0.73
6	0.55	1	0.55
5000 observations			
Model	Number of regimes (M)	Lag length (p)	Both (M and p)
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	1	1

TABLE 3. MODEL SPECIFICATION AND MEMORY PARAMETER ESTIMATION.

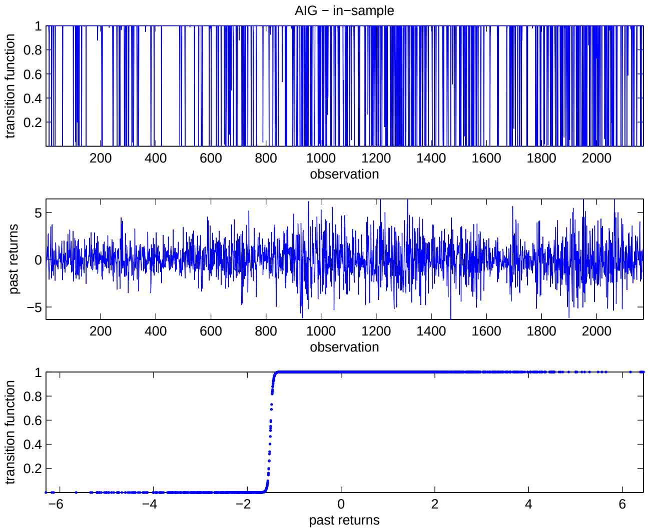
The table shows the ratio of mean squared errors (MSE) from different competing specifications and the heterogeneous autoregressive (HAR-RV) model. Numbers below one indicate that the competing model outperforms the HAR-RV benchmark. The HAR-RV model is estimated with averages over one, five, and 22 days. The competing specifications are as follows: (1) linear ARFIMA; (2) nonlinear ARFIMA with time as transition variable; (3) nonlinear ARFIMA with past daily returns as transition variable; (4) nonlinear ARFIMA with past cumulated returns over five days as transition variable; (5) nonlinear ARFIMA with past cumulated returns over ten days as transition variable; (6) nonlinear ARFIMA with past cumulated returns over 22 days as transition variable; (7) nonlinear ARFIMA with past cumulated returns over 66 days as transition variable; (8) nonlinear ARFIMA with past cumulated returns over 252 days as transition variable; (9) nonlinear ARFIMA with past average volatility over one day as transition variable; (10) nonlinear ARFIMA with past average volatility over five days as transition variable; (11) nonlinear ARFIMA with past average volatility over 22 days as transition variable. The models are estimated using the first 2170 observations.

	Specification																															
	(1)		(2)			(3)			(4)			(5)			(6)			(7)			(8)			(9)			(10)			(11)		
Series	d	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p	d	M	p
AA	0.41	1	0.32	1	1	0.41	1	1	0.41	1	1	0.41	2	1	0.41	1	1	0.41	1	1	0.41	1	1	0.28	1	1	0.30	2	1	0.31	1	1
AIG	0.45	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1
AXP	0.46	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	0	1	0.46	1	1	0.46	1	1	0.46	1	1
BA	0.40	1	0.40	1	1	0.40	1	1	0.40	1	1	0.40	2	1	0.40	2	1	0.40	1	1	0.40	0	1	0.40	1	1	0.40	3	1	0.40	0	1
CAT	0.39	1	0.38	2	1	0.39	1	1	0.39	1	1	0.39	2	1	0.39	1	1	0.39	1	1	0.39	2	1	0.39	1	1	0.39	1	1	0.28	1	1
DD	0.42	1	0.41	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	0	1	0.42	1	1	0.42	2	1	0.41	1	1
DIS	0.44	1	0.44	1	1	0.44	2	1	0.44	1	1	0.44	1	1	0.44	1	1	0.44	1	1	0.44	0	1	0.44	2	1	0.44	1	1	0.44	1	1
GE	0.45	1	0.45	2	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	1	1	0.45	0	1	0.45	1	1	0.45	1	1	0.45	1	1
GM	0.39	1	0.39	1	1	0.39	2	1	0.39	1	1	0.39	1	1	0.38	1	1	0.39	1	1	0.39	0	1	0.39	2	1	0.39	2	1	0.39	1	1
HD	0.43	1	0.43	1	1	0.43	1	1	0.43	1	1	0.43	2	1	0.43	1	1	0.43	1	1	0.43	0	1	0.43	1	1	0.43	1	1	0.43	1	1
HON	0.36	1	0.17	1	1	0.36	2	1	0.35	1	1	0.34	1	1	0.34	1	1	0.34	1	1	0.35	1	1	0.17	1	1	0.21	1	1	0.20	2	1
IBM	0.47	1	0.47	0	1	0.47	2	1	0.47	1	1	0.47	1	1	0.47	1	1	0.47	1	1	0.47	0	1	0.47	2	1	0.47	0	1	0.47	0	1
JNJ	0.42	1	0.42	0	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	0	1
JPM	0.47	1	0.35	1	1	0.47	1	1	0.47	1	1	0.47	1	1	0.47	1	1	0.47	1	1	0.47	0	1	0.47	1	1	0.35	3	1	0.46	1	1
KO	0.44	1	0.44	1	1	0.44	1	1	0.44	1	1	0.44	1	1	0.44	2	1	0.44	1	1	0.44	0	1	0.44	1	1	0.44	1	1	0.44	0	1
MCD	0.38	1	0.38	1	1	0.38	2	1	0.38	1	1	0.38	1	1	0.38	1	1	0.38	2	1	0.38	0	1	0.38	1	1	0.38	0	1	0.38	0	1
MMM	0.42	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1	0.42	1	1
MO	0.40	1	0.40	1	1	0.40	1	1	0.40	1	1	0.40	1	1	0.40	1	1	0.40	0	1	0.40	1	1	0.40	1	1	0.40	0	1	0.40	1	1
MRK	0.38	1	0.38	1	1	0.38	1	1	0.38	1	1	0.38	1	1	0.38	1	1	0.38	1	1	0.38	0	1	0.38	1	1	0.38	1	1	0.38	1	1
PFE	0.39	1	0.39	0	1	0.39	2	1	0.39	1	1	0.39	1	1	0.39	1	1	0.39	1	1	0.39	0	1	0.39	1	1	0.39	0	1	0.39	0	1
PG	0.46	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	1	1	0.46	0	1	0.46	1	1	0.46	0	1	0.46	2	1	0.46	1	1	0.46	1	1
UTX	0.43	1	0.33	2	1	0.43	1	1	0.43	1	1	0.43	1	1	0.43	1	1	0.43	1	1	0.43	0	1	0.50	1	2	0.36	1	1	0.36	2	1
WMT	0.40	1	0.39	2	1	0.40	1	1	0.40	1	1	0.40	1	1	0.40	2	1	0.40	1	1	0.40	0	1	0.40	1	1	0.39	1	1	0.40	1	1

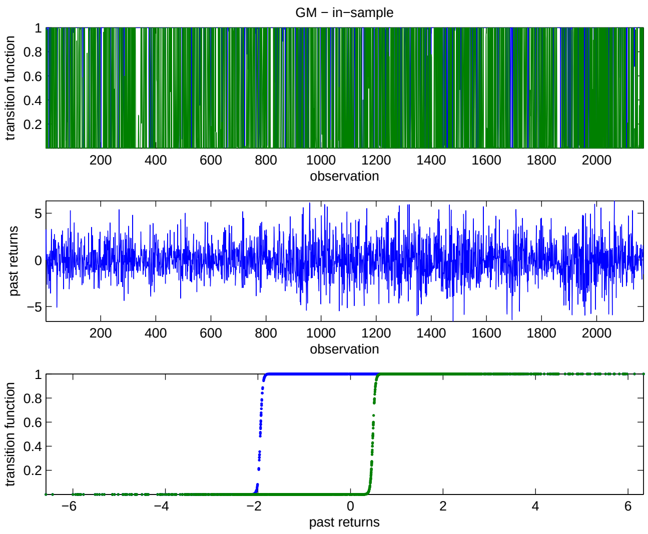
TABLE 4. FORECASTING RESULTS.

The table shows the ratio of mean squared errors (MSE) from different competing specifications and the heterogeneous autoregressive (HAR-RV) model. Numbers below one indicate that the competing model outperforms the HAR-RV benchmark. The HAR-RV model is estimated with averages over one, five, and 22 days. The competing specifications are as follows: (1) linear ARFIMA; (2) nonlinear ARFIMA with time as transition variable; (3) nonlinear ARFIMA with past daily returns as transition variable; (4) nonlinear ARFIMA with past cumulated returns over five days as transition variable; (5) nonlinear ARFIMA with past cumulated returns over ten days as transition variable; (6) nonlinear ARFIMA with past cumulated returns over 22 days as transition variable; (7) nonlinear ARFIMA with past cumulated returns over 66 days as transition variable; (8) nonlinear ARFIMA with past cumulated returns over 252 days as transition variable; (9) nonlinear ARFIMA with past average volatility over one day as transition variable; (10) nonlinear ARFIMA with past average volatility over five days as transition variable; (11) nonlinear ARFIMA with past average volatility over 22 days as transition variable.

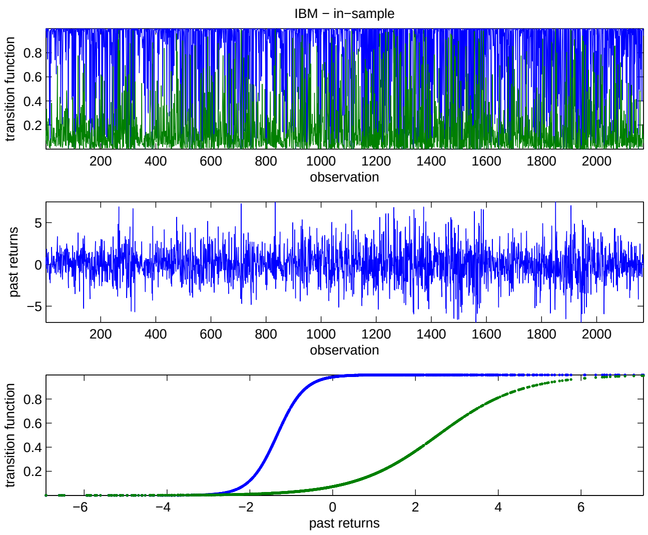
Series	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
AA	0.99	1.09	0.99	0.99	1.00	0.99	1.03	1.02	0.98*	0.99	0.99
AIG	0.99	1.01	0.97**	0.98	0.96*	0.98*	0.98**	0.99	0.98*	0.98*	0.99
AXP	1.01	1.05	0.96**	0.98	1.00	1.00	1.00	1.01	0.97*	0.98*	0.98
BA	1.00	1.06	0.99**	0.98	0.99	0.99	0.97	1.00	0.98*	1.00	1.00
CAT	1.01	1.02	0.98**	0.98	1.00	1.01	0.99	1.01	0.98*	0.99	0.98
DD	1.04	1.08	1.01	1.02	1.02	1.03	1.04	1.04	1.00	0.99	1.00
DIS	1.01	1.07	0.98*	1.00	1.03	1.00	1.02	1.01	0.98*	0.99	0.99
GE	0.99	1.00	0.95*	0.97	0.96	0.98	0.99*	0.99	0.99	0.98	0.98
GM	1.01	1.02	0.99*	0.98	1.00	1.00	0.97*	1.01	1.00	1.00	1.00
HD	1.01	1.04	0.98*	0.98	0.99	1.00	1.00*	1.01	0.99	0.99	0.99
HON	0.95*	1.52	0.93**	0.93*	0.94	0.93	0.94**	0.93*	0.94	0.94	0.98
IBM	0.99	0.99***	0.93**	0.95*	0.96	0.98	0.98*	0.99	1.02	0.99	0.99
JNJ	0.99	0.99***	0.94**	0.98*	0.96*	0.97	0.99*	1.01	0.99	0.98	0.99
JPM	1.00	1.20	0.97**	0.98*	0.98*	0.98	0.99*	1.00	0.98	0.98	0.99
KO	1.01	1.03	0.99**	0.99*	1.00	1.03	1.01	1.01	1.00	1.00	1.01
MCD	1.00	1.03	0.99**	1.00*	1.00	0.99	1.01	1.00	0.99	1.00	1.00
MMM	0.99	1.01	0.96**	0.97*	0.98	0.97	0.98*	0.97	0.97	0.97	0.97
MO	1.00	1.01	0.97**	0.98*	0.98	0.98	1.00	0.98	0.99	1.00	0.98
MRK	0.99	1.01	0.99**	1.00*	1.00	0.99	1.00	0.99	0.99	0.99	0.99
PFE	1.00	1.00	0.99**	1.00*	1.01	1.00	1.03	1.00	0.99	1.00	1.00
PG	1.03	1.04	1.02	1.03	1.03	1.03	1.03	1.03	1.04	1.02	1.04
UTX	0.99	1.17	0.97**	0.97*	0.98	0.98	0.99	0.99	0.99	0.97	1.02
WMT	1.03	0.91***	0.99*	1.03	1.03	1.06	1.05	1.03	0.99	1.02	1.01



(a)



(b)



(c)

FIGURE 1. Transition functions and transition variables. The panels report, top to bottom, the time evolution of the transition function, the transition variable, and the scatter plot of the transition function versus the transition variable. Plot (a): AIG. Plot (b): GM. Plot (c): IBM.

(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